

Université Paris Diderot

HABILITATION À DIRIGER DES RECHERCHES

présentée par

Samuel BOTTANI

le 29 novembre 2019

Structure et dynamique de réseaux biologiques : de la biologie des systèmes aux réseaux neuronaux en culture

JURY

Parrain:

Vincent FLEURY, Directeur de recherche CNRS

Laboratoire Matières et Systèmes Complexes UMR CNRS 7057 – Université Paris Diderot :

Rapporteurs:

Hugues BERRY, Directeur de recherche INRIA

LIRIS UMR CNRS 5205 – Université de Lyon

Annick LESNE, Directrice de Recherche CNRS

Laboratoire de physique théorique de la matière condensée UMR CNRS 7600 – Sorbonne Université

Membres du jury:

Claude MEUNIER, Directeur de Recherche CNRS

Centre de Neurophysique, Physiologie et Pathologie UMR CNRS 8119 – Université Paris Descartes

Reiner VEITIA, Professeur des Universités

Institut Jacques Monod UMR CNRS 7592 – Université Paris Diderot

Catherine VILLARD, Directrice de Recherche CNRS

Centre de Recherches de l'Institut Curie, Institut Pierre Gilles de Gennes pour la microfluidique, Laboratoire PhysicoChimie UMR CNRS 168 – Institut Curie Section de Recherche

Université Paris Diderot

Laboratoire Matières et Systèmes Complexes

UMR CNRS 7057, Paris, France

Table des matières

Table des matières	iii
Liste des figures	v
Liste des encadrés	ix
I Introduction Générale	1
1 Présentation Biographique	7
1.1 Exposé synthétique des recherches	8
1.2 Curriculum Vitae	11
1.3 Références	21
2 Théorie et Biologie	23
2.1 Renouveau de la théorie et de la modélisation en biologie	27
2.2 Développement de la modélisation mathématiques en biologie	29
2.3 Biologie à haut débit et quantitative	35
2.4 La biologie des systèmes	37
2.5 Biotechnologies et biologie de synthèse	41
2.6 Les réseaux complexes	46
2.7 Conclusion du chapitre	58
2.8 Références	58
3 Approches de modélisation	65
3.1 Modélisation statistique	68
3.2 Modélisation théoriques de mécanismes.	69
3.3 Inférence de paramètres	82
3.4 Conclusion du chapitre	89
3.5 Références	90
II Biologie des Systèmes	99
4 Systèmes de régulation génétiques	101
4.1 Introduction au chapitre	102
4.2 Réseaux biochimiques	109
4.3 Modélisation mathématique biophysique de l'expression génétique	114
4.4 Approches mathématiques et réductions de modèles	119
4.5 Conclusion modélisation en biologie des systèmes	127
4.6 Références	127

5	Biologie des systèmes : contributions	135
5.1	Contributions : survol	136
5.2	Caractérisation des réseaux génétique	137
5.3	Circuits génétiques oscillants	149
5.4	Caractérisation de la réponse non linéaire de la transcription.	165
5.5	Dosage génétique, aneuploïdie et déséquilibres génomiques	172
5.6	Variabilité phénotypique et relation génotype/phénotype	194
5.7	Autres travaux de biologie ds systèmes : Réponse au stress osmotique et contrôle temps réel de l'expression	204
5.8	Conclusion du chapitre	204
5.9	Références	205
III	Neurophysique	217
6	Approches physiques pour les cultures neuronales	219
6.1	Présentation	220
6.2	Les cultures de neurones dissociées	222
6.3	Applications des cultures de neurones	238
6.4	Vers des processeurs neuronaux	241
6.5	Activité spontanées des cultures	247
6.6	Programmer les cultures neuronales	255
6.7	Modélisation des cultures neuronales	262
6.8	Conclusions et perspectives : vers la neurobiologie synthétique.	267
6.9	Références	268
7	Neurophysique des cultures neuronales : contributions	287
7.1	Introduction au chapitre	288
7.2	Modèles de percolation	289
7.3	Bouffées synchronisées de potentiels d'actions (Network bursting)	308
7.4	Développement neuronal et croissance de réseau	319
7.5	Conclusion du chapitre	333
7.6	Références	335
8	Conclusion générale	341
8.1	Références	345
IV	Annexe	347
A	Glossaire	III

Liste des figures

1	Carte conceptuelle de mes thèmes de recherche et approches principaux.	6
2.1	Loi de Zipf de la distribution de tailles des villes.	25
2.2	D’Arcy Thomson : transformation de formes de poissons	26
2.3	Reproduction de la première équation du modèle mathématique d’épidémiologie par Daniel Bernoulli publié en MDCCLX.	30
2.4	Citation originale d’Isidore Geoffroy Saint Hilaire	31
2.5	Ngram des termes « Systems Biology », « Integrative biology » et « Synthetic biology ».	38
2.6	Synthetic Aesthetics.	42
2.7	Le « règne synthétique » du vivant d’Alexandra Daisy Ginsberg	42
2.8	Assemblage <i>idempotent</i> de séquences génétiques	43
2.9	Exemple de fiche de « biobrick » de la « Registry of Standard Biological Parts ».	44
2.10	Recouvrement entre l’ingénierie métabolique et la biologie synthétique	45
2.11	Variétés de nature de graphes complexes.	46
2.12	Dessin historique de réseau neuronal	47
2.13	Diagramme conceptuel simplifié des interactions alimentaires dans la baie False Creek.	48
2.14	Réseau d’amitiés dans un lycée américain.	48
2.15	Exemples de réseaux très centralisé, décentralisé et distribué.	49
2.16	Exemples de distributions de connectivité de réseaux complexes réels.	50
2.17	Exemples de trois modèles de réseaux : Erdős-Rényi, Barabási-Albert et hiérarchique.	51
2.18	Carte des interactions protéine-protéine de la levure.	52
2.19	Diamètre d’un réseau complexe	53
2.20	Illustration de la définition du coefficient de clustering).	53
2.21	Exemple de communautés sur un réseau.	53
2.22	Modèle de réseau de Watts & Strogatz.	55
2.23	Clustering moyen et diamètre dans le modèle de petit monde de Watts & Strogatz.	55
3.1	Carte conceptuelle de la démarche de modélisation.	67
3.2	Les simulations multi-physiques.	71
3.3	Exemple de suite logicielle de simulation multi-physique.	71
3.4	Exemple d’environnement de simulation pour les biotechnologies et le pharmaceutique.	72
3.5	Exemple d’environnement modélisation de systèmes biologiques cellulaires : VCell	73
3.6	Résumé graphique de l’article de Henri Markram <i>et al.</i> des premiers résultats de l’Human Brain Project.	74
3.7	Différentes étapes de la modélisation du Human Brain Project.	75
3.8	Comportements dynamiques des différents types de neurones.	83
3.9	Réseau et modèle du cycle cellulaire.	85
4.1	Attracteurs de la dynamique	103
4.2	Modèle physique de pyllotaxie de la fleur de tournesol	104
4.3	Modèle physique de morphogénèse du poumon	105

4.4	Physique du développement du poulet!	106
4.5	Tourbillons de la vie : flux cellulaires durant l'embryogénèse.	106
4.6	Carte du métabolisme cellulaire : le poster Boehringer Mannheim	110
4.7	Réseaux génomique, protéomique et métabolique	111
4.8	À la recherche de modularité dans les réseaux.	113
4.9	Différentes approches de représentation de la courbe de réponse de la régulation génétique	115
4.10	Exemple d'un modèle effectif logique : lyse et lysogénie de phage λ	116
4.11	Illustration du langage Kappa	118
5.1	Réseau de réparation SOS	139
5.2	Profils d'expression des 8 gènes du réseau SOS	140
5.3	Comparaison du réseau SOS connu et inféré	141
5.4	Réseau d' <i>inter</i> -régulation transcriptionnelle chez la levure.	142
5.5	Connectivité du réseau d' <i>inter</i> -régulation transcriptionnelle chez la levure.	143
5.6	Schémas de connexions à 3 gènes et motifs de réseau	145
5.7	Exemple du rôle fonctionnel attendu d'un motif de type « Feed Forward Loop » avec un opérateur « AND » sur le gène en sortie.	146
5.8	Types de motifs d'ordre $n = 2$ et $n = 3$ pour le réseau mixte transcriptionnel et protéines-protéines.	147
5.9	Instances des motifs dans la levure.	148
5.10	Profils phylogénétiques des interactions.	149
5.11	Oscillations pulsées de P53 en réponse à un stress de rayonnement	153
5.12	Réponse de p53 à des stress	154
5.13	Diagramme du circuit minimal p53-Mdm2	154
5.14	Simulations du modèle minimal de l'oscillateur central p53-Mdm2	156
5.15	Diagramme de bifurcation en fonction de la demie-vie de MDM2	156
5.16	Diagramme du circuit de régulation.	157
5.17	Diagramme de bifurcation stochastique et déterministe du modèle complet du circuit p53-Mdm2.	158
5.18	Évolution stochastique du modèle complet du circuit p53-mdm2.	161
5.19	Modèle stochastique complet en absence de retard, avec des taux d'association faibles.	161
5.20	Modélisation déterministe par les équations différentielles avec retard	162
5.21	Distribution des amplitudes des oscillations stochastiques pour le modèle stochastique complet et toutes les réductions.	163
5.22	Fonctions d'autocorrélation pour les différentes réductions.	164
5.23	Régulation du gène GAL4	166
5.24	Schéma de la régulation de la transcription.	167
5.25	Interactions entre un promoteur et des protéines facteurs de transcription coopératives ou synergétiques.	168
5.26	Courbe de réponse transcriptionnelle sigmoïde	168
5.27	Fraction d'occupation	171
5.28	Effet de la sur-expression (en raison de la duplication d'un gène) d'une sous-unité sensible au dosage dans un complexe macromoléculaire.	177
5.29	Compensation locale de dosage.	178
5.30	Compensation de déséquilibre de dosage au niveau des protéines par protection contre la protéolyse, exemple sur un trimère ABA.	180
5.31	Effet tampon sur les effets de dosage dans la transcription dû aux liaisons non spécifiques.	181
5.32	Dynamique de formation d'un dimère MM et duplication de génome.	184
5.33	Conséquences sur une cinétique biochimique d'une duplication de génome suivie de suppression massive d'ADN.	185

5.34 Liaison spécifique de facteurs de régulation en présence de différents niveaux de liaison non spécifique.	185
5.35 Relations non linéaire entre des valeurs représentant le génotype (G) et le phénotype (P)	188
5.36 Modèle simple de dominance génomique lors d'hybridation	190
5.37 Conséquences sur l'occupation d'un promoteur suite à l'hybridation de deux organismes.	192
5.38 Mécanismes de compensation de dosage du chromosome X	192
5.39 Expression de gène avec décroissance exponentielle d'accessibilité de la chromatine	193
5.40 Paysage adaptatif.	195
5.41 Reconstructions de séquences temporelles de profils d'expression génétique dans la réponse au stress hyper-osmotique	199
5.42 Viabilités des différentes souches à un stress intense et rang issu du clustering	200
5.43	201
5.43 Variation continue de la survie durant un stress hyper-osmotique sévère activité de la voie de réponse de cultures post-diauxiques.	202
5.44 Modèle de prise de risque (« bet-hedging ») continu de la réponse au stress hyper-osmotique.	203
6.1 Exemple de réseau de neurone en culture	221
6.2 Exemple de puce microfluidique pour neurones	224
6.3 Méthodes de microfabrication. Principe de gravure photolithographique.	225
6.4 Description d'un système de culture d'organoïdes cérébraux.	227
6.5 Organoïdes cérébraux comme outil d'étude de désordre neurologiques	227
6.6 Culture de neurones avec matrice d'électrodes	229
6.7 Réseau de neurones en culture en développement et liaisons synaptiques potentielles	234
6.8 Suivi au jour le jour de la dynamique de croissance de neurites	240
6.9 Schéma de conception pour la reproduction dans un dispositif de culture microfluidique du chemin initial de propagation la maladie d'Alzheimer.	241
6.10 Dispositif à diode.	245
6.11 Implémentation d'une diode.	245
6.12 Implémentation d'une fonction seuil.	246
6.13 Implémentation de la fonction « AND »	246
6.14 Implémentation d'un oscillateur.	246
6.15 Développement de bouffées de réseau en culture	248
6.16 Activité sous forme d'avalanches au sein de bouffées synchronisées.	251
6.17 L'« unorganized machine » de Turing	255
6.18 Modèle classique de renforcement synaptique suivant Hebb.	257
6.19 Spike Time Dependent Plasticity (STDP)	258
7.1 Dispositif expérimental de l'équipe d'Elisha Moses pour la mesure de la réponse d'une culture à des stimuli d'intensité croissante.	290
7.2 Diagramme de phase de la réponse à une stimulation.	291
7.3 Évolution de la fraction de neurones activés de l'équation auto-consistante.	294
7.4 Comparaison entre les solutions Monte Carlo l'équation auto-consistante.	295
7.5 Comparaison entre les solutions Monte Carlo l'équation auto-consistante.	295
7.6 Comparaison de la simulation explicite Monte carlo et du calcul à champ moyen.	297
7.7 Modèle de percolation à quorum : approche du point critique	297
7.8 Modèle de percolation à quorum : effet de la résolution sur l'approche du point critique.	298
7.9 Solutions de l'équation auto-consistante.	299
7.10 Dépendance de l'activité du réseau pour des désordres du quorum.	299
7.11 Dépendance de la position de la transition du premier ordre (saut) avec le désordre.	300

7.12	Modèle de percolation à quorum avec inhibition	301
7.13	Modèle de percolation à quorum avec inhibition : comparaison des méthodes de résolution.	302
7.14	Modèle de percolation à quorum en fonction de la fraction d'inhibition.	304
7.15	Modèle de percolation à quorum avec un modèle dynamique de neurone.	305
7.16	Dispositif neuronal simple à deux chambres.	306
7.17	Induction et détection optique de bouffées d'activité.	307
7.18	Stimulation locale et propagation de bouffée.	307
7.19	Transmission asymétrique.	308
7.20	Trame des bouffées synchronisées de réseau.	311
7.21	Évolution temporelle du modèle adaptatif exponentiel integrate & fire AEIF.	312
7.22	Attracteur de la dynamique simulée sur un réseau Gaussien	313
7.23	Schéma du modèle équivalent des bouffées synchronisées.	313
7.24	Trajectoire dans l'espace des phases de l'attracteur des bouffées synchronisées. Description dans le texte.	316
7.25	Trajectoire d'une bouffée pour des neurones synchronisés couplés par une synapse « continue ».	317
7.26	Matrice de corrélation pour les caractéristiques principales de la dynamique des bouffées en fonction des paramètres neuronaux du modèle AEIF	317
7.27	Tentative de design pour un mélangeur de connexions neuronales.	319
7.28	Croissance d'axones non contraints	321
7.29	Croissance d'axones non contraints	322
7.30	Affinité des cônes de croissance aux bordures d'obstacles.	323
7.31	Test du modèle de croissance : séparation asymétrique.	324
7.32	Neurites générées par le modèle « run and tumble ».	326
7.33	Schéma de construction d'une diode neuronale en « U »	327
7.34	Schéma de construction d'une diode neuronal en « U »	328
7.35	Comparaison du modèle BEST original et optimisé.	328
7.36	Angles et diamètres des nouvelles branches dendritiques	329
7.37	Étude de la dynamique du modèle de compétition pour 3 cônes de croissance en interaction.	330
7.38	Compétition dépendante du diamètre.	330
7.39	Reconstitutions de différents types de neurones	331
7.40	Exemples de génération de morphologies dans des cultures confinées.	332

Liste des encadrés

2.1	Allométrie	28
2.2	De nombreux types de réseaux complexes.	49
2.3	Descripteurs topologiques principaux.	54
3.1	« Flux Balance analysis » et « Metabolic Control Analysis »	86
3.2	Théorème de Bayes	88
4.1	Représentation schématique d'un régulation activatrice de l'activité d'un gène	117
4.2	L'algorithme de Gillespie	122
4.3	Modélisation de réactions chimiques stochastiques	123
4.4	Simplification des modèles.	125
5.1	Aux origines de la périodicité du vivant : échelles de temps et hors d'équilibre	152
5.2	Dérivation de la fonction de Hill pour la fonction de réponse transcriptionnelle	169
5.3	Définitions autour de la ploïdie, d'après Wikipédia : ploïde et aneuploïde.	173
7.1	Liste de sujets futurs sur les cultures neuronales.	335

Première partie

Introduction Générale

Préface

Mes activités de recherche se situent à la frontière entre biologie et physique. Elles ont porté sur des questions de biologie couvrant un large spectre d'études qui se retrouvent en grande partie dans le courant de la biologie des systèmes et plus récemment dans les neurosciences. Les sujets que j'ai abordé concernent ainsi la régulation génétique, les voies de transduction de signal, la biologie synthétique et la génomique pour le premier ensemble et enfin les cultures neuronales. Ces thèmes assez variés correspondent à une évolution constante de mes activités de recherche guidée par la curiosité pour les mécanismes d'organisation dans le vivant et leur étude sous l'angle de la physique de la complexité.

Des phénomènes comme la synchronisation d'activités électriques de cellules cardiaques et de neurones, et la synchronisation collective de populations d'insectes m'avaient déjà servi il y a longtemps déjà dans ma thèse. Elles avaient servi comme inspiration pour des études théoriques d'auto-organisation dans des systèmes à grand nombre de degrés de liberté hors d'équilibre thermodynamique. L'ordre, spatial, ou temporel apparaît comme un état « attracteur » de la dynamique collective vers lequel le système converge spontanément pour la plupart de ses conditions initiales.

Fondement de cette vision du vivant est l'idée selon laquelle les mécanismes de régulation biologiques correspondent à des dynamiques complexes d'interactions biochimiques ou cellulaires, soumises à des principes de fonctionnement fondamentalement physiques de nature à assurer la fiabilité et la robustesse nécessaire au vivant. C'est vers une exploration de tels principes, et éventuellement du développement de capacités pour les contrôler, que porte mon orientation scientifique. Une recherche marquée par une orientation interdisciplinaire forte vers la biologie dans les secteurs de la génétique, de la biologie des systèmes et synthétique et plus récemment des neurosciences. Des travaux qui ont demandé l'apprentissage d'autres domaines et langages que la physique pour contribuer à des publications dans des revues parfois inhabituelles pour des physiciens.

Depuis mon recrutement comme Maître de Conférences à l'Université Paris Diderot en 1997 je me suis investi dans les différents aspects complémentaires de la carrière universitaire que je considère comme une combinaison indispensable du développement des connaissances scientifiques et savoirs-faire techniques et de leur transfert aux jeunes générations et à la société dans son ensemble. Je me suis efforcé de répartir mes activités sur tous les champs de notre mission universitaire :

- Un engagement en enseignement attentif à la pédagogie, aux technologies et aux recherches en éducation qui m'a amené à créer de nouveaux enseignements mettant en œuvre de nouveaux contenus et approches, ainsi qu'à être porteur ou participant à plusieurs reprises à la création de formations à différents niveaux d'études, tenant compte des aspects professionnalisant des cursus.

- Une prise de responsabilités progressive au sein de nos institutions, comme porteur de projet pour la création de licences professionnelles au sein de l'UFR de physique, à la codirection de l'école doctorale ED 474 de 2008 à 2014 et depuis 2013 la direction du département Centre de Formation des Doctorants à l'insertion professionnelle pour la Comue Université Sorbonne Paris Cité.

Mon parcours n'est donc pas celui d'une spécialisation mais d'une exploration interdisciplinaire avec une démarche théorique d'inspiration physique impliquant simplification et conceptualisation des processus biologiques étudiés et finalement formalisation des modèles ouvrant à des études analytiques ou numériques. Dans l'introduction de ce mémoire je commencerai par discuter le cadre de ces travaux théoriques pour les sciences de la vie, qui a connu un essor considérable ces vingt dernières années. La période même de mon activité. Mon orientation scientifique est résolument aux interfaces entre physique, mathématiques et sciences de la vie où j'ai fait l'expérience du subtil travail à cheval entre des pratiques disciplinaires différentes. Ce positionnement a joué un rôle central dans mon parcours de responsabilités au sein de l'université, j'y reviendrai.

L'idée que notre époque scientifique est portée par et correspond à une intégration de techniques et savoirs de natures et d'origines différentes a souvent été reprise depuis le tournant du millénaire. Même si un laboratoire de biologie moléculaire reste une pièce humide avec ses paillasses, hottes, alignements de bocal de réactifs, éprouvette et pipettes, son activité intègre de plus en plus les mathématiques, l'informatique et l'électronique qui ont permis d'obtenir et modéliser les séquences, les gènes, protéines et autres assemblages nécessaires à ses recherches.

Les domaines de questionnement scientifique en biologie concernés par mes travaux sont d'une part le champ de la biologie des systèmes et les neurosciences d'autre part, plus particulièrement des sous-domaines dans lesquels une approche de physicien peut contribuer. Ce document est donc organisé suivant les domaines biologiques principaux auxquels je me suis intéressé. D'abord sur la biologie des systèmes et la génomique, ensuite sur les questions en liens avec les neurosciences.

Présentation du document Ce document présentera le contour de mon activité scientifique portant sur des approches théoriques de sujets biologiques avec en filigrane un questionnement sur l'organisation et les dynamiques collectives. Les sujets précis auxquels je me suis intéressé avec une méthodologie de physicien appartiennent à des domaines différents de la biologie. En conséquence ce document se compose de deux parties principales suivant les domaines des sciences du vivant concernés, « Biologie de systèmes » dans la partie II et « Neurophysique dans la partie » III, ainsi que d'une « Introduction générale » dans la partie I initiale qui définit le cadre commun. La synthèse des travaux et publications demandée dans le cadre d'une HDR scientifique correspond probablement d'avantage au chapitre 5 pour mes recherches dans le champ de la biologie des systèmes et au chapitre 7 pour celles dans le domaine des neurosciences. Chacun de ces deux chapitres spécialisés est précédé dans la partie correspondante, par un chapitre de mise en perspective relatif à la problématique scientifique et les principes des approches théoriques des sujets que j'ai abordé.

La première partie de ce mémoire d'introduction générale discute de tendances et de pratiques interdisciplinaires traversant de nombreux secteurs des Sciences du vivant depuis une vingtaine d'années et faisant un usage intensif de modélisation mathématique et physique ainsi que de méthodes d'informatiques. Il s'agit du cadre dans lequel se placent les deux directions de mes contributions scientifiques.

La rédaction de ce mémoire, à un stade déjà assez avancé de ma carrière m'a amené à un travail rétrospectif sur les fondements des courants de la recherche et de ses méthodologies qui ont constitué le cadre de mon activité scientifique ces vingt dernières années. C'est l'objet de l'introduction générale de ce document. Écrite avec un esprit légèrement épistémologique, la rédaction de cette partie a été pour moi un travail critique de contextualisation nécessaire pour exprimer les lignes conductrices et l'unité dans les différents sujets et domaines où j'ai évolué scientifiquement. Les sciences de la vie connaissent en effet depuis une vingtaine d'années un essor important d'apports et de pratiques interdisciplinaires faisant un usage intensif de modélisation mathématique et physique ainsi que de méthodes d'informatiques. C'est dans ce contexte que j'ai pu poursuivre les deux directions de mes contributions scientifiques. Les différentes parties de ce

document doivent pouvoir se lire indépendamment, jury d'hdr et lecteurs éventuels pourront se rapporter directement aux parties les concernant le plus!

Avec un parcours de recherche déjà ancien, je n'inclus pas dans ce mémoire les travaux des années de post-docs après ma thèse car très différents de mes activités ultérieures. Portant sur des développements instrumentaux pour le séquençage à haut débit, ces travaux cependant ont eu une influence majeure sur mon orientation de recherche. Se déroulant dans la seconde partie des années 90 ils m'ont permis de découvrir la révolution génomique de cette époque et les nouveaux questionnements interdisciplinaires devenant d'actualité et de me former.

Dans l'époque formidable de foisonnement de la recherche biologique des vingt dernières années, je considère comme un privilège la possibilité d'avoir pu aborder des sujets dans des domaines aussi variés que la génétique, ses développements à grande échelle en génomique et biologie des systèmes, jusqu'au champ très large des neurosciences. La figure 1 résume graphiquement les différents champs et sujets de mes recherches et les démarches communes mobilisées.

Je mentionnerais pour finir cette préface, que les recherches présentées dans ce mémoire ne couvrent qu'une partie, de mes activités universitaires durant la vingtaine d'années pendant laquelle elles se sont déroulées. Je me suis également investi fortement sur le volet enseignement de mon poste d'universitaire. Aussi bien dans la préparation de cours et travail avec les étudiants que dans la création de nouveaux cours et le montage de nouveaux curricula dans tous les cycles d'études de la licence au doctorat. L'orientation et l'ouverture professionnelle des étudiants est en particulier un thème auquel j'ai prêté une forte attention, probablement en raison d'expériences extra-universitaires avec ma participation durant deux années de disponibilité à des projets de création d'entreprise. Mais ces différentes activités universitaires et de valorisation de la recherche sortent du cadre de ce mémoire et de l'habilitation et je n'en parlerai pas davantage hors de leur inclusion dans le CV.

Chapitre 1

Présentation Biographique

Sommaire

1.1 Exposé synthétique des recherches	8
1.1.1 Biologie des systèmes et expression génétique	8
1.1.2 Neurosciences et neurophysique	9
1.2 Curriculum Vitae	11
1.2.1 Formation	11
1.2.2 Publications	12
1.2.3 Invitations et séminaires :	14
1.2.4 Enseignements	14
1.2.5 Ingénierie de formation	16
1.2.6 Encadrement	16
1.2.7 Animation scientifique	18
1.2.8 Organisation de colloques et rencontres scientifiques	18
1.2.9 Responsabilités et implication dans la vie collective de l'établissement	19
1.2.10 Direction des études ED 474	19
1.2.11 Direction Centre de Formation des Doctorant.e.s à l'insertion professionnelle Université Sorbonne Paris Cité	20
1.2.12 Autres responsabilités et implications la vie collective de l'établissement :	20
1.3 Références	21

Ce chapitre présente dans un premier temps de manière synthétique mes différentes activités de recherche depuis environ le début des années 2000 et qui seront repris en détails dans les parties suivantes de ce mémoire. Un parcours de recherche déjà assez long qui m'a amené à m'intéresser à différents domaines d'application aux sciences du vivant, de méthodes théoriques et de modélisation issues de la physique dans les domaines de la biologie des systèmes et les neurosciences. Malgré la diversité des domaines et des sujets un cadre commun général existe autour du paradigme des réseaux d'interactions et des systèmes complexes. Ce document est assez exhaustif de mes travaux, car il a été l'occasion importante de réfléchir rétrospectivement sur les fondements et hypothèses liant les différents sujets traités. Je n'inclus toutefois pas ici les travaux précédant le tournant du millénaire, durant mon post doc à l'Institut de Physique Nucléaire d'Orsay, ayant porté sur l'instrumentation pour le séquençage de l'ADN, une direction que je ne n'ai pas poursuivi. Ces travaux sur le séquençage, au moment même de la course vers le séquençage du génome humain, ont cependant été moteurs à mon orientation ultérieure vers la génomique et la biologie des systèmes.

La seconde partie de ce chapitre présente mon curriculum vitae, publications, activités d'encadrement et responsabilités collectives.

1.1 Exposé synthétique des recherches

Mes activités portent dans le domaine général de la physique statistique et des systèmes dynamiques et appliquées à l'expression génétique, les voies de transduction de signal, la biologie synthétique et les neurosciences. Ces orientations correspondent à une évolution de mes intérêts déjà durant ma thèse théorique (soutenue en 1996) sur l'étude de comportements d'auto-organisation de systèmes à grand nombre de degrés de liberté hors d'équilibre thermodynamique dont des systèmes complexes d'inspiration biologique avec un état attracteur ordonné (synchronisation de cellules, neurones et populations d'insectes). Les mécanismes de régulation biologiques correspondent en effet à des dynamiques complexes d'interactions biochimiques ou cellulaires qui doivent être soumises à des principes de fonctionnement généraux afin d'assurer la fiabilité et robustesse nécessaire au vivant. C'est vers une meilleure compréhension de ces principes et le développement de capacités pour contrôler des dynamiques biologiques que porte mon orientation scientifique.

1.1.1 Biologie des systèmes et expression génétique

Je me suis intéressé dans mes travaux à différentes approches de modélisation pour aider à comprendre l'organisation et la dynamique de l'expression du génome et des phénomènes évolutifs liés. Sur une période désormais assez longue, les thèmes assez variés sur lesquels j'ai mené modélisation et études théoriques ont ainsi concerné la reconstruction puis la caractérisation d'architectures de réseaux de régulation transcriptionnelle incluant aussi un niveau d'interactions de protéines (chapitre 4, section 5.2); le comportement théorique de petits circuits génétiques types impliqués dans des oscillations d'activité (chapitre 4, section 5.3); les contraintes imposées au niveau génomique par les variations de nombre de copies de gènes et leurs conséquences (chapitre 4, section 5.5); ainsi que des questions sur la variabilité dans la relation génotype/phénotype dans le cadre d'une étude de la réponse au stress hyper-osmotique de la levure (chapitre 5, section 5.6).

L'étude de l'expression du génome dispose depuis la fin des années 1990 de moyens toujours plus considérables pour caractériser les niveaux d'expression de grands nombres de gènes. Ces mesures sont possibles au niveau de la transcription par l'observation des niveaux de transcription des ARN messagers, mais aussi au niveau des protéines. Avec la généralisation de ces techniques il est devenu plus facile de prêter attention aux activités corrélées des gènes et de nouveaux paradigmes se sont diffusés. Ainsi plutôt que de mettre l'accent sur les niveaux absolus d'expression, il devient de plus en plus clair que les corrélations sont déterminantes. L'importance des rapports d'équilibres plutôt que des niveaux d'abondance eux même se retrouve également dans des questions classiques de génétique comme l'origine de la dominance et récessivité d'allèles.

Ce sujet aborde notamment la question des effets du dosage de gènes. On entend par là le nombre de copies d'un même gène, c'est-à-dire du nombre de copies d'une séquence codant pour une protéine, dans le génome d'organismes. Les génomes des animaux, insectes et plantes ont généralement deux ou davantage de copies de leurs chromosomes. En conséquence deux fois, ou d'avantage, de copies des mêmes gènes. Ce sont les allèles d'un gène donné. Ces allèles peuvent être identiques, à savoir correspondre à exactement à la même séquence d'ADN, ou porteuses de différences, certaines altérations de la séquence codante donnant des variantes de protéines produites. Le nombre de copies de gènes a donc un rapport direct avec la question classique de génétique des mécanismes et raisons de la dominance de certains allèles sur d'autres. Si un des avantages de la diploïdie pourrait être la simple redondance assurant une continuité fonctionnelle en cas de mutations délétères sur un des allèles, de nombreuses observations discutées dans [VEITIA \[2005\]](#) indiquent que les allèles mêmes identiques ne se substituent pas et que les deux copies doivent être simultanément actives pour une physiologie normale. C'est le sujet de l'haploinsuffisance qui survient quand un allèle en copie unique ne suffit pas à la fonction.

Ces considérations ont conduit à la formulation de la théorie de l'équilibre de dosage contrai-

gnant l'adaptation génétique naturelle. Les gènes étant inter-dépendants et les fonctions dépendantes des niveaux quantitatifs d'expression des protéines associées tout déséquilibre résultant de délétions hétéro-zygotes ou de sur-expression sera défavorisé? L'adaptation naturelle doit avoir tendance à minimiser les déséquilibres qu'elle induirait. En conséquence cette théorie prédit qu'un déséquilibre sur le dosage d'un gène -changement de son nombre de copies dans le génome- diminution de moitié ou gain, serait compensé, au moins partiellement, par une adaptation de l'expression de ses partenaires d'interaction.

Étude de propriétés de la fonction génotype-phénotype sur le cas de la réponse au stress hyper-osmotique Cherchant un modèle d'étude d'une fonction cellulaire avec entrées et sorties suffisamment bien caractérisées pour de la modélisation et l'exploration de la variabilité des réponses, j'ai eu l'opportunité de m'associer aux travaux sur le système de réponse au stress hyper-osmotique de la levure que Pascal Hersen avait montré à partir de 2008 au laboratoire MSC. Ces travaux que je ne développerai pas dans ce mémoire, m'ont amené à poursuivre l'étude de la variabilité de la fonction de réponse au stress hyper-osmotique en collaboration avec Suzannah Rutherford du Fred Hutchinson Cancer Research Center de Seattle.

Son laboratoire avait en effet créé une ressource unique pour l'étude de la variabilité de la relation génotype→phénotype dans le contexte de la réponse au stress hyper-osmotique. Il s'agit d'une bibliothèque de 50 souches hybrides de levures *S. cerevisiae* obtenues par le croisement de levures sauvages issues de 50 habitats différents avec une souche de laboratoire modifiée pour signaler l'activation d'un gène en bout de la voie de signalisation au stress hyper-osmotiques. Ce système de stress a été sélectionné car c'est une voie important pour la survie des cellules qui est facilement manipulable en laboratoire pour des mesures précises de réponse. Les 50 constructions de la bibliothèque représentent ainsi une collection d'une variété de fonds génétiques différents, dont le phénotype de réponse transcriptionnelle au stress hyper-osmotique peut-être mesuré quantitativement. Ces travaux, encore en cours de conclusion, ont permis de mettre en évidence des variations continues entre différents fonds génétiques de l'adaptation au stress et de la survie cellulaire, ils font l'objet de la section 5.6 du chapitre 5.

Pour des niveaux de stress faibles à modérés, les différentes constructions sont capables de s'adapter et de récupérer pleinement la viabilité suivant une dynamique de réponse propre et héritable. Pour des niveaux de stress extrêmes, chaque construction possède une distribution héritable de comportements possibles des cellules. Les cellules des populations iso-génétiques de chaque souche développent une diversification de leur taux de réponse au stress osmotique. Des cellules « lentes » survivent au stress sans s'adapter, d'autres « rapides », essaient de reprendre trop tôt durant le stress leur division cellulaire et meurent. La fraction des cellules de chaque type est propre et héritable à chaque souche.

On peut rendre compte qualitativement de ces observations de la variabilité de la survie suivant différents changements d'environnement osmotiques dans un modèle simple paramétré par la fraction de cellules lentes et rapides au sein d'une souche. La variation héritable entre des souches différentes de la probabilité de générer des cellules lentes est ainsi un moyen de diversification de stratégies de survie de type "prise de risque". Notre étude mets en évidence que cette probabilité peut être ajustée continument par différents schémas de stress environnementaux.

1.1.2 Neurosciences et neurophysique

La thématique des réseaux biochimiques, leur structure et dynamique vue sous l'angle de la dynamique des systèmes, de la physique statistique et de la simulation a été l'axe principal de mes activités de recherche en biologie théorique durant la période 2000-2012. La biologie des systèmes n'est pas le seul domaine de la biologie dans lequel le concept de réseau complexe intervient. D'autres exemples auxquels on peut naturellement penser sont tous les réseaux de transport de

fluides dans les organismes vivants, système veineux, réseau lymphatique chez les animaux, réseau phloémien chez les plantes. Exemples de processus et d'organisation biophysique étudiés par plusieurs collègues du laboratoire MSC (Marc Durant, Adrien Daerr, Julien Derr, Stéphane Douady, Vincent Fleury parmi d'autres.) Enfin, peut-être en réalité le contexte le plus ancien se référant aux réseaux complexes sont les neurosciences. Le thème des réseaux complexes, leur structure, évolution et dynamiques m'a conduit à également m'intéresser à ce domaine très actif, et à revenir sur certaines questions de comportements collectifs synchronisés et critiques qui m'avaient occupés au tout début de mon parcours de recherche en doctorat.

Comprendre l'organisation d'un cerveau et son fonctionnement, les principes de l'apprentissage, de la mémoire et du traitement de l'information sont l'un des enjeux les plus passionnants et difficiles de la science contemporaine. Bien qu'ayant une histoire assez différente de la biologie moléculaire et la génétique, les neurosciences modernes sont aussi de nos jours une entreprise hautement interdisciplinaire tirant partie des avancées informatiques et de progrès instrumentaux propres. Les recherches de neurobiologie couvrent toutes les échelles de la complexité biologique, depuis les facteurs génétiques des processus cellulaires plus faciles à caractériser grâce à la génomique, jusqu'à l'activité de populations neuronales et de circuits possibles à observer à des échelles différentes et qu'on cherche à relier aux comportements des animaux et à la cognition.

L'activité cérébrale a toujours été un sujet d'intérêt dans le champ large d'investigation des sciences de la complexité. Ce domaine qui regroupe des concepts des systèmes dynamiques mathématiques, de la théorie du cas et de la physique statistique se rapporte naturellement à l'organisation et l'activité d'un système neuronal. Cellules nerveuses et tissus sont le siège de dynamiques non linéaires hors d'équilibre dépendant et affectant leur propre structure. Malgré les progrès pour observer à des résolutions et échelles inédites l'organisation et la dynamique cérébrale, il n'existe pas de théorie complète qui peut aider à interpréter et guider les neurosciences et ses applications médicales au sujet du rapport entre la structure du cerveau (anatomie) et son comportement (psychologie), et encore moins ses qualités cognitives (psychologie). Un ensemble de questions dans ce cadre se rapporte par exemple aux relations entre chaos et propriétés cérébrales.

Une approche particulière de ces questions les aborde par le biais de la fabrication de réseaux artificiels de vrais neurones (biologiques et vivants). Il s'agit là de cultures cellulaires de neurones qui permettent des études et des expérimentations impossibles avec les tissus biologiques. Au sein du groupe théorie du laboratoire, Pascal Monceau, Stéphane Mérens et moi-même, avec la contribution inestimable de Renaud Renault en thèse, nous sommes récemment engagés dans cette démarche sous l'angle théorique en lien avec des groupes expérimentaux.

En raison de l'accessibilité de cellule et leur flexibilité, les cultures neuronales sont des dispositifs très utiles pour de nombreuses applications de recherche sur des questions fondamentales neurobiologiques, appliquées sur la physiologie et les troubles neuronaux et également vers de nouveaux usages biotechnologiques.

La connectivité des cultures peut être modifiée de différentes manières : impressions de formes sur une surface déterminant les zones de croissance possible, ou moulage de chambres et canaux confinant les cultures. La construction de géométries de culture permet de concevoir des modèles *in vitro* de structures neuronales complexes et d'étudier les interactions entre activité et connectivité. Associées à ces techniques de micro-fabrication les cultures sont ainsi des plateformes d'expérimentation pour la dynamique et les propriétés de tissus neuronaux, permettant l'étude de questions telles que l'émergence de schémas particuliers d'activité comme la synchronisation. Enfin l'observation en parallèle de l'activité du développement morphologique des cultures aborde des questions d'auto-organisation et de comportements collectifs qui intéressent aussi la physique. Le chapitre 6 de ce mémoire développera davantage le cadre des approches physiques aux cultures neuronales et l'utilisation de ces dispositifs. Mes travaux s'y rapportant sont quant à eux discutés dans le chapitre 7.

1.2 Curriculum Vitae

Samuel Bottani, né le 21 juin 1968

Nationalité Suisse, marié 2 enfants

Carrière

Maître de Conférences H.C. à l'Université Denis Diderot – Paris 7, UFR de Physique,
Recrutement 1997; depuis 2004 au Laboratoire Matières et Systèmes Complexes UMR 7057

- Sept 2013– : **Directeur du Centre de Formation des Doctorants aux initiatives professionnelles** (CFDip) de la COMUE Université Sorbonne Paris Cité (USPC), (département de l'Université Paris 7); directeur adjoint du Collège des Études Doctorales USPC.
- 2008–13 Co-directeur, directeur des Études, École Doctorale ED 474 « Européenne Interdisciplinaire Frontières du Vivant », Universités Paris Descartes et Paris Diderot, Programme Doctoral Liliane Bettencourt. Évaluée A+, AERES fév. 2013.
- 2001–03 Disponibilité pour une création d'entreprise : directeur Biomathématiques de BIOGENIX SA, startup Franco-Italienne dans le domaine de la génétique des populations (Sardaigne) et la génomique fonctionnelle.
- 1997–2001, Groupe de Modélisation Physique Interfaces Biologie, Université Paris 7.
- 1996–97, Chercheur post-doctoral, Institut de Physique Nucléaire d'Orsay.
- 1995–96, ATER, Université Paris 7,
- 1992–95, Doctorant, Laboratoire de Physique Théorique et des Hautes Énergies des Universités Paris 6 et Paris 7.
- 1994–95, Moniteur : Palais de la Découverte (1992-94), Université Paris 7 (1994-95).
- 1988–89 Assistant auxiliaire, École Polytechnique Fédérale de Zürich (ETHZ).

1.2.1 Formation

- 1992–96 : Doctorat en physique théorique de l'Université Paris 6; (très honorable; félicitations du jury); spécialité physique statistique; direction Bertrand Delamotte (CNRS).
- 1989–92 : Magistère Interuniversitaire de Physique, Université Paris 6, Université Paris 11, École Normale Supérieure Paris. DEA Champs Particules Matières.
- 1986–89 : Études propédeutiques de Physiques à l'École Polytechnique Fédérale de Zürich.

Formation continue

(5 dernières années)

- « Management, autorité et professionnalisme », ANGRAM, (3 jours, Oct 2014)
- « Encadrement doctoral : De la connaissance du doctorant et de l'analyse du processus de thèse à l'appropriation des outils d'accompagnement », Philippe Belpaeme-Sud Performance, ED145 Orsay/INRA. (stage de 6 journées, 2011)

1.2.2 Publications

1. S. Bottani, P. De Bernardis, F. Melchiorri, « On the origin of the dipole anisotropy as determined by quadrupole measurements », *The Astrophys. J.*, **384**, L1, (1992).
2. S. Bottani, P. De Bernardis, M. de Petris, S. Masi, B. Melchiorri, F. Melchiorri, P. Tanzilli, « Cosmic Background anisotropies in the millimetric region » dans *The Infrared and Submillimetre Sky after COBE*, M. Signore and C; Dupraz ed., Kluwer, 345-384, (1992).
3. S. Bottani, « Pulse Coupled Relaxation Oscillators : from Biological Synchronization to Self-Organized Criticality », *Phys. Rev. Lett.*, **74**, 4189, (1995).
4. S. Bottani, « Merveilleux Neutrinos », *Revue du Palais de la Découverte*, 23, 41-63, (1995).
5. S. Bottani, « Synchronization of Integrate and Fire Oscillator with global Coupling », *Phys. Rev. E.*, 54 :2334; (1996).
6. S. Bottani, B. Delamotte, « Self-Organized Criticality and synchronization in pulse coupled relaxation oscillator systems : the Olami, Feder and Christensen model and the Feder and Feder model », *Physica D*, 103 : 430-441, (1997).
7. S. Bottani, H. Rebscher, R. Siebert, L. Valentin, « Post-capillary detection without sheath-flow », *J. of Chromatogr. A*, 844 :433-438, (1999).
8. Z. Djouadi, S. Bottani, M.A. Duval, R. Siebert, L. Valentin, « Mobility and activation energy in ssDNA in denaturing crosslinked polyacrylamide ge; electrophoresis », *J. of Chromatography A*, 849 :231-235cd,(2000).
9. Z. Djouadi, S. Bottani, M.A. Duval, R. Siebert, L. Valentin, « Dynamical Study of ssDNA migration in denaturing polyacrylamide slab gel electrophoresis », *Electrophoresis*. 2001 Oct; 22(16) :3527-32, (2001).
10. B. Cazelles, S. Bottani, L. Stone, « Unexpected coherence and conservation » *Proceedings of the Royal Society London B*, 268 :2595-2602, (2001).
11. N. Guelzim, S. Bottani, P. Bourguin, F. Képès, « Topological and causal structure of the yeast transcriptional regulatory network », *Nature Genetics*, **31**(1), 60-63,(2002).
12. B.E. Perrin, L. Ralaivola, A. Mazurie, S. Bottani, J. Mallet, F. d'Alche Buc, « Gene networks inference using dynamic Bayesian network », *Bioinformatics*, **19**, II138-II138, (2003)
13. Ralaivola L, Perrin BE, d'Alché-Buc F, Mazurie A, Bottani S, « Inférence de réseaux d'interactions génique à l'aide de réseaux bayésiens dynamiques », In *Journées Post-Génomiques de la Doua*, La Doua, France, (2003).
14. Ralaivola L, Perrin BE, d'Alché-Buc F, Mazurie A, Bottani S, « Réseaux bayésiens dynamiques pour l'inférence de structure de réseaux de régulation », dans 'Informatique pour l'analyse du transcriptome', Boulicaut JF and Gandrillon O editors, Lavoisier, pp.291-319 (2004)
15. F. d'Alche Buc, P.-J Lahaye, T. Vujasinovic, S. Bottani, A. Mazurie, « A recurrent artificial neural network based on inertia principle for modeling gene regulatory network », chapitre dans *Bioinformatics using Computational Intelligence Paradigms*, Ed. U. Seiffert, World Scientific Publishing, (2004).
16. A. Mazurie, S. Bottani, M. Vergassola, « An evolutionary and functional assessment of regulatory network motifs in yeast », *Genome Biol.* ;6(4) :R35. (2005).
17. Bottani S, Grammaticos B. « Analysis of a minimal model for p53 oscillations » *J Theor Biol.* 2007 Nov 21;249(2) :235-45. 2007 Jul 26
18. S. Bottani, A. Mazurie, « Genetic Networks : between theory and experimentation », chapitre dans "Complex Population Dynamics : Nonlinear Modeling in Ecology, Epidemiology and Genetics" L. Stone and B. Blasius ed., *Lecture Notes in Complex Systems*, World Scientific Press, pp. 215-236 (2007).

19. R. Veitia, S. Bottani, J. Birchler « Cellular reactions to gene dosage imbalance : genomic, transcriptomic and proteomic effects », Trends in Genetics, Vol. 24, No. 8. (August 2008), pp. 390-397.
20. Bottani S, Grammaticos B. «A simple model of genetic oscillations through regulated degradation», Chaos, Solitons & Fractals, Volume 38, Issue 5, December 2008, Pages 1468-1482
21. Veitia RA, Bottani S., Whole genome duplications and a 'function' for junk DNA? Facts and hypotheses., PLoS One. Dec 14;4(12) :e8201, (2009)
22. Uhlendorf J, Bottani S, Fages F, Hersen P, Batt G. Towards real-time control of gene expression : controlling the hog signaling cascade. Pac Symp Biocomput. 338-49.(2011)
23. A. Miermont, S. Bottani, P. Hersen, Signaling cascade dynamics after a hyper-osmotic shock in the yeast *Saccharomyces cerevisiae*. European Biophysics Journal with Biophysics Letters 40, 213 (Aug, 2011).
24. S. Bottani, B Grammaticos , Stochastic and deterministic simulations of a delayed genetic oscillation model : Investigating the validity of reductions, Chaos, Solitons & Fractals "Chaos and dynamics in biological networks" Volume 45, Issue 5,, 588-602. (2012)
25. J. Uhlendorf, A. Miermont, T. Delaveau, G. Charvin, F. Fages, S. Bottani, G. Batt, P. Hersen, « Long-term model predictive control of gene expression at the population and single-cell levels », Proc Natl Acad Sci U S A. 2012 Aug 28;109(35) :14271-6. (2012)
26. A. Miermont, F. Waharte, S. Hu, MN. McClean, S. Bottani, S. Léon, P. Hersen. « Severe osmotic compression triggers a slowdown of intracellular signaling, which can be explained by molecular crowding. » Proc Natl Acad Sci U S A. (2013)
27. RA. Veitia, S. Bottani S, JA Birchler., « Gene dosage effects : nonlinearities, genetic interactions, and dosage compensation. », Trends Genet. (2013)
28. Renault, R., Monceau, P. Bottani, S., « Memory decay and loss of criticality in quorum percolation. » Physical review. E, Statistical, nonlinear, and soft matter physics, 88(6), p.062134. (2014)
29. R. Renault, P. Monceau, S. Bottani, S. Métens, « Effective non-universality of the quorum percolation model on directed graphs with Gaussian in-degree » Physica A (2014)
30. R. Renault ;, S. Descroix, L. Malaquin, JLViovy, JM Peyrin, S. Bottani, P. Monceau, E. Moses ; M. Lucie Vignes, « Combining microfluidics, optogenetics and calcium imaging to study neuronal communication in vitro » PLoS One. Apr 22;10(4) (2015)
31. Veitia RA, Veyrunes F, Bottani S, Birchler JA., X chromosome inactivation and active X upregulation in therian mammals : facts, questions, and hypotheses. J Mol Cell Biol. Feb;7(1) :2-11 (2015)
32. Monceau P, Renault R, Métens S, Bottani S., « Effect of threshold disorder on the quorum percolation model. », Phys Rev E. Jul;94 (2016)
33. Métens S, Monceau P, Renault R, Bottani S., « Finite-size effects and dynamics of giant transition of a continuum quorum percolation model on random networks. », Phys Rev E. Mar; 93(3) : 032112 (2016)
34. Bottani S, Veitia RA., « Hill function-based models of transcriptional switches : impact of specific, nonspecific, functional and nonfunctional binding. », Biol Rev Camb Philos Soc. Apr 8 (2016)
35. Veitia RA, Govindaraju DR, Bottani S, Birchler JA. « Aging : Somatic Mutations, Epigenetic Drift and Gene Dosage Imbalance. » Trends Cell Biol. 2017 Apr;27(4) :299-310. (2016)
36. Bottani S, Zabet NR, Wendel JF, Veitia RA. « Gene Expression Dominance in Allopolyploids : Hypotheses and Models. », Trends Plant Sci. 2018 May;23(5) :393-402. (2018)
37. Fardet T, Ballandras M, Bottani S, Métens S, Monceau P. « Understanding the Generation of Network Bursts by Adaptive Oscillatory Neurons. », Front Neurosci. (2018)

38. Fardet, T., Bottani, S., Méstens, S., Monceau, P. « Effects of inhibitory neurons on the quorum percolation model and dynamical extension with the Brette–Gerstner model. » *Physica A : Statistical Mechanics and Its Applications*, 499. (2018)

Brevets

- Siebert R., Bottani S., Rebscher H., Guillemin M.T., Valentin L., Brevet n°9801007 « Electro-phorèse multicapillaire avec détection à faible bruit par le système OXID », 1998.
- Bottani S. Siebert R., Rebscher H., Cohen-Solal M., Valentin L., Brevet n°9803100 « Perfectionnements aux dispositifs d'électrophorèse multicapillaires du type à détection en sortie des capillaires », 1998.

Autres

- S. Bottani, B. Delamotte, G. Giavarini « Equivalence between the δ -expansion and the weak coupling expansion », Rapport interne LPTHE, (1993).
- S. Bottani, « Merveilleux Neutrinos », *Revue du Palais de la Découverte*, 23, 41-63, (1995).
- S. Bottani, « Des détecteurs pour l'analyse des gènes à l'IPN », *La lettre de l'IPN*, n.122, avril 1999.
- S. Bottani, « L'apport de l'IPN dans le séquençage de l'ADN », *IPN Sciences* n. 2, janvier 2001.

1.2.3 Invitations et séminaires :

(5 dernières années)

2-9 janvier 2011; 27 avril- 5 mai 2013; 22 Février-1 mars 2014 : invitations à l'Institut des Systèmes Complexes, Institut Weizmann, Rehovot, Israël par le Professeur Elisha Moses. Mai 2014 : École thématique Chaos & dynamique dans les réseaux biologiques, Cargèse. 4 nov 2015 : conférence dans le cadre des « Journées européennes en SHS : l'interdisciplinarité ». 15 mars 2016, « Colloque Office de Coopération et d'Information Muséale » intervention invitée et synthèse de la journée. 9 juin 2017 : université d'Orsay séminaire au laboratoire IMNC.

1.2.4 Enseignements

Licence

- 2014-15** Responsabilité cours Physique pour Biologistes L1, UFR Sciences de la Vie, Université Paris 7. Renouveau du module avec mise en œuvre d'outils TICE et de nouvelles pratiques pédagogiques. (64h)
- 2014-15** Responsabilité du module Projets Professionnels Personnels, L1 parcours Physique du Département Sciences Exactes, Université Paris 7. Renouveau du module avec mise en œuvre d'approches pédagogiques participatives. (15h)
- 2012-13-14** TD Physique pour biologistes, L1 Sciences de la Vie (14h, Prof Patrice Flaud)
- 2011-12** Projets Professionnels Personnels, L1 parcours Physique du Département Sciences Exactes, Université Paris 7. (14h)
- 2008-13** Projet de Physique Expérimentale, L3 UFR de physique Université Paris 7. Mise en œuvre de nouvelles approches pour l'apprentissage d'outils de communication numériques en ligne, déploiement et gestion d'un wiki du module pour la formation et les travaux étudiants. (96h)
- 2008-10** Informatique pour Physiciens, L2 Physique, cours de programmation Python pour physiciens.

- 2005-06** Projets Professionnels Personnels, L1 parcours Physique du Département Sciences Exactes, Université Paris 7.
- 2003-06** Physique Numérique et programmation en C, Licence F3 (L3) Physique; responsabilité du module (1/3 de service).
- 2003-04** Informatique pour Physiciens (programmation en C), DEUG 2e année (2/3 de service).
- 2000-01** Projet Documentaire par Internet. Mise en place de ce nouveau module de Licence destiné à la formation des étudiants à la recherche documentaire scientifique et à un usage professionnel d'internet.
- 1999-01 et 2000-01** : Projets Professionnels Personnels. Participé à la création de ce module dans le Deug 1re année Physique à Paris 7, visant à apporter une aide aux étudiants dans leur approche de l'enseignement supérieur et de leur orientation professionnelle.
- 2000-01** Méthode numériques, Licence F3. Travaux dirigés sur ordinateur : initiation à Unix, cours de C, méthodes d'analyse de signaux, transformée de Fourier, application à la propagation d'ondes dans des chaînes d'oscillateurs. Deug 3e année, Travaux pratiques (Prof Jacques Lebourlot)
- 1997-98-99** Projets de Physique Expérimentale du Deug 1ère année Physique (1997 : 1/3 de service; 1998 : 2/3 de service et coordination avec Jean-Marc Esteva)
- 1997-98-99** Travaux dirigés de physique Deug 1re année (Prof Bernard Diu, 1997 : 2/3 de service, 1998 : 1/3 de service)

Master

- 2010-14** Enseignement au stage de rentrée « bootcamp » du M1 Approches Interdisciplinaires du Vivant, cours de modélisation biochimique, thermodynamique et cinétique chimique. Enseignement en Anglais. (15h)
- 2010-12** Enseignement dans le cours Biologie Computationnelle du M1 Approches Interdisciplinaires du Vivant (15h)
- 2010-11-12-13** 1/2 journée de cours sur la biologie synthétique et computationnelle à AgroParis-Tech, module Connaissance et Maîtrise du Vivant (invitation par Prof Agnès Ricoch).
- 2009-13** Responsabilité cours Physique Numérique, M1 Physique de la Matière et Applications, UFR de Physique, Université Paris 7. Enseignement par projet. Mise en œuvre d'une formation hybride en ligne et présentielle. (64h)
- 2008-10** Responsable du cours Biologie Synthétique, M1 Approches Interdisciplinaires du Vivant, Faculté de Médecine Université Paris 5 et UFR Sciences de la Vie Université Paris 7. Enseignement en Anglais. Création du cours. (36h)
- 2003-06** Thermodynamique et Phénomènes de Transport, M1 Physique, Travaux dirigés (cours Prof Noëlle Pottier)
- 2000-01-02** Instrumentation Biophysique, DESS Ingénierie Génomique Université Paris Diderot, Université d'Evry. 1) Imagerie fonctionnelle : mesures d'hybridation, puces à ADN, séquenceurs automatiques. Principes de détection : radioactivité, fluorescence, détecteurs, exemples d'imageurs. Spectrométrie de masse. 2) Perspectives instrumentales : innovations sur les puces, études sur molécules isolées, séquençage non électrophorétique. (12h création du cours pour ce nouveau DESS)

Doctorat (à l'ED 474)

- 2012-13** Critical Analysis of Research Articles; cours pour les doctorants de 1re année de l'ED 474 sur la lecture critique de publications de recherche, basé sur l'analyse et la discussion d'exemples.

2008-09-10-11-12 Workshop Creating Interdisciplinary Research Projects, séminaire d'intégration présentiel annuel des nouveaux doctorants de l'ED 474 portant sur la conception, l'organisation et de projets de recherches interdisciplinaires. Co-encadrement du séminaire avec Ariel Lindner, et animation de l'équipe des chercheurs intervenants.

2008-09-10-11-12 Communication en Séminaires scientifiques, formation à l'occasion des séminaires hebdomadaires des doctorants de l'ED474, préparation pédagogique et commentaires individuels sur la prestation de présentation.

1.2.5 Ingénierie de formation

2008-14 École Doctorale ED 474 « Européenne Interdisciplinaire Frontières du Vivant ». En tant que directeur des études : lancement à partir de septembre 2008 des activités pédagogiques et du programme spécifique de l'école doctorale créée en 2007 (<http://www.fdv-paris.org>).

2009 Organisation du module « Gestion des réseaux électriques » à la création du Master Pro Ingénierie Physique des Energies, Université Paris Diderot; recherche d'intervenants et construction du programme de cours (de 40h) avec les partenaires industriels (Poweo).

2007-08 Création Master 1 AIV « Approches Interdisciplinaires du Vivant » Faculté de Médecine Paris Descartes et UFR Sciences du Vivant Paris Diderot (Comité de projet pour Paris Diderot avec Pascal Hersen; porteur Miroslav Radman, Paris Descartes). Première création en France d'un master de Biologie des Systèmes et Biologie Synthétique.

2003-05 Création de licences professionnelles en alternance (porteur de projet pour l'UFR de Physique Paris Diderot avec Giuseppe Leo). « Analyse des Matériaux », avec l'école Ecole Technique Supérieure du Laboratoire Paris 13e et le CFA AFORP . « Biophotonique » avec l'Ecole National de Chimie et Physique de Paris (13e) et le CFA ID2F

2000-01 Participation à la création du DESS d'Ingénierie Génomique Fonctionnelle, U. Paris 7 et U. d'Evry (Porteur Prof. Marc Fellous).

1.2.6 Encadrement

Encadrements de stages

2017 Janvier-Juillet, Alessio Quaresima, M2 Université de Rome La Sapienza, séjour Erasmus+, « Développement du simulateur de croissance neuronale en culture ».

2016 Janvier-Mai, Mathieu Ballandras, M2 Physique Macroscopique et Complexité UPMC, « Periodic synchronized bursting in neuronal assemblies » (co-encadrant Pascal Monceau 33%, Stéphane Mérens 33%).

2014 , Janvier-Mai, Violaine Lalande, M2 Physics of complex Systems, Université Paris Diderot, « Simulations de la dynamiques de cultures de neurones avec le modèle neuronal d'Izikevitch », encadrement 50% (co-encadrant Pascal Monceau 50%)

2014 Janvier-Avril, Michele Agamennone, M2 Physics of complex Systems, Université Paris Diderot, « Modeling, simulation and study of two-dimensional dissociated neuronal cultures : growth process, topology, activity », encadrement 60% (co-encadrant Pascal Monceau 40%)

2013 Octobre-Décembre, David Davila, M2 Approches Interdisciplinaires du Vivant, « Overview of available Neuron Simulation Packages and their Application in Modeling Neuronal Cultures », encadrement 100%

2013 Juin-Juillet, Mordjane Boukhet, M1 Physique Orsay, « Evolution des amas dans un réseau de culture de neurones in vitro dans le modèle de percolation à quorum », encadrement 40% (co-encadrant Pascal Monceau 60%)

- 2013** Juin-Juillet, Nicolas Fossey, M1 Physique de la Matière et Applications, Université Paris Diderot, « Modélisation du comportement collectif d'un réseau de neurones », encadrement 60% (co-encadrant Pascal Monceau 40%)
- 2010** Octobre-Décembre, Renaud Renault, M2 Approches Interdisciplinaires du Vivant Universités Paris Descartes et Diderot, « La percolation à quorum », encadrement 100%.
- 2012** Juin-Juillet, Nazih Benoumichiara, M1 Physique de la Matière et Applications, Université Paris Diderot, « Modélisation de systèmes dynamiques et mise en œuvre de stratégies d'optimisation de paramètres en application à des réseaux biologiques ».
- 2011** Juin-Juillet, Adrien Berthault, L3 PMA « Comparaison d'algorithmes d'apprentissage de paramètres appliqués à des réseaux génétiques ».
- 2002** avril-décembre, Bruno Edouard Perrin, stage pour le diplôme d'Ingénierie informatique du « Politecnico di Milano » dans le cadre du programme européen TIME,; Titre du mémoire « Une expérience de modélisation de réseau génétique utilisant une approche probabiliste », co-encadré avec F. d'Alché-Buc (LIP6) à 50%.
- 1997** janvier-mars, Yves Hougardy, stage de maîtrise Université d'Evry, « Optimisation de l'excitation laser d'ADN marqué par fluorescence dans une cellule d'électrophorèse capillaire ».
- 1999** sept - 2000 mars, Yuxing Chen, stagiaire visiteur venant de l'Université de Harbin en Chine. Mise au point de méthodes et de programmes pour le traitement de données expérimentales d'électrophorèse.
- 1999** juin-septembre, Sébastien Douet, stage d'école d'ingénieur (ENSEEIH), « Développement de programmes de mise en forme des données expérimentales pour l'analyse avec le logiciel d'assignation de bases Phred; paramétrage et mise en œuvre des analyses. »
- 1998** juin-septembre, Sandrine Herriot, stagiaire en licence de Physique (Orsay), « Caractérisation de la caméra CCD et mise au point des programmes d'acquisition des données sur » les images.
- 1997** mai-août, Philippe Callulier, stage d'IUT informatique (Orsay). « Adaptation de logiciel Bass d'assignation de bases pour le séquençage d'ADN aux données du séquenceur en développement au laboratoire ».

Encadrement de thèses

- 1997-00** Zahia Djouadi, U. Orsay, Etude de l'élargissement de bandes d'ADNss durant l'électrophorèse en gel de polyacrylamide. (direction Marie Alix Duval, Institut de Physique nucléaire d'Orsay, co-encadrement effectif 50%).
- 2001-04** Aurélien Mazurie, U. Paris 6, Des gènes aux réseaux génétiques. Exploitation des données transcriptomiques, inférence et caractérisation de structures de régulation. (Directeur de thèse, Jacques Mallet, Institut de Génétique des Neurotransmetteurs, La Pitié-Salpêtrière; co-encadrement effectif 50%).
- 2009-12** Agnès Miermont, U. Paris 7, « Analyze and measurement of HOG Signaling cascade dynamics in the yeast *Saccharomyces cerevisiae* ». (Directeur de thèse Pascal Hersen; laboratoire Matières et Systèmes Complexes, co-encadrement effectif 33%).
- 2012-15** Renaud Renault ED 518, « Conception Émergente de Processeurs Neuronaux », co-direction 50% (co-encadrant Pascal Monceau 50%).

Thèses en cours : Tanguy Fardet, « Modélisation de la croissance neuronale et dynamique en culture », (Directeur de thèse Pascal Monceau, co-encadrement 50%)

Devenir des docteurs :

- Zahia Djouadi : Maître de Conférences en Physique, Université d'Orsay
- Aurélien Mazurie : Directeur « Bioinformatics Core » à la Montana State University.
- Agnès Miermont : Postdoctoral Research Associate à University College Lond
- Renaud Renault : « Chief of unlisted projects architect » chez Koniku, startup de neuroingénierie à Barkeley, Californie

Encadrement administratif

Dans le cadre de mes responsabilités administratives j'ai assuré l'encadrement managérial de personnels affectés aux services concernés.

- 2008-2013 dans le cadre de l'École Doctorale Frontières du Vivant, 2 personnels, gestionnaire de l'école et responsable administrative.
- 2013- dans le cadre du Centre de Formation des Doctorant.e.s aux Initiatives Professionnelles, 3 gestionnaires de formations et du service, 1 responsable administrative.

1.2.7 Animation scientifique

Au sein du groupe théorie du laboratoire matières et systèmes complexes. Membres de l'équipe : Pascal Monceau (MFC Univ. Evry), Stéphane Métens (MCF Univ. Paris Diderot), Tanguy Fardet (doctorant) : j'ai été le noyau pour la constitution de cette équipe au sein du laboratoire MSC suite à mes travaux dans le domaines initiés à partir de 2011 en faisant la connaissance du Prof. Elisha Moses, Inst; Weizmann Israël (WIS). Nous avons depuis 3 publications, le recrutement de deux doctorants, 4 séjours au WIS; une année de séjour de notre doctorant au WIS, la mise en place de collaborations avec Jean-Louis Viovy et Catherine Villard (Insitut Pierre Gilles de Gennes, Paris) et Jordi Soriano (Université de Barcelone).

1.2.8 Organisation de colloques et rencontres scientifiques

- 2015 - 2016 Coordinateur du groupe de travail inter-Comue d'Île-de-France pour l'organisation d'un salon de l'emploi des docteurs. Action qui a conduit au salon « PhD Talent Career Fair » le 20 octobre 2015 avec 1600 visiteurs et 70 exposants. Mise en place d'un soutien institutionnel et régional pour les éditions successives (100 exposants, plus de 2000 doctorants et docteurs visiteurs en 2017.)
- 2013- 2014 : membre du comité d'organisation pour la première organisation en France du concours de communication scientifique pour jeunes chercheurs « FameLab », en partenariat avec le British Council et Amcsti, finale régionale Ile de France (17 mars) et finale nationale organisées à l'Université Paris Diderot (30 avril).
- « Paris Interdisciplinary PhD Symposium : Shapes in Living Systems », 13-15 juin 2011, ESPCI, supervision du symposium par l'ED 474.
- Workshop « Constructing gene networks : observation, analysis and control », 15 Janvier 2009, organisation avec G. Batt, Université Paris Diderot
- Journée scientifique « Biologie Synthétique », 12 mars 2008, avec D. Mazel, Institut Pasteur
- Workshop « iGEM 2008 European Teachers », 31 mai 2008, avec Meagan Lizarazo, Centre de Recherches Interdisciplinaires Paris Descartes
- Workshop « iGEM in Europe », 29 septembre 2007, avec Randy Rettberg, workshop, Université Paris Diderot

- « Ingénierie du vivant 2.0 La Biologie Synthétique en Question », Cycle 2009 de Débats Publics co-organisés par VivAgora, La Cité des sciences et de l'industrie, L'Institut de recherche et d'innovation (IRI) du Centre Pompidou; membre du comité de pilotage
- Dans le cadre de ma fonction de directeur des études de l'école doctorale Frontières du Vivant, supervision de l'organisation des rencontres doctorales internationales « Paris Interdisciplinary PhD Symposium » par les doctorants de l'école, attirant près de 200 doctorants internationaux : « Shapes in Living Systems », 13-15 juin 2011, à l'ESPCI « From sparse entities to crowded environments : numbers in living systems », 7-9 décembre 2009, à l'Université Paris Diderot

1.2.9 Responsabilités et implication dans la vie collective de l'établissement

1.2.10 Direction des études ED 474

« qBio is the power of math-based reasoning and advanced instrumentation from physics and engineering harnessed to discover fundamental principles of living systems. To make biology quantitative and predictive, it is necessary to draw upon a multitude of approaches from the physical sciences and engineering. These include theoretical concepts developed from studies in statistical mechanics and nonlinear dynamics, and experimental methods such as microfluidics and advanced imaging. Therefore the goal of the qBio graduate program is to provide the students with a mastery of both the theoretical knowledge and experimental skills, and guide them to employ both approaches to address fundamental biological problems during their thesis research. »

UC San Diego qBio Ph.D Specialization [QBIO \[2017\]](#)

En raison de ma recherche à la frontière physique/ biologie et de mon implication pour l'enseignement innovant j'ai été nommé en 2008 directeur adjoint et directeur des études d'une toute nouvelle école doctorale dérogatoire à thématique interdisciplinaire, l'ED 474 « Européenne Interdisciplinaire Frontières du Vivant » issue d'un partenariat public/privé Universités Paris Descartes, Paris Diderot, ENS et Fondation Bettencourt-Schueller. Cette mission a constitué une opportunité unique pour mettre en place un programme doctoral international d'un type nouveau centré sur l'interdisciplinarité et la créativité. Hébergée au Centre de Recherches Interdisciplinaires (CRI) l'ED 474 offre à ses doctorants un esprit d'école fort et un cadre international de formation avancé qui a été distingué en 2013 par l'AERES avec les plus hautes évaluations. J'ai assumé la direction opérationnelle de l'école jusqu'en décembre 2014.

Juin 2008- Dec. 2013 Directeur adjoint et Directeur des Études de l'École Doctorale ED 474 « Européenne Interdisciplinaire Frontières du Vivant », Universités Paris Descartes et Paris Diderot. ED évaluée A+, AERES fev 2013. (Directeur François Taddei, Paris Descartes). École doctorale créée en 2007 bénéficiant par décision ministérielle d'un statut dérogatoire sans équipes d'accueil permanentes, pour un soutien actif de projets de thèse interdisciplinaires autour du vivant (directeur François Taddei, Paris Descartes); effectif 100 doctorants en 2013. Mes activités ont consisté dans la direction opérationnelle de l'école doctorale et le suivi individuel des doctorants : création de toutes les procédures (recrutement, collège des experts internationaux, comités de thèse, évaluations, suivi), organisation de la semaine de formation initiale des nouveaux doctorants à la création de projet de recherche, organisation du programme académique avec choix d'une quinzaine de cours et de groupes d'études, coordination des séminaires hebdomadaires par les doctorants avec retour pédagogiques sur leur communication, suivi des doctorants, aménagement des nouveaux locaux, création et animation du site web pédagogique.

1.2.11 Direction Centre de Formation des Doctorant.e.s à l'insertion professionnelle Université Sorbonne Paris Cité

Construisant sur mon expérience de montage de formations doctorales et de travail interdisciplinaire j'ai été nommé en septembre 2014 directeur du Centre de Formation des Doctorants à l'insertion professionnelle (CFDip) de la Comue Université Sorbonne Paris Cité avec pour mission de développer en partenariat avec les 32 écoles doctorales de la Comue les formations à vocation professionnelle et appliquées proposées aux doctorants USPC et préparant à des poursuites de carrières vers tout secteurs académiques ou autres.

Octobre 2013 – Directeur de département : Centre de Formation des Doctorants à l'insertion professionnelle (CFDip) de la Comue Université Sorbonne Paris Cité (USPC) <http://cfdip.uspc.fr/> visant la promotion et le développement des compétences professionnelles des doctorants et la préparation de leur poursuite de carrière. En partenariat avec les 32 écoles doctorales de la Comue, mutualisation et développement des formations méthodologiques et professionnelles proposées aux doctorants des établissements USPC (soit plus de 6000 doctorants, tous statuts). En 2014 ouverture de 180 formations, suivies par près d'un tiers des doctorants USPC.

Le CFDip est un département de l'Université Paris Diderot assurant une mission de service commun pour la Comue USPC; il a pour attribution aussi la coordination du recrutement des doctorants contractuels avec mission d'enseignement pour l'Université Paris Diderot. En qualité de directeur du centre de formation je suis directeur adjoint du Collège des Écoles Doctorales (CED) de USPC et invité permanent au bureau des représentants des établissements du collège.

1.2.12 Autres responsabilités et implications la vie collective de l'établissement :

Par ailleurs, dans la logique de mon engagement sur les approches innovantes d'enseignement je participe l'automne 2013 au comité exécutif du Service d'Accompagnement aux Pédagogies Innovantes et à l'Enseignement Numérique de Sorbonne Paris Cité (SAPIENS) dont l'objectif est la création de formations et de ressources pédagogiques spécifiques et transversales pour les enseignants universitaires, les responsables de formation, les ingénieurs et conseillers pédagogiques des établissements de la Comue, ainsi que la mise en œuvre d'appels à projets pour des enseignements innovants. En lien avec cette action j'ai été membre du groupe de travail Diderot Usages Numériques (DUNE) missionné par la présidence de l'Université Paris Diderot pour renforcer les pratiques TICE locales.

- 2018 - Membre élu à la Commission Recherche de l'Université Paris Diderot
- 2014-2016 (Mars) - Mission pour la présidence de l'Université Paris Diderot sur les usages pédagogiques numériques (Membre du groupe de travail Diderot Usages Numériques :DUNE)
- 2014-2017 Membre du comité de réflexion du projet européen TRIGGER (TRansforming Institutions by Gendering contents and Gaining Equality in Research) pour le Pôle Égalité Homme Femmes de l'Université Paris Diderot.
- 2013 (Sept.) - Membre du comité opérationnel du Service d'Accompagnement aux Pédagogies Innovantes et à l'Enseignement Numérique de Sorbonne Paris Cité (SAPIENS)
- 21/03/2013- Membre du comité de pilotage du PEPITE CréaJ (Pôle Etudiant pour l'Innovation, le Transfert et l'Entrepreneuriat) de la Comue SPC.
- 2003-2006 : Membre du Conseil des Enseignements, UFR de Physique Paris Diderot
- 2003-2005 Porteur de projet pour l'UFR de Physique Paris Diderot (avec Prof Giuseppe Leo) pour la création de licences professionnelles (deux créations : « Analyse des matériaux et « Biophotonique »)
- 2000-2005 : Membre de la commission de spécialistes CS29 « Constituants élémentaires » de l'Université Paris 7

1.3 Références

QBIO, U. S. D. 2017, «Discover Quantitative Biology (qBio) at UC San Diego», URL <http://qbio.ucsd.edu/>. 19

VEITIA, R. A. 2005, «Gene dosage balance : Deletions, duplications and dominance», *Trends in Genetics*, vol. 21, n° 1, doi :10.1016/j.tig.2004.11.002, p. 33–35, ISSN 01689525. 8

Chapitre 2

Théorie et Biologie

« A theory, or, more concretely, a model, of a mechanism is not a description of reality; it is a description of our assumptions about reality. Models, therefore, have to evolve with our knowledge : they are always wrong but sometimes useful. »

Jeremy Gunawardena « Biology is more theoretical than physics » [GUNAWARDENA \[2014\]](#)

Sommaire

2.1 Renouveau de la théorie et de la modélisation en biologie	27
2.1.1 Théorie et modèles	27
2.2 Développement de la modélisation mathématiques en biologie	29
2.2.1 Systèmes dynamiques mathématiques pour la biologie	32
2.2.2 Théories de la morphogénèse et auto-organisation	33
2.2.3 L'essor de la modélisation mathématique et ses limites	33
2.3 Biologie à haut débit et quantitative	35
2.3.1 La poussée technologique	35
2.4 La biologie des systèmes	37
2.4.1 Quête d'une définition de la biologie des systèmes	38
2.4.2 Pléiotropie et besoin d'intégration	38
2.4.3 Pléiotropie et nombre de gènes, vers les réseaux d'interaction	39
2.4.4 Les réseaux : paradigme scientifique et de société	40
2.4.5 Biologie et informatique : une fertilisation réciproque	40
2.5 Biotechnologies et biologie de synthèse	41
2.5.1 De la nouveauté de la biologie de synthèse	42
2.5.2 Les technologies propres à la biologie de synthèse	43
2.6 Les réseaux complexes	46
2.6.1 Neurosciences et réseaux complexes	57
2.7 Conclusion du chapitre	58
2.8 Références	58

Parmi les disciplines scientifiques d'exploration du monde naturel, la biologie est une science expérimentale, rarement qualifiée d'« exacte », de « formelle » ou « dure ». Ces adjectifs sont plus couramment associés à des disciplines comme la physique, la chimie, l'informatique et plus généralement toute discipline reposant sur un usage intensif des mathématiques. Dans ces domaines, c'est l'usage important des mathématiques qui motive leur qualification avec ces adjectifs exhaltant la précision, la rigueur et la robustesse de leur démarche scientifique. Les mathématiques y sont en effet utiles pour capturer les mécanismes naturels découverts par l'observation et l'expérience, et pour les exprimer dans un langage formel permettant d'anticiper des comportements

non encore observés. Les « modèles » sont alors des fragments de connaissances, des explications formalisées de situations et exemples particuliers, construits dans le cadre d'une théorie, c'est-à-dire le cadre conceptuel général définissant notre compréhension actuelle des principes fondamentaux du domaine. Depuis le 16^e siècle et Galilée, en physique, les théories et modèles sont sous forme mathématiques.

Les sciences de la vie semblent échapper à cette tradition. Celles-ci ne sont généralement pas reconnues comme fondées sur les mathématiques. Même si très fortement technologiques, elles se rapprochent davantage des grands questionnements et de la créativité propre des « sciences humaines », lesquelles revendiquent même parfois une absence de mathématisation¹. Les explications biologiques sont souvent exprimées sous forme de textes décrivant avec des mots et des phrases les agents en jeu, les facteurs qui influencent et les actions possibles. Les textes sont souvent accompagnés des schémas et de dessins illustrant graphiquement ces descriptions. Tout manuel de biologie est rempli d'illustrations sophistiquées et très soignées représentant notre compréhension des structures et des objets impliqués dans le processus étudié et ses interactions. Tous ces schémas dont des modèles, des représentations du savoir scientifique qui se lisent dans un cadre de référence plus général, introduit progressivement dans les ouvrages ou le long d'un cursus d'étude, qui donne la signification aux différents éléments dessinés et le contexte du mécanisme disséqué. Cet ensemble conceptuel est la théorie scientifique relative au champ d'étude, dont le modèle est l'application dans une situation particulière. On voit bien ainsi que la biologie repose elle aussi des théories et l'utilisation de modèles.

Néanmoins il est incontestable que les sciences de la vie sont communément perçues comme clairement moins théoriques et abstraites que les autres sciences de la nature. La biologie apparaît comme une science avant tout de l'observation. Il suffit d'interroger les étudiants et étudiantes de biologie et souvent aussi leurs enseignants, pour constater la prévalence de ce point de vue. La lecture des manuels scolaires concoure dans ce sens². Ceux de sciences de la vie, par ailleurs passionnants, sont une somme considérable de descriptions, de définitions, de schémas, de descriptions de situations et cas d'application. L'étude de la discipline exige une grande mémorisation, la capacité d'associations entre nombreux éléments et de déduction. Les ouvrages de physique, ceux de mathématiques, quoique de nos jours richement illustrés et enrichis d'exemples, sont moins prolifiques d'illustrations et de données. Ils définissent davantage un nombre restreint de notions et objets d'étude, quelques principes et théorèmes qui s'y appliquent suivis d'illustrations avec des applications sur des exemples divers. La mémorisation est plus réduite par rapport à la biologie. La recherche d'associations entre faits différents puis de déductions moins centrale. Bien que les nouveaux programmes du lycée survolent un champ d'applications les plus diverses, dans l'enseignement scolaire des sciences physique et chimique on s'attache encore à l'identification d'hypothèses de départ et le déroulé d'un raisonnement logique fait de calculs et application de lois naturelles pour obtenir un résultat. Pour les étudiantes (étudiants essentiellement... l'équilibre des genres est loin d'être vrai parmi les inscrites) en L1 de sciences de la vie de mes cours et TD de physique que j'interroge depuis plusieurs années, la physique est abstraite, c'est des mathématiques et des calculs appliqués. Quelques-unes se plaignent qu'elle est trop théorique. La biologie au contraire est pour elles concrète et l'apprentissage d'un grand nombre de faits et de détails est inéluctable pour composer des explications réelles.

Les mathématiques et leur place dans ces disciplines semblent être responsables de ce clivage. Dans un cas, calcul et théorie, dans l'autre observation, synthèse et déduction. Les communautés des chercheurs de biologie et de physique ont des positionnements différents envers la manière d'exprimer des savoirs. (On discutera plus bas sur les langages de modélisation utilisé pour représenter la nature, en biologie plus rarement mathématiques qu'en physique.) Cependant si la théorie semble moins présente en biologie que dans d'autres sciences naturelles c'est peut-être

1. Notons que des domaines traitant des sociétés humaines comme l'économie, ou encore certaines approches de sociologie, sont également fortement basées sur les mathématiques. Certains courants de sciences humaines développent des approches mathématiques mais font objet de vives polémiques.

2. La scolarité de mon fils cette année en terminale a été instructive!

plus pour des fondements philosophiques. Une situation pouvant provenir de différences historiques de perception de la démarche scientifique entre les praticiens des différents domaines.

Dans un commentaire récent dans *Nature* l'écrivain scientifique auteur de nombreux ouvrages de vulgarisation Philipp Ball, décrit comment la démarche scientifique contemporaine comporte encore deux tendances fortes aux racines anciennes BALL [2013] :

« Plato and Aristotle stand proxy for two types of scientists : those who see similarities, patterns and universals, and those who see differences, variation and specifics. Both viewpoints are needed, but they are hostage to disciplinary and cultural fashions. It's probably safe to say that physics stresses the former, biology the latter. »

Le premier type de démarche scientifique cherche avant tout à identifier des similarités entre objets et phénomènes naturels. Elle scrute de schémas généraux et universels. La seconde démarche de l'investigation scientifique est davantage à la recherche des différences pouvant exister au sein de systèmes, elle vise à identifier leurs variations et les traits particuliers. L'hypothèse de Philippe Ball de racines philosophiques profondes serait intéressante à approfondir, je ne m'égarerai pas davantage pour l'instant. Toujours est-il que les deux approches évoquées, sont bien présentes dans les champs de la Science. Des modes de pensée peut-être caractéristiques de stéréotypes des auteurs anciens peuvent avoir influencé de manière différente l'orientation collective de communautés différentes. Les théories scientifiques, par nature, relèvent plutôt de la première tendance et elle a prévalu dans l'inspiration en physique. L'énorme diversité du monde vivant a probablement conduit la seconde démarche à dominer davantage en biologie.

L'étude des organismes vivants, d'environnements naturels de sociétés humaines, de villes, de l'économie nécessite de scruter les traits particuliers de chaque sujet. C'est le propre des disciplines scientifiques spécialisées, détaillant les objets d'étude. Au même moment un autre regard, philosophiquement différent, peut chercher à faire apparaître les propriétés communes dans ces structures, celles génériques au-delà des traits spécifique de chaque exemple. Même des objets humains et technologiques aussi complexes que des villes se prêtent à ces vues différentes. À l'analyse circonstanciée des géographes et historiens peut s'ajouter la recherche de caractéristiques du phénomène « ville » en général. Ainsi, si chaque ville est différente, en fonction de sa géographie et de son histoire, la création et le développement des villes obéit à des principes généraux que met en évidence par exemple la loi empirique de George Kingsley Zipf, figure 2.1.

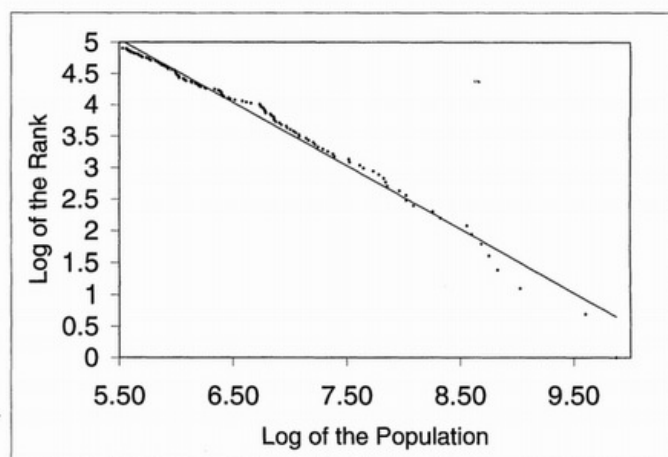


FIGURE 2.1 – **Loi de Zipf de la distribution de tailles des villes.** Au-delà de toutes les spécificités géographiques, historiques, économiques, d'entités complexes et très différentes telles que des villes, leur création et leur fonctionnement suit des propriétés globales. Pour la plupart des pays, la distribution de la taille des villes en fonction de leur population correspond à une loi de puissance : le nombre de villes dont la population est supérieure à S est proportionnel à $1/S$. Source GABAIX [1999]

FIGURE I
Log Size versus Log Rank of the 135 largest U. S. Metropolitan Areas in 1991
Source: Statistical Abstract of the United States [1993].

C'est une expérience commune de physiciens et mathématiciens s'intéressant au vivant, de voir des interlocuteurs biologistes outrés par les représentations considérées simplistes avec laquelle les premiers essaient de décrire et de modéliser des systèmes naturels. Une suspicion pro-

fonde a longuement prévalu en biologie envers la théorie. En témoigne le point de vue d'un des fondateurs de la neurobiologie moderne, Santiago Ramon y Cajal rapporté par Jeremy Gunawardena sur son site web [GUNAWARDENA](#).

Dans le chapitre 5 intitulé « Maladies de la volonté » dans un court ouvrage destiné à former et conseiller les jeunes chercheurs « Reglas y consejos sobre investigación científica » [RAMÓN Y CAJAL, 1899], Santiago Ramon y Cajal identifie, non sans humour acerbe plusieurs grandes classes de scientifiques « mentalement dérangés ». Il identifie des « dilettantes ou contemplateurs » (dilettantes o contempladores), des « érudits et bibliophiles » (eruditos o bibliófilos), des « fétichistes de l'instrumentation » (organófilos), des « mégalomanes » (megalófilos) et enfin des « désaxés et théoriciens » (descentrados y teorizantes) [RAMÓN Y CAJAL, 1899, chap. 5]. Son opinion sur ces derniers n'est guère flatteuse : « Au fond, le théoricien est un homme paresseux déguisé en diligent. Sans s'en rendre compte, il obéit à la loi de l'effort minimum. Parce qu'il est plus facile de forger une théorie que de découvrir un phénomène. » (« En el fondo, el teorizante es un perezoso disfrazado de diligente. Sin percatarse de ello, obedece a la ley del mínimo esfuerzo. Porque es más fácil forjar una teoría que descubrir un fenómeno. ») Avec le succès, sur lesquels on reviendra, de la modélisation mathématique en neurosciences ce point de vue de Santiago Ramon y Caja est particulièrement ironique!

Aux biologistes la méthodologie réductrice de la physique paraît ingénue et naïve par rapport à la caractéristique vécue du nombre d'exceptions à la règle vécue dans les phénomènes biologiques. La pratique commune de la biologie s'interroge sur ce que des lois générales, « universelles », pourraient vraiment dire d'intéressant sur un système donné. Les objets et phénomènes biologiques seraient-ils trop « complexes » pour permettre une abstraction mathématique ou des expériences de pensée? Trop sujet à imprécisions, fluctuations et désordre? Comme le fait remarquer Suzannah Rutherford dans un éditorial intitulé « Towards a physical biology » de la revue Bioessays [RUTHERFORD, 2011] ces interrogations toujours présentes s'accompagnent cependant d'une nouvelle considération pour les approches théoriques.

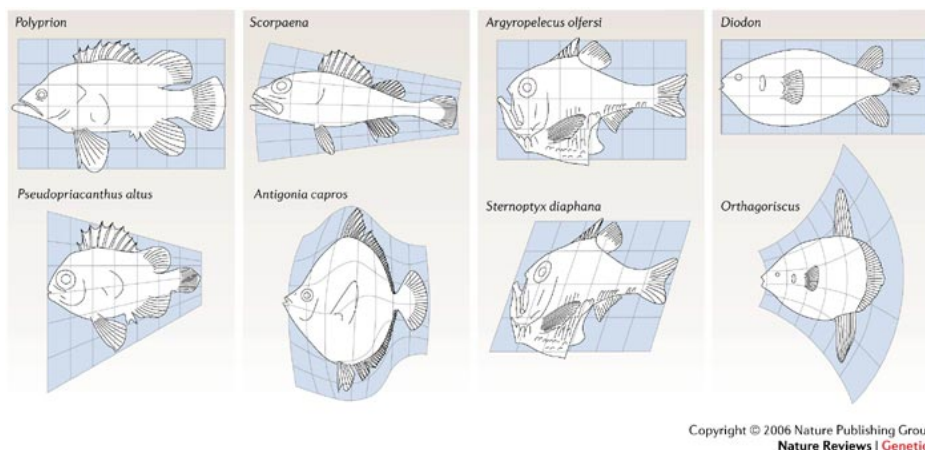


FIGURE 2.2 – D'Arcy Thomson : transformation de formes de poissons. Transformations mathématiques des formes d'espèces de poissons entre elles. Source :Wallace Arthur Nature Reviews Genetics 7, 401-406 (May 2006)

Ces dernières décennies l'importance des principes mathématiques en biologie est en effet de plus en plus reconnue. Tout comme aussi dans d'autres champs de recherche qui y étaient également éloignés comme les sciences sociales. Une tendance à une plus grande perméabilité entre les disciplines, et certainement aussi la puissance des applications des mathématiques dans la société moderne, à permis de diffuser l'idée selon laquelle des lois et des relations simples et générales auraient un rôle déterminant dans la croissance, la forme et la dynamique de systèmes complexes en dépit de la diversité et des subtilités des systèmes vivants.

Même si mal reconnues et non valorisées, ces idées ne sont pas neuves. Elles courraient en

biologie au moins depuis la fin du 19^e siècle. Par exemple le concept d'allométrie, voir l'encadré 2.1, introduit par le zoologiste écossais D'Arcy Thompson révèle des relations générales au travers des espèces vivantes. Il en va de même pour les principes universels des formes des organismes mis en évidence dans son célèbre ouvrage « On Growth and Forms » figure 2.2. Des propriétés générales qui suggèrent fortement comment la richesse du vivant s'exprime dans le cadre des lois naturelles et mathématiques qui en déterminent des caractéristiques fondamentales. Cette vision a grandement inspiré des physiciens mais semble avoir eu une influence moins profonde dans la communauté des biologistes en dehors de quelques courants plus interdisciplinaires. Cette situation semble être en profond changement.

2.1 Renouveau de la théorie et de la modélisation en biologie

« One of the most perplexing questions in connection with models is how they relate to theories. The separation between models and theory is a very hazy one and in the jargon of many scientists it is often difficult, if not impossible, to draw a line. So the question is : is there a distinction between models and theories and if so how do they relate to one another? »

Frigg, Roman and Hartmann, Stephan, « Models in Science » Stanford Encyclopedia of Philosophy [FRIGG et HARTMANN, 2017]

Conséquence des progrès technologiques époustouflants et des découvertes en biologie se succédant sans ralentissement depuis près d'un quart de siècle, la théorie, la modélisation et les approches numériques connaissent un développement important inédit en sciences du vivant. La fin des années 1990 a en effet été marquée par un tournant majeur dans les sciences de la vie : la publication des résultats des grands programmes de séquençage rendant disponible l'information des génomes d'espèces de plus en plus nombreuses. De nouvelles méthodes et outils d'analyse devenaient nécessaires, pour traiter de données jamais aussi abondantes et complexes, ainsi que de nouveaux cadres conceptuels pour questionner et interpréter les données.

À partir du tournant du millénaire de nombreux éditoriaux et rapports de sociétés savantes ou de commissions gouvernementales se sont donc penchés sur les transformations à mettre en place de programmes scientifiques et de politiques de formation à la suite de cette nouvelle donne [BAUERLE et collab., 2009; COMMITTEE ON UNDERGRADUATE BIOLOGY EDUCATION, 2003; HUTTER, 2000; MALAKOFF, 1999; SERVICE, 1999a,b]. Les programmes et projets discutés dans les références précédentes font tous apparaître la nécessité d'ouverture interdisciplinaire plus importante pour tenir compte des nouvelles masses de données et des progrès technologiques, ainsi que du renforcement des sciences formelles (mathématiques, informatique, physique) dans les cursus universitaires de sciences de la vie. Il s'agit d'évolutions cependant lentes et toujours d'actualité même après plus d'une décennie. Outre les changements de pratique, la nature même des savoirs scientifiques est interpellée : quelles sont les explications satisfaisantes quand nous avons accès à toute l'information génétique d'un organisme? Se pose en plein la question des théories scientifiques et des modèles d'explications en usage ou nécessaires pour les sciences du vivant.

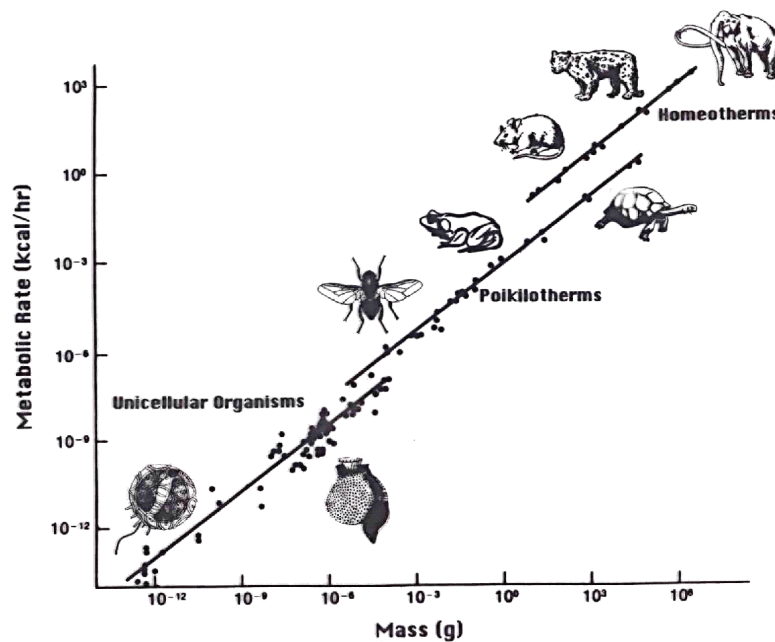
2.1.1 Théorie et modèles

Dans un récent éditorial du journal ELIFE, lancé en 2011 par le Howard Hughes Medical Institute, la Max Planck Society et le Wellcome Trust, les éditeurs Wenying Shou, Carl T. Bergstrom, Arup K. Chakraborty et Frances K. Skinner insistent sur le rôle essentiel du questionnement théorique dans les sciences de la vie [SHOU et collab., 2015] dont la communauté scientifique concernée doit davantage prendre conscience. De nombreux faits historiques montrent en effet comment théorie et modélisation ont joué un rôle essentiel pour de nombreuses avancées fondamen-

L'allométrie, est l'étude du changement des caractéristiques des organismes vivant avec leur taille fournit de nombreux exemples de relations générales. Caractérisées statistiquement de nombreuses relations ont été établies entre les dimensions des organismes, taille, masse, et leurs propriétés morphologiques et physiologiques. Ces relations établies depuis plus d'un siècle indiquent clairement comment de nombreux traits génétiques, morphologiques et fonctionnels doivent être interdépendants. Que l'évolution d'un de ces traits entraîne, ou bien est combinée avec l'évolution des autres. Les mécanismes de ces relations sont inconnus. Ils sont cependant, ils doivent cependant être, généraux, valables pour toutes les espèces dont les relations d'échelle mettent en évidence le partage de même principes communs [SHINGLETON, 2010].

Le graphique ci-dessous illustre l'allométrie dans le cas du taux métabolique, la vitesse de transformation de l'énergie, des organismes vivants. La loi statistique de Keibler établit que les taux sont répartis suivant une loi de puissance de la masse avec un exposant de valeur 3/4 :

(Taux métabolique B) ~ (Masse M)^{3/4}.



Allometric scaling of metabolic rate for a selection of homeotherms (birds and mammals), poikilotherms (fish, reptiles, amphibians, and invertebrates), and unicellular organisms. The solid lines all have a slope of .75. Modified from Hemmingsen, 1960.

L'existence de telles relations statistiques révèle des principes d'organisation subjacents universels à tous les organismes. Dans le cas de l'allométrie les relations géométriques fonctions de la taille des organismes sont déterminantes.

Encadré 2.1 – Allométrie

tales des sciences de la vie. Aussi bien d'ailleurs de théories et modèles basées sur des mathématiques que sans support d'un tel langage formel.

Théories En effet, l'élaboration d'une « théorie » consiste à l'identification et la description des mécanismes fondamentaux d'un domaine qui fournissent un cadre explicatif. La construction d'une théorie est l'aboutissement de la démarche scientifique basée sur l'expérimentation et l'observation. La théorie correspond à l'élaboration de notre compréhension des phénomènes observés et manipulés. Les sciences de la vie moderne, comme toutes les sciences naturelles – physique, géophysique, chimie, etc. – de nos jours sont déjà fondamentalement platoniciennes.

Les chercheurs analysent observations et collectes de données à la recherche de schémas, de tendances caractéristiques, de régularités dans les phénomènes étudiés. Cette recherche est de

nos jours essentiellement numérique. Elle consiste à scruter mesures et données avec les outils statistiques pour en extraire les tendances et descriptions principales. Des outils de plus en plus étendus au-delà des méthodes classiques par les apports d'algorithmes issus de l'intelligence artificielle. Les algorithmes de « machine learning », de « data mining » multiplient ainsi les panoplies d'outils à disposition des chercheurs pour extraire des signaux, des corrélations des tendances dans des données produites massivement.

Les résultats de ces analyses, l'identification de caractéristiques, n'est qu'une étape vers la compréhension, c'est-à-dire vers une théorie qui donne les outils pour rendre compte, expliquer, les observations. Il s'agit pour cela de formuler des hypothèses sur les mécanismes en action. D'imaginer et exprimer les processus conduisant aux phénomènes et comportements mis en lumière expérimentalement. On aura progressé vers leur compréhension scientifique quand on saura formuler un jeu d'explications mécaniques, c'est-à-dire basé sur des processus décrits, cohérent avec d'autres résultats et dépendant le moins possibles de différentes suppositions. Le nombre réduit et la simplicité des hypothèses nécessaires et cruciales, car leur multiplicité est source potentielle d'incertitude. Une nouvelle explication théorique aura toute sa valeur quand elle ne sera pas restreinte à l'explication d'une expérience unique particulière, mais constitue un cadre conceptuel plus large permettant l'explication de nombreuses observations. On pourra alors parler d'une véritable théorie scientifique valable pour le champ de phénomènes naturels en question.

Modèles La modélisation est une mise en œuvre de la théorie sur des cas d'applications. La théorie donne le cadre conceptuel et les méthodes avec lesquels construire les modèles de phénomènes et situations particulières observables. Le modélisateur doit décider du cadre théorique qu'il va appliquer pour décrire un phénomène particulier. Pour cela il devra aussi préciser avec attention l'objet d'étude. Délimiter le phénomène qu'il tentera d'expliquer qui rentrera dans le champ du modèle. La modélisation concernera ainsi seulement certaines propriétés retenues dans la tentative d'explication. Les modèles sont souvent présentés comme des représentations du réel. Une description de celui-ci dans un certain langage. L'affirmation n'est que partiellement correcte. Comme le dit très clairement Jeremy Gunawardena dans la citation en tête de ce chapitre et dans l'article d'où elle est extraite [GUNAWARDENA, 2014], un modèle ne représente au mieux que des facettes de la réalité, choisies avec bien des présupposés par le modélisateur.

Quel est le sens alors de la modélisation si d'emblée la portée des explications attendues est limitée? Rentre en jeu la question du pouvoir prédictif. Dans la construction d'un modèle, par le choix du langage, du formalisme utilisé, et du cadre théorique utilisé, on espère que l'explication constituée par le modèle admette une flexibilité d'application. C'est la possibilité d'utiliser des paramètres différents et des valeurs différentes des ingrédients du modèle pour l'utiliser dans des variantes des conditions. Le modèle a un pouvoir prédictif, d'une part si déjà il possède un champ d'applications large pour des conditions différentes de celles qui ont servi à le construire. Et d'autre part, s'il rend compte correctement, en conformité avec des observations autres, de propriétés des systèmes avec ces autres conditions. On peut alors attribuer à ce modèle une valeur explicative. On peut donner foi aux hypothèses qui ont conduit à sa formalisation, donc croire avec un fondement scientifique que le modèle et ses hypothèses ont un sens pour l'interprétation du réel.

2.2 Développement de la modélisation mathématiques en biologie

« I believe the day must come when the biologist will – without being a mathematician – not hesitate to use mathematical analysis when he requires it. »

Karl Pearson, dans l'épigraphe à la première édition de « On Growth and Form » de D'Arcy W. Thompson. D'après P. Ball [BALL, 2013]

L'approche théorique met souvent en œuvre des mathématiques. Mais ce n'est pas exclusif ni nécessaire. Dans le contexte des Sciences de la Vie, W. Shou et ses co-auteurs discutent largement dans l'éditorial déjà cité [SHOU et collab., 2015] cet aspect non mathématique de la théorie en biologie. Face aux transformations de la génomique ils appellent de nos jours à davantage de formalisation.

o

Les mathématiques ne sont pourtant pas étrangères à l'étude du vivant. James D. Murray, éminent bio-mathématicien, auteur de l'ouvrage monumental « Mathematical biology » [MURRAY, 2002] indique dans deux revues récentes sur l'évolution des rapports entre mathématiques et biologie [MURRAY, 2012] plusieurs exemples historiques anciens de scientifiques classiques, mathématiciens et biologistes, s'étant intéressés à des questions mixtes. Le premier usage de systèmes d'équations différentielles sur des questions d'épidémiologie date ainsi de 1760, avec un modèle de Bernoulli de l'immunisation à la variole [BERNOULLI, 1760; GABRIEL et DE LA HARPE, 2010], figure 2.3. Un travail précurseur de modélisation en épidémiologie qui n'eut pas de conséquences pratiques immédiates. La pratique de la variolisation ayant continué à susciter l'hostilité de nombreux médecins et la démarche innovante de Bernoulli a été oubliée.

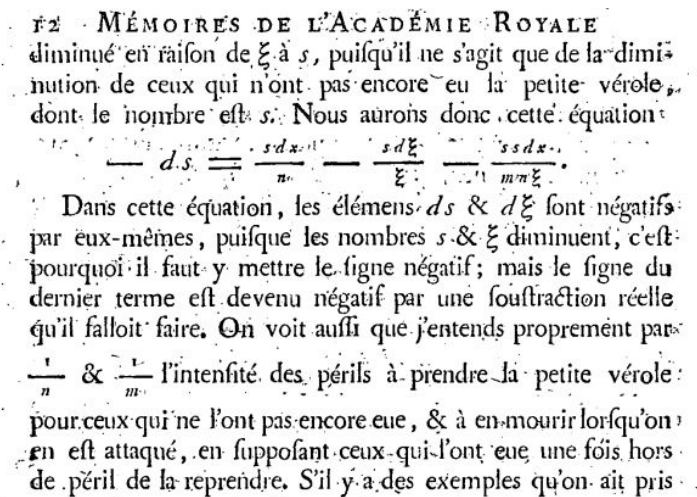


FIGURE 2.3 – **Reproduction de la première équation d'un modèle mathématique d'épidémiologie par Daniel Bernoulli.** Le travail précurseur de Daniel Bernoulli établit un modèle d'équations différentielles pour la propagation d'épidémie et l'immunisation plus de 160 ans avant leur réintroduction dans un cadre proche par MCKENDRICK. Source : « Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, & des avantages de l'inoculation pour la prévenir. »; disponible sur Gallica [BERNOULLI, 1760]

Un autre exemple d'usage ancien de la modélisation mathématique est dû à Isidore Geoffroy Saint-Hilaire³ en 1836 [SAINT-HILAIRE, 1832] dans son « HISTOIRE GÉNÉRALE ET PARTICULIÈRE DES ANOMALIES DE L'ORGANISATION CHEZ L'HOMME ET LES ANIMAUX, OUVRAGE COMPRENANT DES RECHERCHES SUR LES CARACTÈRES, LA CLASSIFICATION, L'INFLUENCE PHYSIOLOGIQUE ET PATHOLOGIQUE, LES RAPPORTS GÉNÉRAUX, LES LOIS ET LES CAUSES DES MONSTRUOSITÉS, DES VARIÉTÉS ET VICIES DE CONFORMATION, OU TRAITÉ DE TÉRATOLOGIE » à l'occasion duquel il a été un des premiers scientifiques à mettre en parallèle l'étude des formes physiques et biologiques, voir à ce sujet sa magnifique argumentation tirée de la préface de l'ouvrage reproduite dans la figure 2.4 .

Mathématiques et théorie de l'évolution La théorie de l'évolution fournit aussi un exemple particulièrement significatif de modélisation mathématique même si initialement elle n'a pas été formulée de cette manière et qu'elle ne se décrit pas nécessairement formellement. La formalisation

3. À ne pas confondre avec son père –et professeur– Etienne Geoffroy Saint-Hilaire, le célèbre naturaliste participant à l'expédition d'Égypte.

D'un autre côté, en même temps qu'une science, par l'accroissement numérique de ses faits particuliers, tend à se diviser, d'autres progrès lui font éprouver un autre besoin, lui impriment une tendance en apparence contradictoire : celle d'une *association* avec toutes les branches analogues des connaissances humaines. A mesure qu'elle s'élève à des généralités plus nombreuses et plus vastes, l'intervalle, d'abord immense, qui l'isolait, se comble et s'efface peu à peu ; et bientôt une alliance intime, féconde, également utile à toutes, ne permet plus de voir entre les sciences de même ordre que des rameaux distincts, mais étroitement unis, d'une même tige. Ainsi, par la grande Loi de l'attraction newtonienne, l'histoire tout entière des corps inorganiques repose sur des bases communes et ne semble plus qu'un vaste et immense corollaire du même principe. L'étude des êtres organisés, plus variés, plus complexes, modifiés à chaque instant par les phénomènes encore inexplicables de la vie, n'a pu être embrassée dans une aussi haute généralité ; mais déjà des principes communs à tout le règne animal, à tout le règne végétal et même à l'ensemble des deux règnes organiques, sont les magnifiques préludes des succès futurs. Ajouterai-je que déjà même il est peut-être permis d'entrevoir l'instant de haut progrès scientifique où, par les lois des courans, un admirable lien s'étendra sur la nature entière, et où se trouvera réalisé cet enchaînement de toutes les parties du grand ensemble, vers lequel tendent depuis si long-temps les efforts prématurés d'esprits audacieux ?

FIGURE 2.4 – Citation originale d'Isidore Geoffroy Saint Hilaire extraite de la préface de son monumental ouvrage sur la tératologie [SAINT-HILAIRE, 1832].

mathématique de la théorie de Darwin est en réalité la contribution majeure des premiers généticiens des populations comme Pearson, Wright, Fisher, Haldane dans la première moitié du XX^e siècle. Les fondements mathématiques solides de la théorie de l'évolution qu'ils ont établis ont eu un rôle déterminant pour sa confirmation (et ce faisant ont contribué aussi fortement au développement du champ de la statistique). Ce sont bien les mathématiques qui ont permis la démonstration fondamentale pour l'évolution de la fixation des mutations génétiques dans une population. C'est-à-dire une explication, justifiée mathématiquement, de l'existence de variations héréditaires sur lesquelles la sélection naturelle de Darwin peut agir.

À l'origine de la théorie de l'héritage génétique Une autre contribution centrale historique des mathématiques aux sciences du vivant peut-être reconnue dans les fondements même de l'hérédité génétique avec les recherches de Georges Mendel. Jeremy Gunawardena du département de biologie des systèmes de Harvard dans ses recherches sur l'historique de la biologie théorique et de la biologie des systèmes (voir par exemple [GUNAWARDENA]) fait effectivement noter que l'établissement des lois de l'hérédité par Gregor Mendel est le résultat d'une démarche fondamentalement quantitative et mathématique. Mendel n'était pas le seul biologiste de son époque expérimentant des croisements de plantes. L'originalité de son approche a consisté dans le dénombrement des résultats, dans la recherche de régularités et l'identification de schémas numériques les caractérisant. Et finalement dans l'élaboration d'un modèle de mécanisme pour en rendre compte. Une démarche qui a construit une théorie de l'héritage supposant l'existence de caractéristiques (« Merkmale ») héréditaires, qui ont été appelées « gènes » seulement plus tard en 1905 par le danois Wilhelm

Johannsen [JOHANNSEN, 1905, 1909] et identifiées chimiquement près d'un siècle après. Même si très élémentaires par rapport aux calculs usuels en physique, l'article encore très lisible de Mendel [MENDEL, 1866] (joliment accessible en ligne <http://www.mendelweb.org/MWpaptoc.html>) présente de nombreux tableaux et des schémas de relations représentant le travail quantitatif sur les données numériques et sa conceptualisation.

Mutations aléatoires, réactions enzymatiques, etc... De nombreux autres exemples du rôle central de la modélisation mathématique pour des résultats classiques de la biologie du 20^e siècle sont encore discutés par Jeremy Gunawardena [SHOU et collab., 2015] ainsi que W. Shou et ses co-auteurs [SHOU et collab., 2015]. Quelques cas marquants intéressants à souligner sont :

- La preuve de l'origine aléatoire des mutations génétiques qui est venue compléter les mécanismes moléculaires nécessaires à la théorie d'évolution de Darwin. Il s'agit de la démonstration par Luria et Delbrück en 1943 du caractère aléatoire des variations génétiques bactériennes dont l'origine était auparavant controversée. Une hypothèse en cours supposait que les bactéries des changements héréditaires pouvaient s'induire en réponse spécifique à des conditions de l'environnement plutôt que de manière purement aléatoire. C'est par un raisonnement de modélisation mathématique sur les prédictions des distributions de probabilité de survie attendues suivant les deux types d'hypothèses mécaniques que le débat a été tranché.
- Dans le domaine de la biochimie, l'explication des réactions enzymatiques par Leonor Michaelis et Maud Menten en 1913 par le biais d'un modèle mathématique faisant l'hypothèse de l'existence d'un complexe intermédiaire substrat-enzyme. L'idée d'un complexe transitoire lors de réaction enzymatiques avait déjà été introduite en 1889 par Savante Arrhenius [CORNISH-BOWDEN, 2013; WORTHINGTON, 2018]. Le succès du modèle de Michaelis-Menten repose sur la déduction algébrique d'une formule reproduisant des séries de mesures précises de cinétique qui a donné crédit à l'existence hypothétique de l'assemblage transitoire de l'enzyme avec le complexe. Une idée théorique qui a attendu trente ans avant sa preuve expérimentale par Britton Chance en 1943 pour l'enzyme peroxidase [CHANCE, 1999; KRESGE et collab., 2004]

À ces exemples de contributions de modélisation mathématique à des résultats classiques de la biologie moderne on peut aussi inclure les analyses mathématiques poussées de cristallographie qui ont été indispensables pour l'établissement de la structure en double hélice de l'ADN. De nos jours sous le terme de modélisation mathématique en biologie il est souvent question de descriptions de systèmes évoluant dans le temps. C'est l'objet de l'essentiel des chapitres d'ouvrages de références de « Biomathématiques » comme l'ouvrage de James D. Murray « Mathematical Biology » [MURRAY, 2012]. Il est intéressant de discuter plus spécifiquement cette catégorie de modèles fortement diffuse dans les pratiques actuelles.

2.2.1 Systèmes dynamiques mathématiques pour la biologie

Dans certains champs d'études biologiques la modélisation mathématique de l'évolution des systèmes dans le temps est pleinement présente. C'est le cas de l'écologie où l'usage de modèles théoriques formels s'est développé depuis l'introduction de modèles de compétition prédateurs-proies en 1925 par Alfred James Lotka [LOTKA] et Vito Volterra en 1926 [VOLTERRA, 1926] connus actuellement sous le nom de modèles Lotka-Volterra [CHAZOTES et MONTICELLI, 2013]. À la même époque Anderson Grey McKendrick et William Ogilvy Kermack ont introduit une démarche similaire de modèles sous forme d'équations différentielles pour décrire la transmission d'infections [KERMACK et MCKENDRICK, 1991a,b,c]. Des idées réinventées près de 160 ans après Bernoulli. Depuis, l'épidémiologie n'est pas le moindre secteur faisant usage des mathématiques avec un impact critique pour la santé publique.

Les neurosciences sont un autre secteur fortement marqué par la modélisation mathématique dynamique. Cela fera l'objet d'un chapitre de ce mémoire. En dépit de l'opinion de Santiago Ramon y Casal présentée plus haut, les mathématiques y sont devenues essentielles depuis l'explication de l'activité électrique des neurones par Alan Lloyd Hodgkin et Andrew Fielding Huxley en 1952 [HODGKIN et HUXLEY, 1990]. Des travaux qui sont encore un des exemples encore le plus stupéfiant de démarche combinant expérimentation et modélisation théorique. Partant d'hypothèses sur les mécanismes ioniques, Hodgkin et Huxley ont modélisé le système responsable de l'initiation et de la propagation des potentiels d'actions sous forme de schéma électronique traduit ensuite en équations différentielles non linéaires dont ils ont déterminé les paramètres par des ajustement avec des séries d'expériences sur l'axone géant du calamar. Le modèle d'Hodgkin-Huxley est toujours à la base de modélisations détaillées actuelles et indispensable pour la compréhension du neurone.

2.2.2 Théories de la morphogénèse et auto-organisation

L'après-guerre et le début des années cinquante a été très marqué par le développement des premiers ordinateurs stimulés par les besoins de calcul militaire. Le développement des technologies matérielles et logicielles de calcul a été une source d'inspiration pour de nombreux chercheurs qui ont commencé à aborder les phénomènes naturels, et en particulier le vivant sous l'angle de processus calculatoire, avec notamment le mouvement de la cybernétique⁴.

Au même moment que Alan Lloyd Hodgkin et Andrew Fielding Huxley publiaient leurs résultats, Alan Turing publiait une exploration pionnière sur le vivant des bases chimiques et mathématiques de la morphogénèse [TURING, 1952]. Mais d'après le biomathématicien James D. Murray, grand spécialiste de ces travaux, contrairement au modèle de neurone de Hodgkin-Huxley, les idées théoriques de Turing auraient été largement ignorées jusqu'à la fin des années soixante quand elles furent redécouvertes par Ilya Prigogine et Gregory Nicolis [MURRAY, 2012; PRIGOGINE et NICOLIS, 1967]. L'école de Bruxelles sur la thermodynamique hors d'équilibre et l'auto-organisation initiée par Ilya Prigogine a été un des endroits les plus actifs de modélisation biomathématique de systèmes biochimiques durant les années 1970 [PRIGOGINE et collab., 1969]. S'en est suivi un engouement d'un grand nombre de théoriciens pour des modèles biophysiques de génération de formes et structures. Murray fait correspondre à cette période la véritable origine de l'essor de la modélisation mathématique en biologie au-delà des applications d'épidémiologie et d'écologie [MURRAY, 1977]. Depuis cette époque la modélisation mathématique s'est constamment développée, notamment au tournant du millénaire.

2.2.3 L'essor de la modélisation mathématique et ses limites

« In the thirteen years since the first edition of this book appeared the growth of mathematical biology and the diversity of applications has been astonishing. Its establishment as a distinct discipline is no longer in question. One pragmatic indication is the increasing number of advertised positions in academia, medicine and industry around the world; another is the burgeoning membership of societies. People working in the field now number in the thousands. Mathematical modelling is being applied in every major discipline in the biomedical sciences. »

J.D. Murray [MURRAY]

Dans l'éditorial de la troisième et plus récente édition de son ouvrage « Mathematical Bio-

4. La cybernétique, introduite en 1948 par le mathématicien Norbert Wiener désigne un ensemble de recherches sur les principes de fonctionnement autonome de systèmes vivants et artificiels impliquant des processus de contrôle, de communication et de rétroaction. Ce courant a eu des répercussions importantes sur le développement des sciences cognitives, de l'intelligence artificielle, et des théories biologiques de l'auto-organisation

logy » publié en 2003 d'où est tirée cette citation d'ouverture, J.D. Murray souligne à quel point la pratique s'était développée et allait encore se développer et connaître une notoriété jusqu'à maintenant sous un florilège de dénominations comme « bio-mathématiques », « mathématiques biologiques », « biologie théorique » et « biologie des systèmes ».

Pourquoi alors en dépit de cette reconnaissance et des exemples de plus en plus nombreux d'applications a-t-on encore autant la sensation d'une division en biologie entre recherches théoriques et expérimentales? Encore à la soumission d'articles dans des revues scientifiques biologiques on est rappelé d'enlever autant que possible toute formule du corps du texte sous peine de faire fuir des lecteurs potentiels (une recommandation d'un éditeur vécue récemment à la soumission d'un article). Ou de renvoyer les formules dans les annexes, seul lieu possible d'ailleurs pour des démonstrations.

W. Shou et ses co-auteurs suggèrent qu'une raison à cette situation peut simplement venir de la *difficulté*. Les systèmes biologiques sont complexes, ils comportent de nombreuses parties en interaction, sont très hétérogènes et impliquent des échelles temporelles et spatiales très larges. Une question biologique même simple exige des mathématiques relativement sophistiquées. En conséquence la lecture des articles théoriques est probablement souvent trop difficile pour une communauté insuffisamment formée aux mathématiques. Les publications théoriques trop abstraites, rédigées pour des spécialistes et ne faisant pas suffisamment de lien avec la réalité expérimentale sont loin des préoccupations de laboratoires appliqués.

Mais cette situation évolue, les groupes mixtes sont de plus en plus nombreux combinant sur un même projet expérimentation et théorie. Dans l'enseignement universitaire également la formation des biologistes intègre souvent davantage de mathématiques plus avancées. Des revues de biologie de premier plan se prononcent dans ce sens comme l'éditorial cité de W. Shou et ses co-auteurs. On va résolument vers une présence plus significative de la théorie mathématique en biologie. Les discours actuels sur la « nouvelle » place de théories mathématiques en biologie vient donc surtout des secteurs de la biologie qui en avaient perdu l'habitude et qui concernent les fondements moléculaires et génétiques. Des secteurs qui vivent actuellement une frénésie de prodigieuses avancées à tous les points de vue, donnant lieu à de nouvelles démarches d'investigation. Différents qualificatifs ont vu le jour et sont utilisés depuis le tournant du millénaire. On parle de « Biologie Quantitative », de « Biologie des systèmes », de « Biologie de synthèse ». Ces thèmes sur lesquels on s'étendra un peu plus bas sont le cadre dans lequel s'est manifesté un nouveau besoin de théorie mathématique pour la biologie.

Pourquoi une modélisation « mathématique », « informatique », « physique »? Les mathématiques ont un double rôle. En modélisation elles sont à la fois un langage à disposition des modélisateurs pour décrire les systèmes et un outil de calcul pour exprimer le potentiel prédictif des modèles. De nombreux formalismes mathématiques sont possibles. Dans le concept de « modélisation mathématique » on peut de manière générale comprendre toute description de manière formelle d'un phénomène, suivant un langage conventionnel obéissant à des règles clairement définies et opérations logiques. Automates cellulaires, agents de calculs, codes Monte-Carlo sont ainsi des démarches permettant de décrire rigoureusement, de formaliser, des modèles sans nécessairement faire intervenir explicitement des équations. Il peut s'agir de listes de règles, des schémas opérationnels ou des codes informatiques de simulation, sans correspondre forcément à une écriture algébrique (quoique des mathématiques d'un niveau d'abstraction élevé devraient le permettre).

Dans la plupart des modélisations l'informatique est désormais incontournable pour effectuer les nombres de calculs nécessaires pour des modèles sophistiqués, mais pas seulement. L'informatique, en tant que science de l'information traite également de langages de modélisation, de l'établissement de formalismes suffisamment expressifs pour capturer et représenter des situations complexes. Les modèles ne sont pas écrits uniquement sous forme d'équations, il peut s'agir de listes d'instructions, de règles de schémas codifiés mis sous une forme informatique qui permet leur exploitation. Les interactions génétiques et biomoléculaires sont en effets complexes, et

la plupart du temps nombreuses. L'action d'une enzyme sur un substrat, d'un facteur de transcription sur un gène ou d'une protéine au sein d'une association n'est généralement pas simplement proportionnelle, linéaire. Or les systèmes non linéaires, même à quelques composants seulement sont notoirement difficiles à analyser. Ils sont capables de comportements « complexes », avec de nombreux régimes dynamiques et stationnaires. Les mathématiques associées à l'informatique sont indispensables à leur analyse. Elles le sont encore davantage quand les systèmes biologiques ont une plus grande taille.

Dans cet ensemble comment intervient la physique? Sa contribution est pragmatique, le raisonnement physique et les lois physiques servent à la construction des modèles. En tant que science expérimentale dite « exacte », s'appuyant traditionnellement déjà depuis longtemps sur une formalisation mathématique, la physique à l'habitude de la modélisation de phénomènes naturels, de leur résolution mathématique et de la confrontation avec l'expérimentation. Elle est à même d'assurer le lien entre l'abstraction mathématique et la réalité des observations.

On vient de voir une histoire diversifiée de la modélisation mathématique en sciences de la vie. J'ai déjà fait déjà allusion aussi en faisant plusieurs citations à des articles d'opinion récents aux transformations récentes en cours dans le domaine qui renouvellent l'importance des approches théoriques mathématiques en biologie. Il est temps de développer et discuter maintenant ces nouvelles tendances. Il s'agit dans un premier temps de la biologie à grande échelle, qui alimente le courant plus théorique de la biologie des systèmes.

2.3 Biologie à haut débit et quantitative

« After all, until now, biological complexity could only be understood through the study of the individual parts. In molecular biology and biochemistry, the questions have always been more of a qualitative nature, such as, have we cloned the relevant gene or not? Is it present and active? In which cells? Systems biology is bound to take away that burden, to become much more quantitative, and to offer a more integrated perspective on the inner workings of a cell, without having to resort to vagueness. »

Geert Potters « Systems Biology of the Cell » Nature Education (2010) [[POTTERS, 2010](#)]

La fin du millénaire a en effet été marquée par la possibilité technologique de connaître l'information génétique complète d'organismes vivants et de suivre l'activité génétique et moléculaire à des échelles radicalement nouvelles. Alors que la vision des chercheurs était auparavant concentrée sur les propriétés et les fonctions de quelques gènes isolés, il devenait envisageable de chercher à comprendre comment tous les gènes et leurs produits associés d'une cellule travaillent ensemble.

La découverte de la structure de l'ADN en 1953 par Watson et Crick, a donné naissance à la biologie moléculaire moderne. Cette caractérisation faisait suite et parachevait la synthèse moderne de la théorie de l'évolution qui avait conduit durant la première moitié du XX^e siècle à établir le lien entre la théorie de l'hérédité de Mendel et l'évolution naturelle selon Darwin. Il s'en sont suivi plusieurs décades ayant conduit à la connaissance des structures de gènes, de leurs protéines associées, de leurs mécanismes d'expression et des effets de mutations sur leur fonctionnement. Ces recherches ont été nourries d'une succession d'avancées des outils et méthodes qui ont finalement à aborder le fonctionnement d'un système biologique dans son ensemble.

2.3.1 La poussée technologique

La biologie a naturellement bénéficié des progrès technologiques de notre civilisation. Ses progrès correspondent à une conjonction d'avancées interdépendantes de nombreux secteurs. Des

développements inspirés par d'autres disciplines scientifiques et d'ingénierie, tel que l'aéronautique, les télécommunications, le spatial, le nucléaire et la physique des particules et la chimie de synthèse ont permis l'électronique, l'informatique, la robotique et l'optique moderne : le panel de technologies de base faisant les instruments de la biologie moderne. À ces ressources se rajoutent des étapes cruciales propres aux bio-technologies, mettant en œuvre des entités biologiques et chimiques liées à la biochimie du vivant. En premier lieu le progrès technologique fondamental a été la capacité de déchiffrer l'ADN grâce aux méthodes de séquençage de Maxam et Gilbert 1977 suivie de celle de Sanger la même année. Également primordiale, la capacité de multiplier l'ADN, en premier grâce au clonage de micro-organismes, et plus tard avec un gain d'efficacité et de précision à partir de la seconde moitié des années 80 la possibilité d'amplifier *in-vivo* la quantité d'une séquence d'ADN souhaité par la PCR (Polymerase chain reaction) développée par Kary Mullis et Michel Smith.

Le séquençage de génomes Cette évolution a été époustouflante. La première séquence d'un virus *phiX174* de 5,368 paires de bases (bp) a été laborieusement obtenue par Fred Sanger en 1977, suivie dans ce qui allait être une compétition passionnante entre groupes aux approches très différentes du premier procaryote séquençé *Haemophilus influenza* en 1995 à la surprise générale par l'équipe de Craig Venter et Hamilton Smith en dehors des initiatives publiques, rapidement suivi du premier organisme eucaryote *Saccharomyces cerevisiae* en 1996 par une collaboration de 74 laboratoires, du premier multi-cellulaire *Ceanorabditis elegans* en 1998 et finalement la publication d'un premier brouillon du génome humaine au tournant du millénaire en 2001 et sa version complète en 2003. L'évolution des technologies de séquençage s'est encore accentué depuis cette période glorieuse. Ces techniques ont elles même été révolutionnées par différentes méthodes massivement parallèles durant la dernière décennie, que l'on a d'abord désigné par « next-generation sequencing », maintenant plutôt qualifiées de méthodes à haut, ou ultra-haut débit (« (Ultra) high-throughput »). Le séquençage atteint le niveau personnel. Pour un millier d'euros on peut désormais séquencer le génome d'un individu humain. L'efficacité est telle que des cellules individuelles animales peuvent être séquençées et avoir accès aux quelques variantes génétiques et mutations qui ont pu s'établir sur des lignées cellulaires différentes [SHENDURE et collab., 2017].

La *génomique* décrit la structure des l'ensemble de l'information génétique. Non seulement tous les gènes sont connus, mais aussi leurs positions sur les chromosomes et les séquences non codantes –majoritaires en taille– entre les gènes. À cette description de l'intégralité des séquences et des gènes se sont ajoutées d'autres champs complémentaires. Les mêmes fondements technologiques qui ont conduit aux séquenceurs ont permis le développement d'instruments à très haut débit permettant d'analyser l'activité de tous les gènes.

De la structure à l'activité des génomes Un florilège de méthodes tirant parti des avancées de l'imagerie, et notamment des innovations biochimiques en marquage fluorescent, permet non seulement d'identifier les gènes actifs, mais aussi leurs niveaux d'activité. Les gènes sont identifiés, leurs niveaux de transcription mesurés ainsi que les abondances des protéines associées. On parle alors de nouveaux secteurs d'étude. La *transcriptomique* désigne ainsi l'observation et l'étude de la transcription génétique, première étape de l'expression des gènes. La *protéomique* caractérise l'étape suivant et étudie l'ensemble des protéines présente résultat final de l'expression, modulé encore par de nombreux processus cellulaires.

De nombreuses limitations et incertitudes expérimentales existent bien entendu. Tous les éléments et tous les détails sont loin de pouvoir être caractérisés. La réalité des possibilités expérimentales est plus nuancée et contrastée. Si la structure du génome par le séquençage est assez fermement établie, les mesures d'activité sont plus aléatoires. Les biais expérimentaux plus nombreux. Suivant les objets, organismes d'étude c'est l'activité d'un grand nombre des gènes qui peut être mesurée, mais approximativement avec de grandes barres d'erreur et des faux positifs. La plu-

part des méthodes de mesure doivent composer entre quantité et qualité.

Mais des possibilités inédites sont apparues aussi sur le front de la précision. Les fondements technologiques qui ont permis d'approcher l'intégralité de l'activité génomique ont aussi conduit à la possibilité de faire des mesures à très grande précision, mais de quelques éléments seulement. La mise au point de la GFP (Green Fluorescence Protéine) a été une étape majeure de l'imagerie biologique. Différentes manières génétiques et moléculaires existent pour fusionner la séquence de cette protéine fluorescente à la suite de la séquence de toute protéine ciblée. L'expression de la protéine d'intérêt se fera alors sous la forme d'une protéine hybride qui combine l'unité d'intérêt biologique à l'unité fluorescente servant à la détection. La détection est positive que dans les conditions d'expression de la protéine fusionnée cible. De nombreux développements ont raffiné cette technique pour l'appliquer dans une diversité de situations. Avec des caméras de microscopie elles aussi toujours plus performantes on peut suivre avec une haute précision les niveaux d'expression de protéines au cours du temps tout en enregistrant des informations sur la localisation spatiale.

Exemple récent parmi les plus spectaculaires le Genotype-Tissue Expression (GTEx) Consortium suit l'activité de près de 10000 traceurs génétiques dans 44 emplacements de tissus cellulaires différents pour 449 individus [ARDLIE et collab., 2015]. On est donc dans une ère de production sans précédent soit de très grandes quantités de données sur la structure et le fonctionnement des génomes soit de données de très haute précision. L'abondance de données conduit donc à parler d'une nouvelle « biologie à haut débit ». Des outils d'analyse appropriés sont clairement nécessaires pour manipuler et explorer ces masses de données, mais aussi de nouveaux paradigmes de questionnement pour explorer la complexité d'un système vivant dans sa globalité. Cette échelle d'étude se trouve désignée de plusieurs termes « Biologie des Systèmes », « Biologie systémique », « Biologie Intégrative », ou parfois « Biologie holistique ». On va maintenant discuter davantage de ce qu'il s'agit. Je retiendrais la désignation « Biologie des Systèmes ».

2.4 La biologie des systèmes

« Scientific fields, like species, arise by descent with modification, so in their earliest forms even the founders of great dynasties are only marginally different than their sister fields and species. It is only in retrospect that we can recognize the significant founding events. » « Systems biology offers an opportunity to study how the phenotype is generated from the genotype and with it a glimpse of how evolution has crafted the phenotype. »

Marc Kirschner, « The meaning of systems biology » [KIRSCHNER, 2005]

La biologie des systèmes désigne le champ d'étude visant la compréhension d'un système biologique dans son ensemble. Un objectif qui dans ces termes est bien imprécis, voir arrogant. De nombreux auteurs voient la biologie des systèmes comme la discipline qui étudie comment la physiologie et les phénotypes émergent à partir des interactions moléculaires [ALON, 2007; GUNAWARDENA, 2010; KIRSCHNER, 2005]. Il s'agit alors de comprendre comment les caractéristiques principales d'un organe, d'un organisme, d'une cellule, d'un compartiment biologique sont déterminés à partir des caractéristiques microscopiques génétiques et moléculaires.

Cette démarche est issue des technologies de transcriptomique et protéomique qui produisent des quantités massives discutées plus haut et s'appuient fortement sur des méthodes statistiques et informatiques pour déduire des corrélations et établir des liens de causalité probables. Un travail dit de fouille de données (« datamining ») dont l'objectif est de mettre en évidence des interactions entre objets biologiques. Mais l'identification de corrélations n'est qu'une étape vers la compréhension des phénomènes. Prédire l'action d'un élément sur un autre est certes intéressant, mais ne sera pas très utile si cette action possible est seulement l'une parmi de nombreuses, ou

bien si ses conditions d'activation ne sont pas connues avec précision.

2.4.1 Quête d'une définition de la biologie des systèmes

La définition de la biologie systémique a fait l'objet de nombreux articles et éditoriaux depuis les années 2000. L'analyse par l'outil Ngram de Google (<https://books.google.com/ngrams>) qui répertorie l'usage de termes lexicaux dans les ouvrages recensés sur Google books, fait apparaître le début d'une croissance linéaire de l'usage du terme « systems biology » depuis les années 2000. L'évolution du terme « integrative biology » suit une autre tendance suggérant des usages différents. Le sujet est encore régulièrement abordée dans les principales revues scientifiques, soit pour en faire la promotion, la critique ou une mise en perspective épistémologique. Il est curieux que sur une période déjà assez longue la communauté scientifique en science de la vie continue à s'interroger sur la signification d'un de ses axes. D'autant plus d'une orientation qui pourtant se veut aussi englobante. Peut-être que son positionnement par rapport à la continuité des pratiques n'est finalement pas aussi radicalement différent? Ou bien que l'importance scientifique, hautement revendiquée par les tenants de la démarche, des nouvelles perspectives de questionnement n'est finalement pas, ou pas encore, aussi claire? Ou encore que l'appropriation large des pratiques de biologie systémique est pour des raisons techniques, scientifiques, structurelles encore très difficile et demande plus de temps? Toujours est-il que le domaine prête à controverses et fait débat. Il en est ainsi car on est en train de connaître un changement de paradigme dans la vision du fonctionnement moléculaire du vivant.

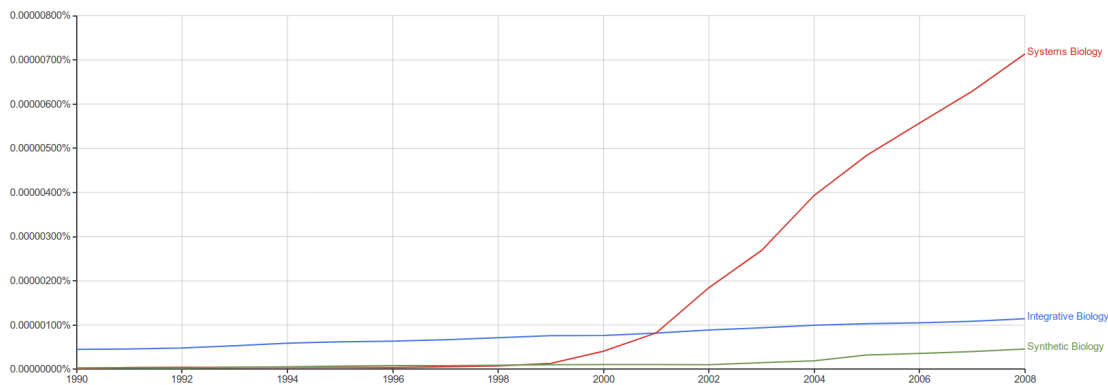


FIGURE 2.5 – Ngram des termes « Systems Biology », « Integrative biology » et « Synthetic biology ». Graphique de la fréquence par année d'usage des termes « Systems Biology », « Integrative biology » et « Synthetic biology » dans les corpus textuels de Google, ici entre 1990 et 2008. Ces graphiques montrent l'évolution de tendance de popularité d'expressions et de termes. Le champ de la biologie synthétique suit la croissance de la biologie des systèmes.

2.4.2 Pléiotropie et besoin d'intégration

Marc Kirschner, fondateur du département de Biologie des Systèmes à la Harvard Medical School, relève dans un éditorial dans *Cell* sur la nature de la biologie des systèmes [KIRSCHNER, 2005] qu'un élément est probablement central dans la vision proposée par la biologie des systèmes. Depuis l'origine de la génétique prévalait en effet un paradigme simple d'une relation directe entre un gène, sa protéine et sa fonction. Or la biologie est en train de s'éloigner de plus en plus de cette vision simpliste. L'image moderne fait apparaître des rôles multiples, un fonctionnement en réseau dont le sens apparaît au niveau d'un système composé et non de chaque pièces individuelle. La biologie des systèmes vient de ce nouveau besoin d'intégration.

L'hypothèse introduite par Georg Beadle et Edward Talum en 1941 de la relation « un gène, une enzyme (donc une fonction) » a été le moteur essentiel à l'origine du développement de la

biologie moléculaire. Une idée fondamentale, qui sera assez rapidement reformulée dans la vision déjà plus large de « un gène, une protéine, un rôle dans une fonction ». Et qui surtout aura guidé la génétique moléculaire pendant les décennies suivantes pour en découvrir les mécanismes et les recherches d'association génétiques avec des phénotypes.

Or de plus en plus de biomolécules sont caractérisées finement et sont désormais associées à des processus différents suivant les contextes. Marc Kirschner cite des exemples comme la tubuline, les enzymes kinase PI3, les protéines de signalisation Rac, qui peuvent jouer des rôles multiples dans la constitution des cellules, la croissance, le développement, les déplacements cellulaires et de nombreuses autres fonctions cellulaires. Tous ces constituants sont pléiotropes, ils interviennent dans différentes fonctions et déterminent différents phénotypes. Face à cette pluralité d'implications comment prédire les phénotypes résultant de mutations des gènes associés ?

La somme de connaissance et de données de plus en plus disponibles venant de la biologie à haut débit discutée précédemment a conduit à prendre pleinement conscience de la multiplicité des rôles des gènes et de leur protéines. D'une réutilisation de ces briques dans différents processus suivant le contexte cellulaire. La biologie moléculaire est une réalité beaucoup plus imbriquée que l'on pouvait croire pour pouvoir espérer assigner des fonctions précises à des gènes que l'on aurait correctement annoté dans une base de données.

La biologie des systèmes s'est développée sur ce constat car elle propose de travailler au niveau d'ensembles d'unités moléculaires contribuant à une fonction. Dans cette vue le phénotype est le résultat d'une action de groupe. Il faut s'intéresser à la structure et la dynamique de l'ensemble, les pièces individuelles ne suffisent pas. L'objet d'étude change, d'une focalisation traditionnelle sur le gène et la protéine, la biologie des systèmes élargit le regard à l'ensemble de pièces et leurs interactions : les nouveaux objets d'étude seront les réseaux biologiques.

2.4.3 Pléiotropie et nombre de gènes, vers les réseaux d'interaction

Cette constatation d'une pléiotropie de l'implication des gènes dans de multiples fonctions vient compléter la réponse à un paradoxe issu des résultats des programmes de séquençage. La constatation que l'humain n'a qu'un nombre en apparence très limité de gènes, sur les 30000 a grandement surpris au début des années 2000. Un nombre de gènes à mettre en regard des plusieurs centaines de milliers de gènes découvert dans certaines plantes. Les espèces vivantes ne semblent pas avoir des nombre de gènes en rapport à la sophistication, que nous percevons, de leur développement et au nombre de fonctions de l'organisme. Le fait même d'une surprise sur le nombre de gènes, de considérer qu'il s'agit d'un paradoxe, peut s'expliquer par l'ancrage à la vision ancienne linéaire entre un gène et une fonction clairement associée. Les multiples usages des gènes et de leurs produits suivant les conditions démultiplie les possibilités d'un ensemble donné d'éléments par effet combinatoire. Un argument supplémentaire parlant pour l'imbrication des pièces du vivant et leur réutilisation dans des applications multiples.

Un élément supplémentaire pour expliquer les possibilités de multiples combinaisons d'un nombre de gènes limité est d'avoir pris conscience qu'une grande partie de la complexité du génome résidait dans la régulation de celui-ci. En effet l'analyse des séquences fait apparaître que les régions cis-régulatrices des gènes, c'est-à-dire en amont de ceux-ci ont souvent une grande sophistication permettant un contrôle combinatoire par de nombreux facteurs. L'activité de chaque élément est déterminée par des règles de combinaisons de facteurs qui permettent de l'adapter pour différentes situations.

Une idée clé que l'on peut discerner dans les arguments ci-dessus c'est le rôle central des interactions multiples possibles entre gènes et protéines. Les rôles des gènes et des protéines ne peuvent se concevoir qu'au travers de leurs interactions avec une multiplicité d'autres agents, soit pour affecter leur expression ou leur activité, soit pour déterminer en s'associant à d'autres partenaires la réalisation ou non d'un processus. La biologie des systèmes possède ainsi un paradigme central, la notion de réseau d'interactions biochimiques, aussi appelé « circuit biochimique » ou

encore « graphe d'interactions », dans lesquels gènes, leurs produits et toutes les entités moléculaires interagissent pour assurer la vie cellulaire. On reviendra sur la notion de réseau biochimique plus en profondeur dans le prochain chapitre.

2.4.4 Les réseaux : paradigme scientifique et de société

La biologie n'est pas le seul domaine scientifique qui a connu l'essor d'une vision à grande échelle avec un vision intégrative mettant en avant les réseaux complexes de nombreux agents en interactions. Sciences sociales, urbanisme, technologies sont autant de domaines qui ont développé des approches analogues à la biologie des systèmes. La même explosion technologique de l'informatique, des capacités de stockage de l'information et de la mise en relation via internet (un réseau aussi...) a stimulé et soutenu l'émergence de visions équivalentes dans des domaines bien distincts. Les sciences humaines peuvent ainsi s'alimenter de la recherche sur des communautés en ligne d'internet, permettant l'exploration des habitudes, de la formation de groupes et l'anticipation de tendances.

Ces résultats et méthodes ont été centraux, et le sont encore, pour le développement des géants de l'économie numérique, dont le modèle d'affaires dépend de l'analyse des réseaux des utilisateurs et consommateurs et de leur dynamique. On comprend que l'informatique au sens de la science de l'information concerne ici bien directement toutes les disciplines où de l'information circule et est transformée. En première ligne naturellement la biologie. Les transformations en sciences de la vie, l'essor de l'idée de biologie des systèmes, serait la transposition dans ce domaine scientifique des grands courants technologiques et sociétaux qui traversent notre époque.

2.4.5 Biologie et informatique : une fertilisation réciproque

Si le développement de l'informatique et des réseaux numérique a été primordial pour ces transformations transversales dans tous les secteurs, il serait imprudent d'en attribuer la responsabilité à eux seuls. Une plus grande imbrication de l'avancée des sciences et technologies dans tous domaines existe. Les ordinateurs, l'informatique des réseaux ont certes amené les matériels et les méthodes indispensables à l'essor des études des réseaux. Mais ces technologies sont elles-mêmes sous l'influence de l'évolution des idées des sciences, de la biologie et de la génomique et des neurosciences en particuliers. Il existe en effet une longue histoire de la contribution de la pensée biologique à l'informatique qui remonte aux fondateurs même de cette science et les travaux pionniers sur les ordinateurs.

La biologie de servi de référence, comme en réalité seul exemple disponible, pour les réflexions sur l'architecture nécessaire pour un système artificiel capable de traitement d'information et programmable, c'est-à-dire capable d'être adapté à des situations différentes. Cette notion a été formalisée par Alan Turing dans ses travaux mathématiques fondamentaux pour le développement de machines de calcul programmables. Par extension c'est ce cadre qui l'a amené par la suite dans l'après guerre et avant sa fin tragique à être un pionnier de la modélisation mathématique de la morphogénèse.

À la même époque, un autre mathématicien Norbert von Neumann allait définir l'architecture toujours actuelle des ordinateurs contemporains en réfléchissant explicitement à la structure logique d'un système vivant capable de s'auto-reproduire. Von Neumann a également introduit la notion d'automates cellulaires, comme une architecture possible de processeurs numériques. Directement inspiré du vivant, le nom le suggère également, il s'agit de systèmes de calcul constitués d'unités autonomes capables d'interactions simples avec leur voisinage. Avec certains jeux d'interaction le comportement d'ensemble d'un collectif d'automates est une « Machine de Turing », à savoir un objet avec les capacités formalisées par Alan Turing de pouvoir tout calculer, tout simuler en fonction d'une programmation adéquate.

Les automates cellulaires sont toujours un paradigme très puissant pour la représentation de systèmes naturels et ont donné lieu à des approches scientifiques et technologiques nombreuses. Des applications qui ont connu cependant jusqu'à présent une moindre destinée que des approches plus classiques. L'essentiel de l'électronique que nous connaissons suit le paradigme de von Neumann d'un processeur central de calcul. En simulation scientifique, les automates cellulaires sont régulièrement utilisés plutôt dans des explorations conceptuelles que des applications de production.

Le développement logiciel a également connu des fertilisations diverses avec la biologie, notamment en intelligence artificielle. Les neurosciences ont inspiré le développement des algorithmes d'apprentissage automatique des réseaux de neurones formels mettant en œuvre les règles d'apprentissage de Hebb (qui seront discutées dans le chapitre 6). La théorie de l'évolution a conduit aux algorithmes génétiques reproduisant des mécanismes de mutations aléatoires et de sélection par rapport à un objectif dans une vue d'optimisation. La biologie, et notamment la génétique est en grande partie une science de l'information car portant sur l'étude de l'hérédité, de l'information transmise entre génération et de la séquence des gènes. Les rapports entre biologie et informatique ne doivent pas surprendre et ne se limitent pas aux exemples précédents. La biologie synthétique dont il sera question plus bas en est une autre illustration.

L'évolution des sciences et des technologies est dans un même éco-système produisant co-évolution et inspirations croisées. La biologie n'est pas isolée et la biologie des systèmes est le développement dans le secteur du vivant d'un courant global. Une manifestation importante de la convergence technologique conduisant à la biologie des systèmes est son versant biotechnologique : le domaine actuellement connu sous la désignation de « biologie de synthèse », dont on va discuter à présent étant donné ses liens avec les thèmes de mes recherches.

2.5 Biotechnologies et biologie de synthèse

Se basant sur les données, les réseaux biochimiques et les méthodes et les modèles mathématiques venant d'approches de la biologie des systèmes, la démarche maintenant notoirement désignée de « biologie de synthèse » (ou aussi « biologie synthétique ») vise la fabrication de dispositifs biologiques vivants, réalisés à façon, suivant une conception d'ingénierie.

C'est un courant de recherche qui s'est rapidement développé durant la dernière décennie avec une forte visibilité médiatique recherchée par les principaux protagonistes et de nombreux milieux scientifiques, techniques et industriels intéressés. S'affichant volontiers comme l'avenir des biotechnologies, la biologie de synthèse ambitionne la création d'applications très variées dans tous les secteurs technologiques. En premier lieu naturellement dans les domaines d'application traditionnels de biotechnologies comme l'alimentaire et la santé. Mais les champs d'application visent également la chimie fine, la fabrication de nouveaux matériaux, voir même l'exploitation de l'ADN en guise de système de stockage d'information à très long terme. La biologie de synthèse interpelle aussi les relations de l'humanité avec la nature du vivant elle-même dont une illustration symptomatique est son adoption dans des explorations artistiques avec des formes de création employant gènes, cellules et tissus. Le projet « Synthetic Aesthetics » présenté dans la figure 2.6 en est un exemple parmi de nombreux. Des artistes et des designers n'hésitent pas à se projeter comme parties prenantes pour la conception de nouvelles espèces vivantes, dont on imagine la création d'un nouveau règne synthétique, figure 2.7.

La biologie de synthèse est la possibilité d'exploiter les connaissances des systèmes biologiques ouvertes par la génomique à haut débit et sa mathématisation pour concevoir de nouveaux systèmes biologiques artificiels. Mais au-delà du « buzz » et des déclarations visionnaires quelle est sa réelle nouveauté?

d'étapes de biosynthèse et impliquant des contrôles le long du processus. On voudra par exemple aussi, plus ambitieux, construire tout un organisme vivant conçu avec juste le nombre de gènes nécessaire pour accomplir une tâche utile cible unique dans des conditions déterminées, sans autres capacités, mais avec le plus haut rendement possible pour son objectif.

La biologie synthétique a donc pour objet de manipulation les réseaux biochimiques et systèmes biologiques dans leur ensemble. Elle est alors une émanation de la biologie des systèmes dont elle tire parti des données, concepts et méthodes. Le développement de cette nouvelle biotechnologie provient des avancées technologiques de l'informatique, électronique et de la robotique dont on a déjà fait état. Mais aussi de développements technologiques spécifiques à la biologie moléculaire elle-même, démultipliant les capacités de clonage et de synthèse.

2.5.2 Les technologies propres à la biologie de synthèse

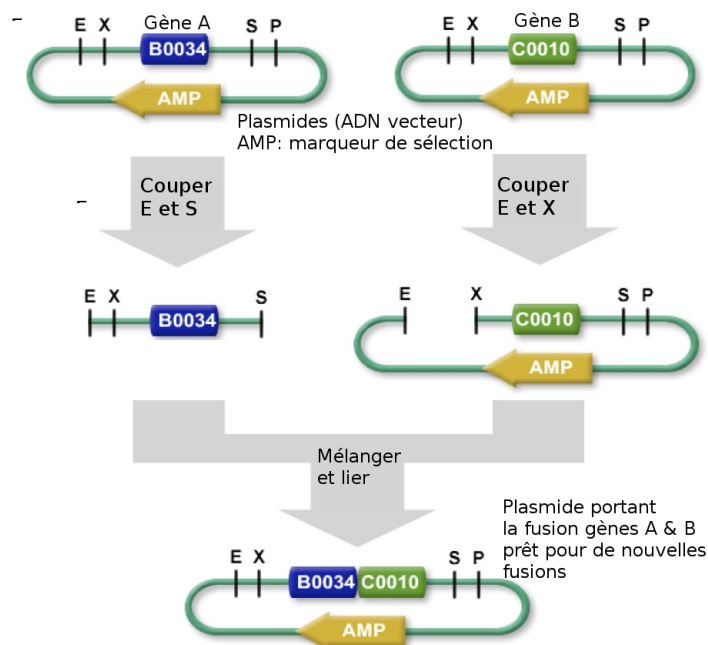


FIGURE 2.8 – **Système d'assemblage *idempotent* imaginé pour la compétition iGEM.** Les deux plasmides du haut portent deux séquences génétiques A et B (gènes) dont on veut réaliser l'assemblage. Les plasmides vecteurs utilisés ont été conçus pour porter des séquences préfixes et suffixes pour des enzymes de restrictions standardisés particuliers. Le résultat de la fusion de A et B est porté dans un plasmide vecteur de même structure que les vecteurs originaux. Le système est *idempotent* : le résultat peut être réutilisé dans un assemblage du même type avec d'autres séquences pour la fabrication d'un système plus complexe.

La biologie synthétique se construit sur le répertoire des technologies dites de l'« ADN recombinant » développées durant les années 1970. Permettent d'isoler des fragments d'ADN d'un organisme, de l'associer à d'autres fragments et de le placer dans des cellules receveuses pour le multiplier, ces méthodes ont conduit lors de leur développement à des controverses importantes sur leur utilisation. Craignant le potentiel d'effets ou d'usages dangereux d'organismes contenant de l'ADN recombinant un moratoire international sur la recherche a été décrété par la communauté scientifique lors de la conférence restée célèbre d'Asilomar sur l'ADN recombinant de 1975.

Trente années plus tard tout avait changé. La technologie. Et sa diffusion.

Ce n'est pas l'endroit pour s'étendre davantage sur les changements passionnants de diffusion, acceptation et controverses liés aux biotechnologies de l'ADN. Un sujet capital en soi que les tenants de la biologie synthétique essaient d'aborder via de nombreuses initiatives, dont le terrain artistique, en pour prévenir tout moratoire ou obstacle à son essor.

D'un point de vue technologique la mise au point durant les années 1990 de la PCR a été une étape majeure qui a permis de simplifier considérablement, systématiser et automatiser l'amplification de séquences d'ADN. Une avancée bouleversant la biologie moléculaire et capitale aussi pour toute la biologie à haut débit discutée plus haut.

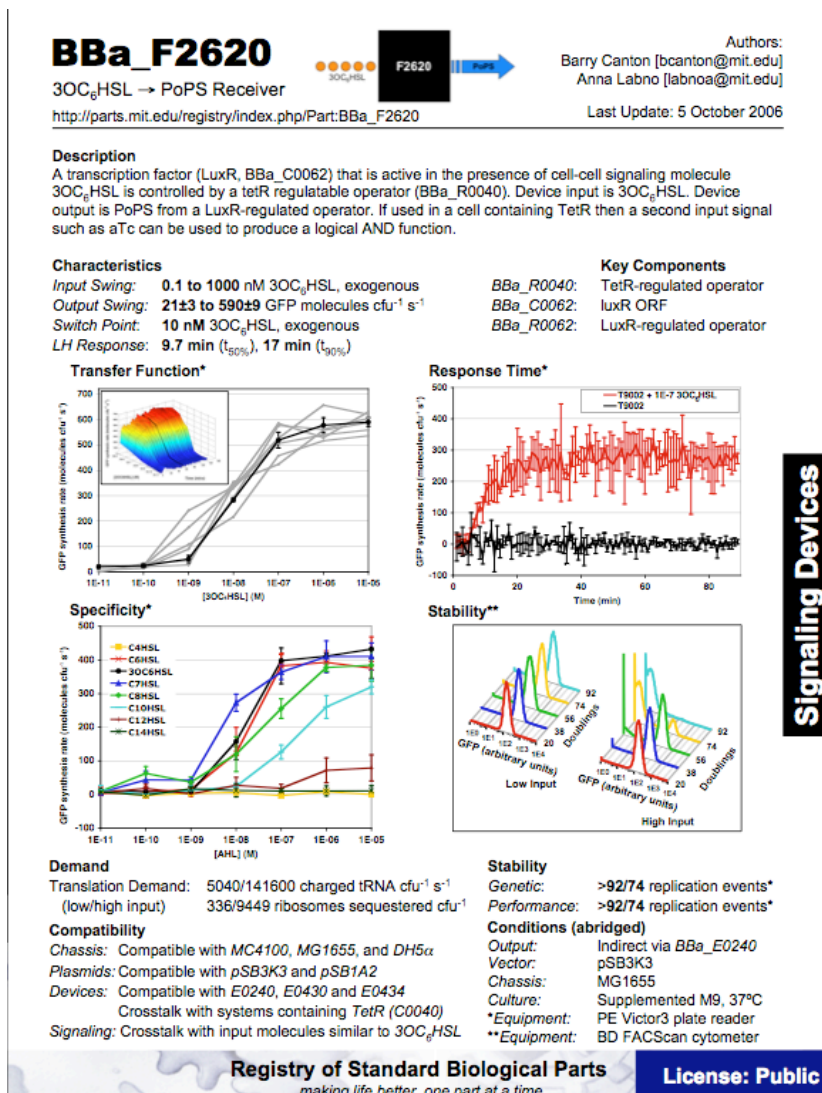


FIGURE 2.9 – Exemple d'une fiche de description d'une « biobrick » une composante biologique standardisée du « Registry of Standard Biological Parts » de la compétition iGEM. La fiche décrit la structure et les propriétés fonctionnelle du composant à l'image des fiches de description de composants électronique. Les biobricks sont portées sur des plasmides vecteurs standardisés facilitant leur assemblage dans des systèmes complexes. Chaque équipe iGEM reçoit une distribution physique de cette librairie en début de compétition.

Un autre facteur majeur a porté sur la capacité à produire de l'ADN. Depuis les années 2000 de nombreux efforts de nature différentes ont révolutionné les capacités d'assemblage de séquences d'ADN de plus en plus longs et complexes pouvant être implantées dans des cellules hôtes. Il est désormais possible et avec des coûts en diminution constante de synthétiser chimiquement des molécules d'ADN en plus longues à partir simplement de l'information de la séquences de nucléotides souhaité.

La synthèse chimique d'ADN est en train de s'imposer dans les laboratoires supplantant les méthodes de paillasse usuelles de clonage et assemblage. Les techniques ont également été développées pour insérer les fragments synthétisés de plus en plus longs dans des cellules receveuses. Celles-ci porteront un génome partiellement modifié pouvant implémenter un programme génétique, ou bien même leur génome intégralement remplacé par un génome conçu a façon. Ce

scénario correspond au programme effectivement réalisé par l'équipe de Craig Venter, qui après le séquençage et la génomique s'est résolument orienté vers les perspectives de la biologie de synthèse. Avec le Nobel Hamilton Smith, Craig Venter et leurs équipes ont achevé en 2010 la fabrication d'une cellule de levure dont tous les chromosomes avaient été synthétisés *ex informatio*, à partir de l'information, par synthèse chimique.

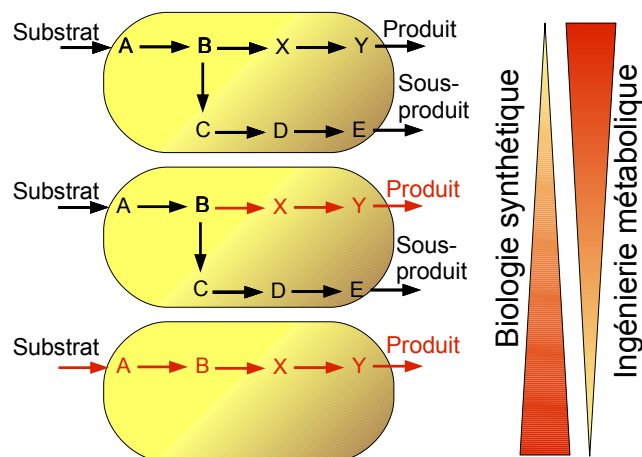


FIGURE 2.10 – **Recouvrement entre l'ingénierie métabolique et la biologie synthétique.** En haut : schéma d'un organisme possédant une voie métabolique transformant un substrat en un produit d'intérêt, le long de transformation intermédiaires A, B, X, Y et la production d'un sous-produit suivant une voie parallèle A, B, C, D, F dont il convient de minimiser la production. Milieu : l'ingénierie métabolique permet d'augmenter le rendement en optimisant les étapes X et Y de la voie de biosynthèse. Bas : la biologie synthétique vise à concevoir un organisme optimal en ayant par exemple supprimé *toute* la voie parallèle C-D-E. Source : d'après J Nielsen, JD Keasling, Nature Biotech, Août 2011

Si ces résultats restent des prouesses scientifiques encore hors de portée de laboratoires usuels, depuis 2003 déjà les outils et ressources génétiques sont suffisamment accessibles pour permettre à de jeunes étudiants du monde entier de jouer chaque été à la création d'organismes vivants génétiquement modifiés. C'est la compétition iGEM « international Gene Engineered Machines Competition » initiée par un certain nombre d'ingénieurs et informaticiens visionnaires venus à la biologie Tom Knight, Randy Rettberg et Drew Endy. Ce groupe de chercheurs au MIT allait donner notoriété et un élan presque fondateur à la biologie synthétique en promouvant l'application dans les biotechnologies de processus de standardisation suivant les pratiques des autres sciences d'ingénierie, notamment électronique et informatique. Suivant leur vision la biologie en tant que domaine basé sur de l'information et des processus de transformation, devait permettre la mise au point d'outils et de pièces standardisés de fabrication, facilitant la conception de systèmes biologiques complexes et leur réalisation matérielle. La figure 2.8 montre le système de construction génétique proposé dans le cadre de iGEM pour assembler de manière standardisée dans une bactérie receveuse des circuits génétiques consistants de plusieurs gènes capables d'interactions. Le processus exploite des techniques usuelles de biologie moléculaire d'insertion d'ADN dans des cellules, mais mises en œuvre par les ingénieurs à l'origine d'iGEM d'une manière standardisée pour que chaque construction génétique réalisée puisse être réutilisée par le même protocole pour des assemblages plus complexes. La figure 2.9 montre un exemple de pièce génétique de base de construction, nommées *biobricks*, disponibles sur le vecteur de construction standardisé et caractérisées comme des composants standards de circuit à l'instar de composants électronique.

La biologie de synthèse correspond à cet effort de standardisation de pièces et méthodes biologiques pour une conception rationnelle et une fabrication facilitée de systèmes biologiques. Beaucoup d'objectifs sont analogues à ceux du génie génétique et l'ingénierie métabolique issus des

années 1970. Les stratégies sont différentes. La biologie de synthèse consiste avant tout au développement de standards et processus facilitant les constructions génétiques complexes. Avec les facilités de production la biologie de synthèse vise à concevoir toute une voie de biosynthèse optimisée. La figure 2.10 illustre cette évolution vers une transformation de tout un système. Son objectif est la conception et fabrication de l'objet de la biologie des systèmes : l'ensemble de gènes et protéines en interaction qui constituent les réseaux biochimiques des cellules. Il est temps de donner quelques précisions sur la notion de réseau qui apparaît si centrale dans ces développements.

2.6 Les réseaux complexes

En biologie, comme dans de nombreux autres domaines le paradigme est aux réseaux. Dans tous les domaines scientifiques, les outils expérimentaux et les ressources informatiques actuelles à haut débit remplissent des bases de données, parfois nommés entrepôts de données, tant les volumes et la diversité des données produites sont grands. Ces données ne sont rien sans leur analyse pour en extraire de l'information pertinente. La fouille de données (datamining) est l'activité consistant à rechercher de l'information utile dans les stocks de données brutes faisant appel à un florilège d'outils informatiques avec des méthodes statistiques, mathématiques et d'intelligence artificielle.

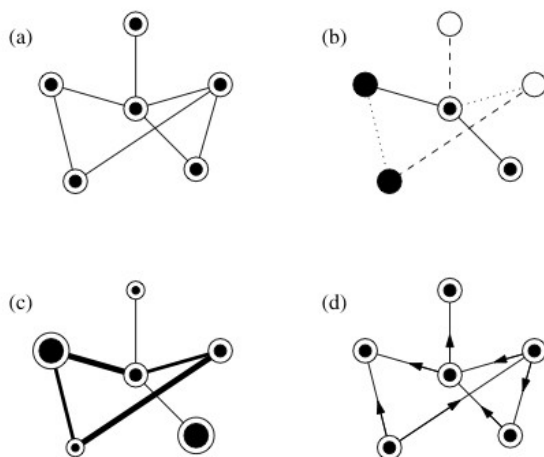


FIGURE 2.11 – Variétés de nature de graphes complexes. Des grandes variétés possibles de nature des graphes. a) Nœuds identiques. b) Nœuds et liens de types différents. c) Nœuds et liens avec des poids différents. d) Liens dirigés. Voir aussi l'encadré 2.2.

Dans ce contexte d'exploration de données multiples le concept de réseau complexe est central. Le réseau est l'ensemble des relations entre les éléments d'une base de données. Les réseaux sont dits complexes car ils sont constitués d'un grand nombre d'éléments, appelées nœuds ou aussi vertex, reliés entre eux par des liens en nombre et de nature pas nécessairement uniforme (figure 2.11). Suivant les cas d'application les connexions peuvent être uni- ou bi-directionnelles, et les nœuds peuvent avoir chacun des nombres différents de liens, le cas échéant des nombres différents de relations entrantes ou sortantes, voir l'encadré 2.2. Ces réseaux sont donc des structures de grandes tailles, régulières ou non, dont les propriétés affectent le fonctionnement de l'ensemble.

La représentation du contenu d'une base de données sous la forme de structure mathématique d'un réseau permet aux scientifiques et ingénieurs de rechercher des propriétés sur les systèmes à l'origine. Soit des caractéristiques globales, par exemple sur l'interdépendance de chaque partie du système, ou bien l'existence de parties quasi-indépendantes. Ou bien des caractéristiques particulières à certains éléments, comme des nœuds auxquels les relations avec le reste du réseau affectent des caractéristiques spéciales : par exemple, tout simplement, des nœuds très connectés peuvent davantage affecter l'ensemble que des nœuds plus isolés. Les concepts des réseaux complexes sont appliqués pour étudier des réseaux artificiels comme le réseau reliant

serveurs et clients d'internet, le réseau de hyperliens entre pages web, les réseaux de distribution électriques ou de fluides.

Réseaux complexes en biologie En biologie moléculaire et génétique le concept de réseau n'a jamais été une nouveauté et apparaissait naturellement dès qu'on s'éloigne du niveau de chaque structure moléculaire individuelle pour regarder les interactions ou certains organes. La biologie connaît en effet des réseaux et en a représenté des dessins et cartes bien avant que l'étude de leur topologie et propriétés d'ensemble devienne un sujet à la mode. Certains systèmes macroscopiques majeurs sont d'ailleurs matériellement des réseaux comme les systèmes de transport chez la plupart des organismes multi-cellulaires. Pour assurer l'alimentation, l'oxygénation et l'évacuation des résidus à toutes les cellules d'organismes de grande taille des systèmes sophistiqués de distribution et circulation de matière ont du se développer et former les réseaux vasculaires, lymphatiques, respiratoire des bronches, phloémien chez les plantes. (On pourrait aussi citer les lignes de forces d'acupuncture dans la pharmacopée chinoise. . .) Et pour coordonner les informations sur l'environnement, les traiter et coordonner la réaction de l'ensemble s'est développé un système nerveux ramifié dans tout de l'organisme, ainsi qu'un processeur central de calcul, le cerveau, figure 2.12 . Réseaux de transport et réseau nerveux, ont une réalité tangible, mécanique. Leur développement et propriétés dépendent aussi des caractéristiques physiques, forces, élasticité, viscosité des cellules qui les composent matériellement, et de leurs interactions avec l'environnement où les réseaux se développent.

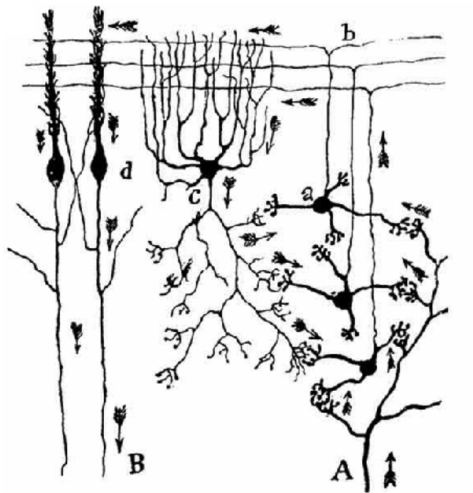


FIGURE 2.12 – **Dessin historique de réseau neuronal.** Dessin par Santiago Ramon y Casal d'un réseau de neurones observé au microscope et révélé grâce à la technique de coloration développée par Camillo Golgi mettant en évidence la nature discrète du réseau neuronal composé de cellules différentes, voir section 2.6.1. De gauche à droite la figure montre trois types de neurones différents. Trois neurones pyramidaux sont représentés à droite avec les points de contacts entre les neurones distincts, lieu d'établissement des synapses.

D'autres réseaux biologiques sont plus abstraits. Ils représentent les interactions entre objets biologiques. Reprenant les termes du biologiste des plantes Geert Potters [POTTERS, 2010] :

« Rien en biologie n'agit seul : chaque chose agit en conjonction, opposition et synergie avec d'autres éléments. Les gènes régulent leurs activités les uns les autres, les produits des gènes joignent leurs forces ou bien s'inhibent les uns les autres, les cellules communiquent, les métabolites bougent d'un tissu à un autre et tout est interconnecté. ».

Au niveau le plus macroscopique, d'écosystèmes naturels, il s'agit par exemple des interactions entre espèces vivantes, notamment des liens alimentaires représentés depuis longtemps en écologie par des dessins et formant les réseaux dits « trophique ». Au niveau microscopique des molécules ce sont toutes les interactions possibles entre molécules qui rassemblent dans un réseau les informations du comportement moléculaire d'ensemble, objet d'étude de la biologie des systèmes.

En génétique et biologie moléculaires, champ de la biologie des systèmes, les réseaux ne sont pas matériels, ils représentent les interactions possibles entre gènes différents, protéines et biomolécules. Ces réseaux d'interactions biochimiques seront discutés plus particulièrement dans le

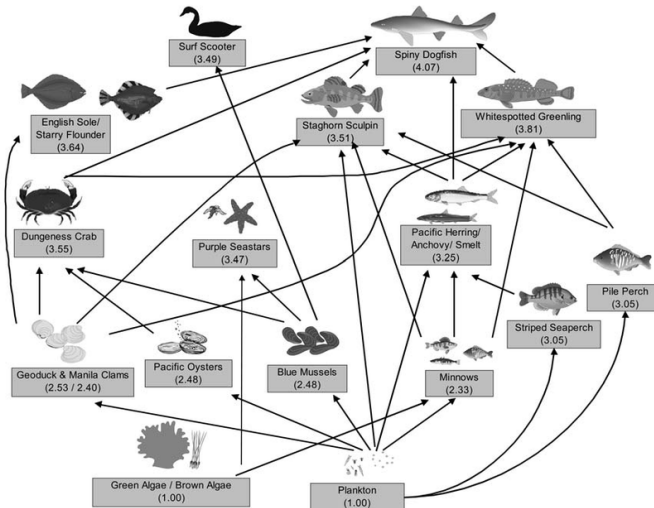


FIGURE 2.13 – Diagramme conceptuel simplifié des interactions alimentaires dans la baie False Creek. [GOBAS et collab., 2003]

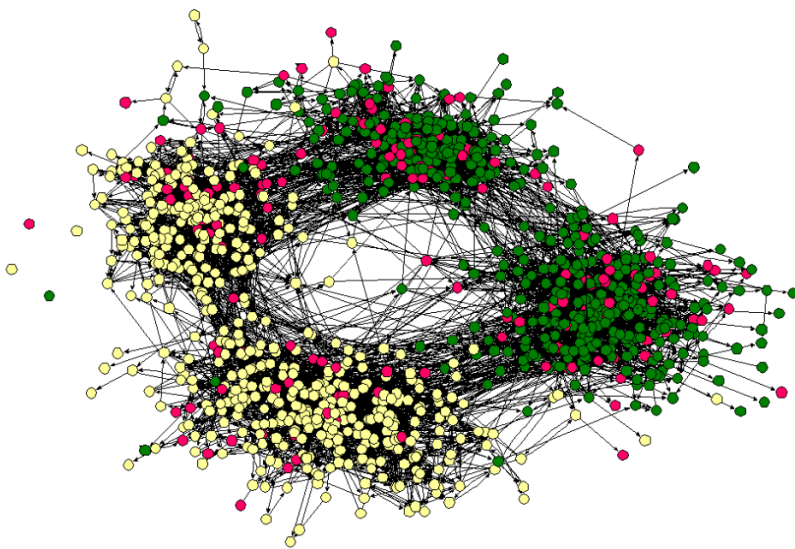


FIGURE 2.14 – Réseau d'amitiés dans un lycée américain. Les nœuds du réseau correspondent aux élèves et sont colorés par couleur de peau. Jaune : blancs; vert : noirs; rouge : d'autres types [MOODY, 2001].

chapitre 4.

L'analyse de la topologie et de la morphologie des réseaux est nécessaire pour comprendre les déterminants des activités qui s'y déroulent. De nombreux réseaux complexes mettent en relation des objets sans notion de localisation spatiale. Seule les propriétés des schémas de connections jouent. Dans ce cas on peut aussi utiliser le terme mathématique de *graphe* pour désigner le réseau. Le domaine mathématique pour l'étude de ces objets est la « théorie des graphes » inspiré par les travaux fondamentaux de la fin des années 50 du mathématicien hongrois Paul Erdos et qui fournit les méthodes pour analyser et décrire les réseaux complexes. Ces contributions des mathématiques ont trouvé des applications considérables avec la disponibilité de réseaux complexes de grande taille au courant des années 1990 et le développement d'une véritable « science des réseaux ».

Ces travaux mathématiques ont conduit à un corpus de concepts et méthodes pour caractériser les propriétés des réseaux qui sont maintenant appliqués sur les grands réseaux complexes de nos jours avec de nombreux outils informatiques devenus disponibles. On peut se référer pour des détails à la revue classique extensive par M.E.J. Newman [NEWMAN, 2003].

Dans tous les différents domaines où on les retrouve, les fonctions portées par les réseaux complexes peuvent dépendre de nombreux facteurs :

- les propriétés spécifiques des interactions entre les éléments; les interactions peuvent être de nature différentes dans un réseau complexe.
- les propriétés propres à chaque élément, indépendamment de ses liens avec d'autres; les nœuds d'un réseau ne sont pas de nature identique.
- la structure du graphe de connections, c'est-à-dire le schéma de l'ensemble des liens en nœud indépendamment de leur nature, ce qui correspond à la topologie du réseau.
- l'existence ou non d'une métrique sur le réseau; c'est-à-dire si les nœuds sont placés dans un espace avec une notion de distance entre eux, indépendamment de la topologie des connexions. On parle parfois de morphologie d'un réseau pour désigner l'intégralité de sa forme, topologie des connections et répartition spatiale des nœuds.
- la possibilité de liens de poids différents dans le cadre de réseaux pondérés (« weighed networks ») qui rendent compte d'intensité différentes des relations; un cas important notamment pour le réseau neuronal.

Encadré 2.2 – De nombreux types de réseaux complexes.

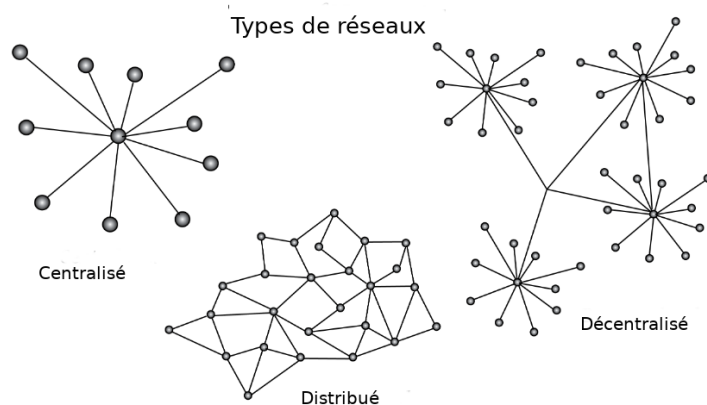


FIGURE 2.15 – **Réseaux centralisés, décentralisés et distribués.** Dans les réseaux centralisés, un nœud ou un sous ensemble de nœuds sont connectés à un grand nombre d'autres nœuds faiblement interconnectés entre eux. Dans les réseaux décentralisés il existe des sous-réseaux non, ou faiblement interconnectés sans sous-ensemble central. Dans les réseaux distribués les nœuds ont des degrés à peu près identiques.

Descripteurs topologiques

De nombreuses mesures sur les réseaux peuvent être définies et calculées qui permettent de les caractériser les uns par rapport aux autres et en particulier par rapport à des réseaux dits « aléatoires ». Différentes caractéristiques de réseaux, les principales sont présentées dans l'encadré 2.3, et différents modèles théoriques de réseau construits par le hasard servent de références pour décrire les propriétés de réseaux réels.

La première propriété caractéristique d'un réseau est sa *distribution du nombre de connexions* par nœud, la distribution des degrés 2.3. C'est la propriété immédiate, la plus simple qui les distingue des *réseaux réguliers*, les plus simples avec le même nombre de connections pour chaque nœud. Les réseaux complexes naturels et technologiques ont des distributions de degrés non homogènes, ni mêmes centrées sur une valeur de degré caractéristique. Dans ces réseaux la plupart des nœuds participent uniquement à quelques interactions, alors que d'autres sont très fortement connectés.

Comme visible dans la figure 2.16 de nombreux réseaux de grande taille réels semblent avoir des distributions du nombre de leurs connexions par nœud suivant des lois de puissance avec

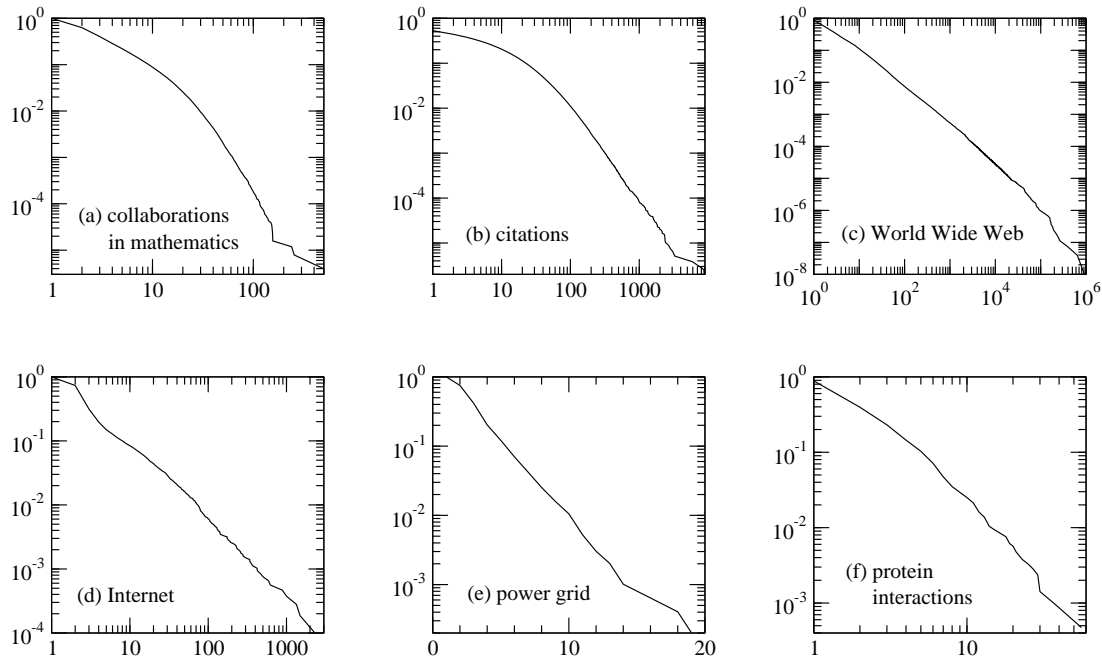


FIGURE 2.16 – **Exemples de distributions de connectivité de réseaux complexes réels.** Figure 6 de la revue de M.E.J. Newman sur les réseaux complexes [NEWMAN, 2003] illustrant les différentes distributions de connectivité de six réseaux complexes réels. L'axe horizontal représente le degré k des nœuds (le degré entrant si le réseau est dirigé) et l'axe vertical la distribution de probabilité cumulée des degrés $P_k = \sum_{k'=k}^{\infty} p_k$ avec p_k la fraction de nœud de degré k . P_k correspond à la fraction des nœuds de degré supérieur ou égal à k . Les exemples sont les réseaux : a) des collaborations entre mathématiciens, chaque nœud est un auteur et les liens est l'existence d'au moins une publication commune [GROSSMAN et ION, 1995]; (b) des citations de tous les articles entre 1981 et 1997 dans le catalogue de l'Institut for Scientific Information, éditeur du « citation index », chaque nœud est une publication avec les liens pointant sur ses articles cités [REDNER, 1998]; (c) des hyperliens www internet en 1999 entre 300 million de pages [BRODER et collab., 2000]; (d) des connections internet entre serveurs en avril 1999 [CHEN et collab., 2002]; (e) de distribution électrique sur la côte Ouest des États-Unis [WATTS et STROGATZ, 1998]; (f) des interactions entre protéines du métabolisme de la levure *S. cerevisiae* [JEONG et collab., 2003]. Les réseaux (c), (d) et (f) montrent des distributions en loi de puissance comme indiqué par la ligne approximativement droite dans le graphe double logarithmique; (b) possède une queue en loi de puissance avec un comportement différent pour les petits degrés; (e) a une distribution exponentielle des degrés, le graphe est log-linéaire dans ce cas; (a) est une distribution en loi de puissance tronquée ou bien suit deux régimes en lois de puissance avec des exposants différents.

certaines valeurs d'exposant $p_k \sim k^{-\alpha}$, avec α l'exposant de la distribution. D'autres distributions ont des queues exponentielles $p_k \sim e^{-k/\kappa}$, où κ est alors un degré caractéristique. Pour d'autres réseaux les degrés de distribution peuvent-être plus compliqués notamment si les nœud et liens sont de nature hétérogène. Les distributions en loi de puissance expriment la possibilité d'invariances d'échelle. C'est-à-dire d'absence d'une valeur moyenne significative si $\alpha > 1$. En effet dans ce cas la valeur moyenne $\int_1^{\infty} k \cdot p_k dk$ diverge. Or la plupart des exposants mesurés ont des valeurs proches de 2 [CLAUSET et collab., 2009].

Ces lois de puissance laissent penser aux phénomènes dits « critiques » de la physique statistique qui correspondent à des corrélations infinies dans le système, que l'on caractérise aussi comme invariants d'échelle⁵. Cette connexion possible avec des propriétés physiques particulières et l'observation de nombreux réseaux semblant caractérisés par une loi de puissance a conduit à de nombreuses études et spéculations sur l'universalité de cette propriété [ALBERT et BARABÁSI, 2002; CALDARELLI, 2007; DOROGOVTSSEV et MENDES, 2002; STROGATZ, 2001]. Les distributions de degrés ont également été étudiées pour les réseaux biochimiques ce qui fera l'objet d'une

5. La loi de puissance est la seule fonction $f(x)$ telle qu'un changement d'échelle d'un certain facteur a est proportionnelle à la même fonction : $f(ax) = bf(x)$.

discussion particulière dans la section 4.2.

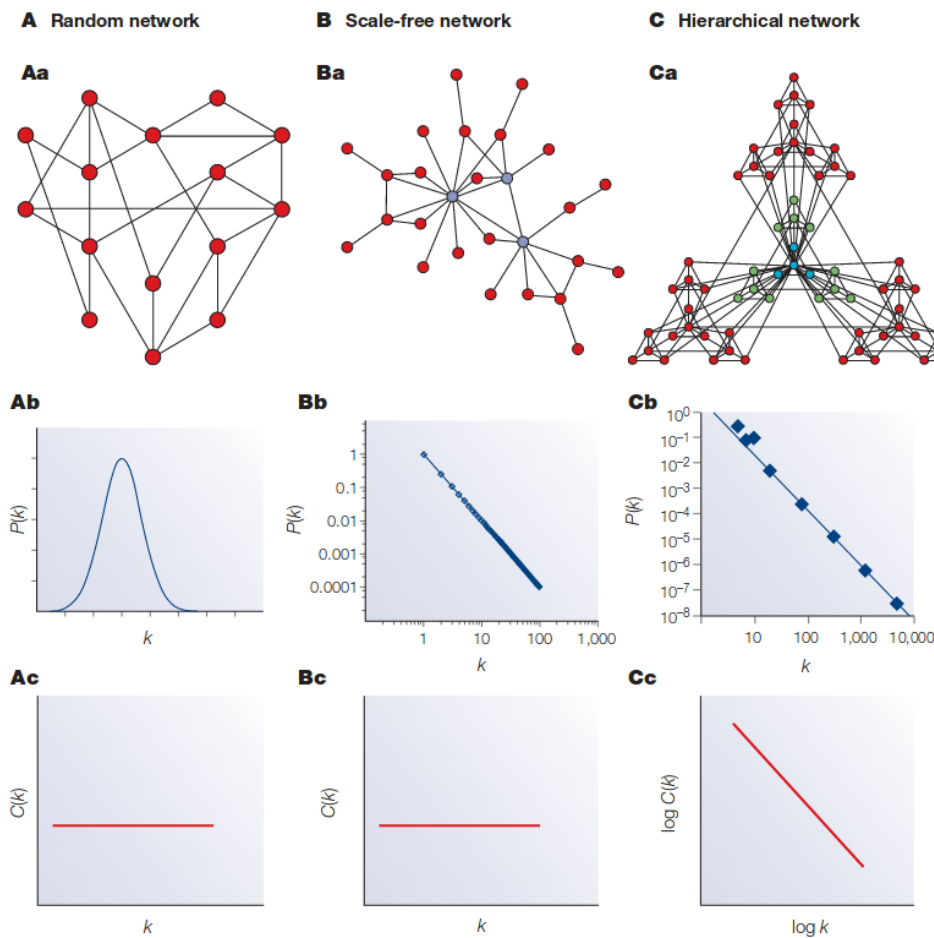


FIGURE 2.17 – **Exemples de trois modèles de réseaux aléatoires.** A) aléatoire d’Erdős-Rényi : allure typique du réseau (Aa) ; la distribution des degrés (Ab) est une distribution de Poisson et la fonction $C(k)$ du coefficient de clustering de nœuds de degré k est indépendante de k (Ac). B) en loi de puissance suivant le modèle de Barabási–Albert : allure (Ba) ; la distribution des degrés est une loi de puissance $P(k) \sim k^{-3}$ (Bb) ; le modèle de Barabási–Albert ne crée pas de modularité hiérarchique, la fonction $C(k)$ est uniforme. C) modèle en loi de puissance hiérarchique de RAVASZ et BARABÁSI [2003] ; RAVASZ et collab. [2002] où le réseau se développe itérativement par multiplication d’un motif initial central : allure typique (Ca) ; loi de puissance $P(k) \sim k^{-2,26}$ généré par ce modèle dont l’exposant dépend du nombre de nœud du motif initial et du nombre de multiplications à chaque étape ; fonction de clustering $C(k) \sim k^{-1}$ manifestant une hiérarchie à toutes les échelles. Source : [BARABÁSI et OLTVAI, 2004]

Les distributions de connectivité ne sont pas les seules caractéristiques d’un réseau. Une autre mesure utile est le *diamètre* du réseau, à savoir la moyenne des distances entre deux points quelconques, où la distance correspond au nombre minimal de liens successifs connectant les deux nœuds. Watts et Strogatz ont publié en 1998 [WATTS et STROGATZ, 1998] l’observation très surprenante que pour la plupart des réseaux complexes naturels et technologiques connus ce rayon augmente seulement logarithmiquement avec la taille du réseau plutôt qu’une augmentation linéaire comme dans des réseaux réguliers avec une connectivité moyenne bien définie.

Connue sous le nom de propriété, ou phénomène *petit monde* (« small-world ») cette propriété est interprétée en termes d’efficacité de transfert d’information dans le réseau avec un nombre limité d’intermédiaires. Elle avait été mise en évidence et baptisée ainsi en sociologie par Robert Milgram en 1969. L’expérience visait à sonder la distribution des chemins dans le réseau de connaissances personnelles en demandant aux participants de transférer une lettre à une personne directement connue, essayant de la faire parvenir à un destinataire individuel particulier. La plupart de lettres ont été perdues, mais parvenues a destination étaient passées au travers de

6 personnes en moyenne. Cette propriété petit monde a été par la suite mise en évidence dans de nombreux réseaux complexes de nature très différente [NEWMAN et WATTS, 1999; WATTS, 2003; WATTS et STROGATZ, 1998].

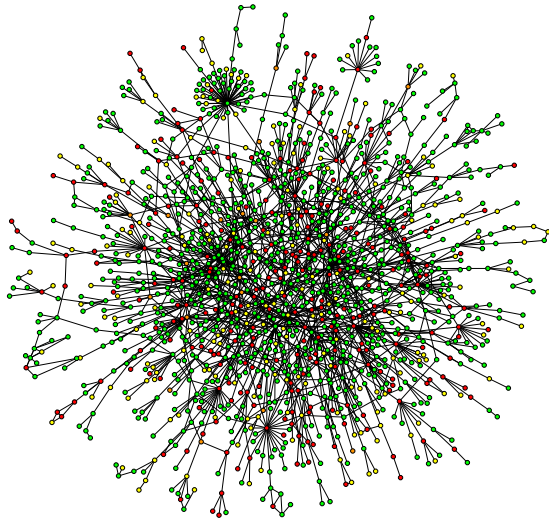


FIGURE 2.18 – **Carte des interactions protéine-protéine de la levure.** Caractéristique du protéome : la carte des interactions protéine-protéine de la levure. Est représenté le regroupement connecté le plus grand qui contient ~ 78 de toutes les protéines. La couleur de chaque nœud indique l'effet phénotypique de la suppression de la protéine correspondante (rouge : létal; vert : non létal; orange croissance lente; jaune : inconnu). Source : [JEONG et collab., 2001]

La *coefficient de clustering* exprime quant à lui le degré d'interconnexion de nœuds proches c'est-à-dire le degré de clustering de trois nœuds interconnectés. Cette propriété dite de *transitivité* caractérise une forte structuration locale nécessaire dans le cas d'une organisation modulaire des réseaux nécessaire à leur fonction et développement [HARTWELL et collab., 1999; KITANO, 2002]. Le coefficient de clustering moyen caractérise la tendance générale du réseau à former des groupes. Le coefficient de clustering $C(k)$ en fonction k , qui mesure cette grandeur pour les nœuds de degré donné donne une indication de l'*organisation hiérarchique* du réseau. De nombreux réseaux réels ont une dépendance $C(k) \sim k^{-1}$ [RAVASZ et BARABÁSI, 2003; RAVASZ et collab., 2002]. Une architecture hiérarchique implique l'existence de sous-structures de nœuds fortement connectés entre eux avec des liens entre différents groupements impliquant seulement quelques nœuds charnières.

Une question liée directement à la structuration globale des réseaux, distribution des degrés, transitivité, modularité est leur robustesse à des dysfonctionnements. La **résilience** des réseaux à la suppression de nœud et de liens est une préoccupation majeure de nombreuses études. Les fonctions des réseaux dépendent de leur connectivité et de l'existence de chemins entre différents nœuds. Si des nœuds ou des liens intermédiaires disparaissent les chemins sont perturbés, ils peuvent être rallongés, ou bien disparaître et déconnecter différentes parties. Le niveau de résilience des réseaux dépend de leur structure et de la nature des perturbations. En effet, nœuds ou liens peuvent par exemple disparaître au hasard, ou bien certains types de nœud, comme ceux les plus connectés, être plus particulièrement la cible d'attaques. La question de la résilience se pose donc naturellement chaque fois que la structure d'un réseau est interprétée dans son contexte.

De nombreux autres descripteurs existent pour caractériser les réseaux complexes, on se rapportera à la revue très claire de M.E.J. Newman [NEWMAN, 2003] et de BOCCALETTI et collab. [2006]. Quelques descripteurs supplémentaires sont brièvement décrits dans l'encadré 2.3. Certains de ces indicateurs sont très utiles dans des perspectives d'applications et visent non seulement à décrire le réseau mais aussi à identifier en son sein des éléments particuliers comme des nœuds ou des chemins entre nœuds avec des positions centrales stratégiques ou des regroupements qui se distinguent de l'ensemble par des propriétés spéciales. Interviennent à cet effet différents indicateurs dits de *centralité*. Enfin les *motifs*, permettent de rechercher des schémas de réseaux particuliers identifiés par leur abondance dans le réseau étudié par rapport à un réseau aléatoire pris comme référence. On les retrouvera dans la section 5.2.3 du chapitre 5 dans le cadre de mon étude du réseau mixte transcriptionnel et d'interactions de protéines de la levure. Plusieurs de ces descripteurs sont difficiles à calculer sur des réseaux de grande taille exigeant des

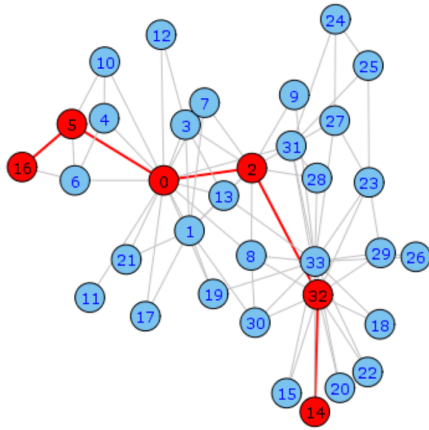


FIGURE 2.19 – Illustration du diamètre d'un réseau complexe : exemple du diamètre dans le réseau des partenaires au sein d'un club de Karaté à New York [ZACHARY, 1977]

calculs numériques et besoins de mémoire importants. Un champ important de la recherche porte encore sur le développement d'algorithmes informatiques efficaces pour ces applications.

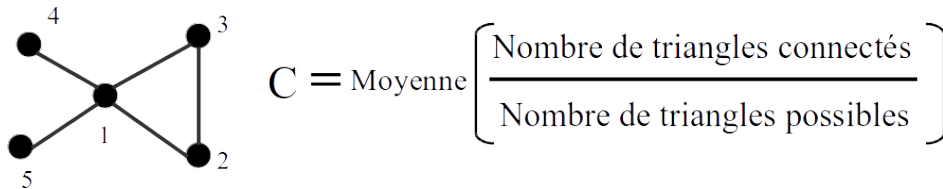


FIGURE 2.20 – **Coefficient de clustering (coefficient d'agglomération)**. Ce coefficient caractérise la fraction des plus proches voisins d'un nœud interconnectés entre eux. Il se calcule en comptant le nombre de triangles connectés sur le nombre total de triangles possibles.

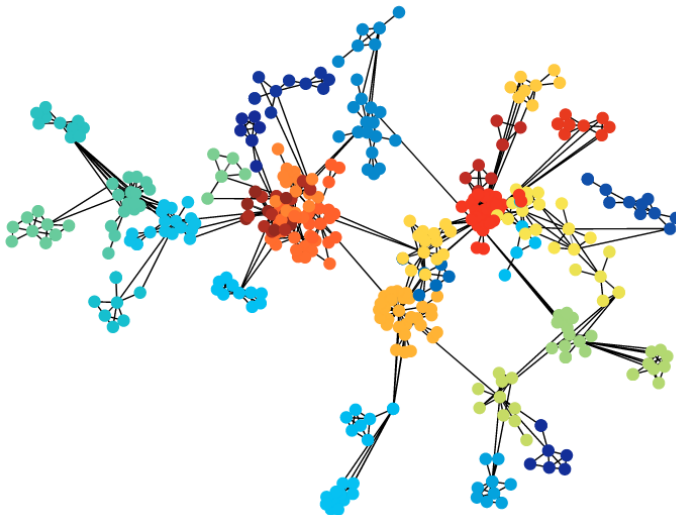


FIGURE 2.21 – **Exemple de communautés sur un réseau**. Communautés de co-auteurs scientifiques dans le domaine de la recherche sur les réseaux (!). Les différentes couleurs représentent des auteurs ayant l'habitude de publier ensemble [PORTER et collab., 2009].

Réseaux aléatoires

La théorie des graphes a introduit un certain nombre de modèles de réseaux théoriques construits par des formes différentes de hasard et aboutissant à des distributions de connectivité différentes. Comme aucune considération technologique ou naturelle contraint les réseaux aléatoires, ils constituent des références idéales contre lesquelles comparer les réseaux issus de la réalité. La plupart des réseaux réels semblent avoir des organisations qui ne sont ni complètement régulières ni complètement construits par le hasard. Les différents modèles de réseaux aléatoires servent ainsi de référence pour des comparaisons.

<p>Degré (connectivité) nombre de ses liaisons d'un nœud avec d'autres ; si les liens sont dirigés on distingue les degrés entrants et sortants.</p> <p>Distribution de degré (de connectivité) distribution sur l'ensemble du réseau du nombre de nœud possédant un degré donné. La fonction de la distribution de connectivité permet de quantifier des types de graphes différents.</p> <p>Chemin, chemin le plus court, chemin moyen séquence de nœud adjacents successifs qui relie deux nœuds donnés. Le nombre de liens d'un chemin est sa longueur. Comme de nombreux chemins peuvent relier deux nœuds existent aussi les notions de chemin le plus court et chemin de longueur moyenne. Le <i>diamètre</i> d'un réseau correspond au plus long des chemins les plus courts entre deux nœud, figure 2.19.</p> <p>Coefficient de clustering (coefficient d'agglomération)⁶ coefficient qui permet de caractériser le niveau d'interconnection moyen entre les voisins d'un nœud (figure 2.20), on parle aussi de <i>transitivité</i>; quand sa valeur élevée, si un nœud A est connecté à deux nœuds B et C alors ceux-ci ont une probabilité élevée de l'être aussi. Une notion associée est celle de <i>clique</i>, correspondant à un sous-réseau dans lequel toutes les paires de nœud sont inter-connectées.</p> <p>Modularité, communauté descripteurs qui quantifie à quel point le réseau peut-être divisé en des sous ensembles distincts, figure 2.21.</p> <p>Centralité mesure locale (pour chaque nœud) qui correspond au nombre de chemins les plus courts qui contiennent un nœud donné. Les nœuds avec une centralité élevée participent à de nombreux chemins et probablement donc à de nombreux processus.</p> <p>Assortativité coefficient de corrélation des connections entre nœud de même degré. Dans les réseaux sociaux les personnes très connectées ont des connections directes. Dans des réseaux biologiques et technologiques les nœud très connectés sont rarement liés les uns aux autres.</p> <p>Efficacité caractéristique du temps ou du nombre d'étapes nécessaires pour relier deux nœuds par une marche au hasard sur le réseau.</p> <p>Motifs schémas de connectivité dans un réseau dont la statistique est plus grande que leur abondance dans un réseau aléatoire de même distribution de degrés. La question des motifs dans les réseaux biochimiques sera abordées dans mon étude de la section 5.2.2 du chapitre 5.</p>
--

Encadré 2.3 – Descripteurs topologiques principaux. De nombreuses variantes et des descripteurs supplémentaires existent et sont par exemple résumés avec des codes de calcul dans [RUBINOV et SPORNS, 2010].

« **Le** » réseau « **aléatoire** ». Dans les années 1950 Rapoport [SOLOMONOFF et RAPOPORT, 1951] puis Paul Erdős et Alfréd Rényi [ERDOS et RENYI, 1959] ont introduit le modèle aléatoire de référence, souvent connu du nom de modèle d'Erdős-Rényi. Dans celui-ci, chaque élément d'un ensemble donné de N nœuds est connecté aléatoirement, indépendamment des autres avec une probabilité p fixe. Une des propriétés importantes de ce réseau est son fractionnement ou non en îlots dissociés, qui dépend du produit $N \times p$. Pour de grands réseaux la distribution de connectivité tend vers une distribution de Poisson centrée sur une valeur de connectivité moyenne. Le modèle est pour cela aussi parfois appelé réseau aléatoire de Poisson, bien que la plupart du temps simplement nommé *réseau aléatoire*⁷. De nombreuses généralisations du modèle ont été introduites depuis les années 50 pour admettre notamment des distributions de connectivité autres que Poisson et sont notamment discutés dans la revue de M.E.J. Newmann [NEWMAN, 2003].

Le modèle petit monde Ce modèle introduit à la fin des années 1990 par Steve Strogatz et Duncan Watts permet [J.WATTS, 2000; STROGATZ, 2001; WATTS et STROGATZ, 1998] de faire le lien entre le modèle aléatoire ci-dessus et un modèle régulier de connectivité fixe. Ce dernier cas cor-

7. Même si communément simplement appelé réseau aléatoire, il est utile de remarquer que le tirage au hasard uniforme de nœuds de ce modèle est néanmoins une forme stochastique particulière, quoique la plus simple.



FIGURE 2.22 – **Modèle de réseau de Watts & Strogatz.** Transition entre un réseau régulier à gauche, vers un réseau aléatoire à droite. Le réseau intermédiaire au centre est obtenu en réorganisant au hasard une fraction p des connections du réseau régulier. Le réseau régulier est fortement interconnecté localement (clustering élevée) mais un diamètre élevé. Le réseau aléatoire a un clustering faible mais un petit diamètre. Le réseau intermédiaire possède à la fois un petit diamètre et un clustering élevé : on dit qu'il a une caractéristique *petit monde* [WATTS et STROGATZ, 1998]

respond à la topologie de réseaux spatiaux réguliers avec des nœuds connectés selon les mailles d'une grille. Dans de nombreux cas on comprend que la localisation géographique des nœuds influence leur connectivité, privilégiant les liens avec leurs plus proches voisins. Le modèle petit monde part ainsi de la topologie des nœuds d'une grille régulière. Ensuite chaque lien est modifié avec une certaine probabilité pour déplacer une des extrémités sur un autre lien tiré au hasard dans le réseau et donc potentiellement aussi très loin, figure 2.22. Des raccourcis sont ainsi créés permettant de relier directement des parties éloignées de la grille. Plusieurs variantes ont été depuis introduites permettant notamment de simplifier l'analyse théorique.

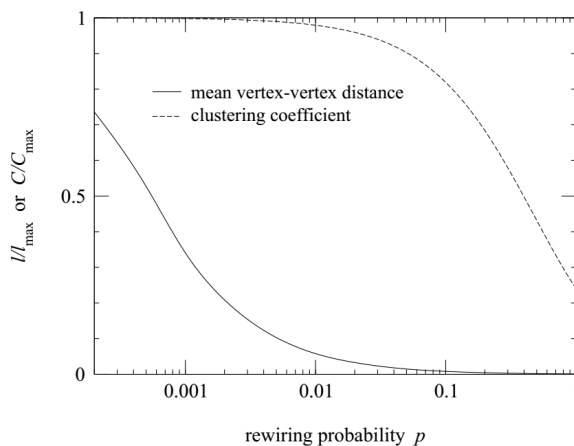


FIGURE 2.23 – **Clustering moyen et diamètre dans le modèle de petit monde de Watts & Strogatz.** Le coefficient de clustering moyen $\langle C \rangle$ et le diamètre l dans de réseaux du modèle petit monde de Watts et Strogatz en fonction de la probabilité p de réorientation de lien. Les coefficients ont été normalisés par leur valeur maximale dans le modèle avec $p = 0$. Entre les deux extrêmes, réseau régulier $p = 0$ et réseaux aléatoire $p = 1$ se trouve une région dans laquelle le clustering est à la fois élevé et le diamètre petit (caractéristique de la propriété petit-monde). Source : [WATTS et STROGATZ, 1998]

La contribution majeure de ces modèles qui organisent une transition continue entre l'ordre et le désordre est de rendre compte de la combinaison dans les réseaux réels de propriété petit monde (d'où le nom du modèle) et d'un fort niveau de transitivité (clustering), figure 2.23. Les deux cas extrêmes du modèle, complètement régulier $p = 0$ et complètement aléatoire $p = 1$ ont des caractéristiques qui ne se retrouvent pas dans les réseaux complexes. Le premier est fortement transitif (haut clustering), mais son diamètre est très grand (pas du tout « petit monde »). Le second n'est ni transitif ni petit-monde. Une situation intermédiaire mêlant ordre et désordre est davantage compatible avec les situations observées des réseaux complexes.

Modèles de croissance de réseaux La structure des réseaux complexes réels est généralement le résultat d'un processus de construction progressif se déroulant dans le temps. Les grands réseaux technologiques comme le réseau de distribution électrique, de gaz, de connections téléphoniques, internet sont le résultat d'une histoire de développement se déroulant sur plusieurs générations.

Les liens d'amitié ou de relations professionnelles des réseaux sociaux s'établissent progressivement au cours de la vie. Et enfin, les réseaux biologiques sont tous le fruit de l'histoire évolutive.

De nombreux modèles ont alors été proposés pour décrire des processus de croissance de réseau et conduisant à différentes topologies pouvant reproduire ou servir de comparaison avec les structures réelles. La plupart des modèles proposés visent à expliquer les nombreuses observations de réseaux avec des distributions de degré en loi de puissance.

Dans ces modèles de nouveaux nœuds sont introduits progressivement et connectés au réseau suivant différents types de règles d'attachement. Dans la récente vague d'intérêt pour les réseaux complexes ces modèles ont été popularisés notamment à partir de 1999 par Barabási et Albert par leur article dans Science [BARABÁSI et ALBERT, 1999] selon l'idée d'un processus d'« attachement préférentiel ». Un principe qui se retrouvait déjà en 1976 chez D. Price [PRICE, 1976] sous la désignation d'« avantage cumulatif » et qui explique comment un processus d'évolution où les « riches deviennent plus riches » génère les lois de puissance reprenant des idées des années 1950 par H. Simon [SIMON, 1955]. La seule différence du modèle de Barabási et Albert par rapport au modèle de Price est que les liens du premier ne sont pas directionnels, ce qui facilite le traitement théorique [DOROGVTSEV et collab., 2000]. Ces modèles ont donné lieu durant les années 2000 à de très nombreuses études et des nombreuses variantes [ALBERT et BARABÁSI, 2002]. Outre la distribution de connectivité ces modèles développent des réseaux qui se distinguent par d'autres caractéristiques d'organisation comme la transitivité, les corrélations entre âge des nœuds et degré ainsi que les corrélations de degré entre nœuds.

Les modèles d'attachement préférentiel ne sont pas les seuls permettent d'expliquer l'émergence de réseaux complexes suivant une lois de puissance des degrés. D'autres types de modèles de construction de réseau conduisant également à de telles distribution ont été introduits qui représentent notamment mieux les processus de création des réseaux biochimiques qui seront discuté au chapitre 4. Les modèles de duplication et évolution représentent en effet davantage des mécanismes que l'on peut imaginer pour le développement de réseaux d'interaction entre gènes et protéines [DOROGVTSEV et collab., 2000; JUNG et collab., 2002; KLEINBERG et collab., 1999; SOLE et collab., 2002; VAZQUEZ et collab., 2001]. Les mécanismes à l'œuvre dans ces modèles comportent typiquement les étapes suivantes de : i) duplication d'un nœud tiré au hasard et de toutes ses connections, ii) divergence des connections dupliquées, un certain nombre au hasard d'entre-elles étant éliminées, iii) mutation avec la création de nouveaux liens au hasard entre le nœud dupliqué et le reste du réseau. Ces mécanismes représentent les idées courantes pour l'évolution du génome par duplication de gènes et diversification. Le mécanisme de duplication donne naissance à des distributions en loi de puissance avec des exposants dépendant des paramètres de divergence et mutation et typiquement dans l'intervalle $2 \leq \alpha \leq$ des observations dans les réseaux réels.

La discussion ci-dessus a porté sur les caractéristiques topologiques et des modèles de réseaux sans considération d'une métrique spatiale (uniquement une notion de voisinage d'origine dans le modèle petit-monde). Cependant quand les nœuds sont répartis géographiquement dans un espace on s'attend naturellement que leur répartition spatiale intervient dans les propriétés du réseau et des processus qui s'y déroulent. Les descripteurs et propriétés de tels réseaux ont également fait l'objet de nombreuses études [BARRAT et collab., 2005; BULLMORE et SPORNS, 2009; CRUCITTI et collab., 2006; GUILLIER et collab., 2017] qui ne seront pas discutées dans ce mémoire et constituent un champ actuel de recherche. On en trouvera une revue partielle dans la section 2.5 de la revue extensive de Boccaletti *et al.* [BOCCALETTI et collab., 2006] sur les réseaux complexes. On retrouvera la question des contraintes spatiales sur les réseaux dans le chapitre 6 à l'occasion de mes travaux de modélisation sur la croissance de cultures neuronale.

2.6.1 Neurosciences et réseaux complexes

Les exemples et secteurs d'application des réseaux complexes que nous venons de passer en revue sont très divers. C'est un paradigme puissant étroitement lié avec les technologies et la société actuelle. On le retrouve donc naturellement avec force dans le domaine des neurosciences où la structure même du cerveau correspond à un réseau extrêmement complexe avec de très nombreux éléments. Partant de la biologie des systèmes et des réseaux biochimiques ce contexte général m'a conduit également à venir m'intéresser aux réseaux neuronaux et mener les études qui seront présentées dans le chapitre 7.

Un cerveau, un système nerveux d'un animal même le plus élémentaire est une structure extrêmement organisée d'éléments très sophistiqués en interaction. Cet organe centralise des signaux de l'environnement, sait les traiter, en garder trace - mémoriser - et générer, on dira, calculer, une réponse adéquate. Le cas échéant il est capable de s'adapter pour réagir au mieux à des situations, ces-à- dire apprendre et mémoriser des comportements. Comprendre l'organisation d'un cerveau et son fonctionnement, les principes de l'apprentissage, de la mémoire et du traitement de l'information sont l'enjeu des neurosciences.

La notion de réseau neuronal semble en effet naturelle de nos jours. Les approches de théories des graphes et des réseaux complexes que nous venons de discuter ont depuis le courant de la dernière décennie eu un impact croissant dans la vision et l'analyse de données cérébrales [BARRAT et collab., 2005; BULLMORE et SPORNS, 2009; CRUCITTI et collab., 2006; GUILLIER et collab., 2017]. La neurobiologie est peut-être la discipline moderne étant venue la première à voir le paradigme d'un réseau complexe.

La découverte des neurones, de leur variété et complexité de formes, des interactions multiples qu'ils établissent entre eux et de l'organisation en réseau est le résultat d'une aventure scientifique de la fin du XIX^e siècle et début du XX^e rendue possible par les progrès de la microscopie à partir de 1873 avec la technique de coloration au nitrate d'argent inventée par Camillo Golgi.

Celle-ci a permis la découverte stupéfiante dans les tissus cérébraux des multiples structures aux formes déconcertantes avec d'innombrables ramifications reliées à d'autres structures analogues. Déjà une image du système nerveux en réseau complexe, mais dont la nature a fait polémique à l'époque. Ce système était-il un assemblage d'entités cellulaires distinctes mais interconnectées comme l'avait proposé Santiago Ramon y Cajal? Ou bien était-il un objet réticulaire diffus et continu comment le suggérait Camillo Golgi? Le terme « neurone » allait être introduit par l'anatomiste Heinrich Wilhelm Gottfried von Waldeyer-Hartz pour désigner les entités cellulaires interconnectées mais distinctes du tissu nerveux de la théorie « neuronale » de Santiago Ramon y Cajal dont il a été un des majeurs soutiens. Charles Scott Sherrington et Santiago Ramon y Cajal allaient par la suite faire la démonstration cruciale pour la neurologie moderne des neurones en tant que cellules distinctes. Les connexions directionnelles entre les neurones ont été nommées « synapses » en 1897 à l'initiative de Sherrington à partir d'un terme grec ancien dont la signification pourrait être « agrafe » [FESSARD et collab., 2017].

Aujourd'hui de nombreuses techniques expérimentales très sophistiquées existent pour observer l'activité neuronale de populations de neurones ou de cerveaux entiers. Les neurosciences expérimentales sont un des secteurs de haute technologie les plus avancées avec des applications en recherche fondamentale et médecine. Il est entièrement hors du cadre de ce mémoire de discuter de l'éventail de ces technologies. Quelques unes spécifiques au champ spécialisé de mes recherches sur les cultures neuronales sont discutées dans le chapitre 6. Il est cependant utile d'introduire une extension conceptuelle importante au paradigme de réseau complexe, c'est la notion de réseau « fonctionnel ».

La connectivité fonctionnelle est définie par l'observation de corrélations dans le temps de l'activité de zones distinctes dans des populations de neurones ou des zones cérébrales. L'existence de ces corrélations est interprétée comme l'existence d'un lien dans le cadre d'une fonc-

tion cérébrale car les zones corrélées se trouvent actives dans le mêmes processus. Des échanges doivent avoir lieu entre ces régions ou bien avoir été les cibles d'une activation commune.

Ces liens fonctionnels issus de corrélation sont notamment établis dans l'analyse des données d'imagerie résonance magnétique nucléaire fonctionnelle (RMN, en anglais MRI) qui mesure des signaux dépendant de l'oxygénation. En particulier les corrélations dans les fluctuations de ces signaux au repos permettent de mettre en évidence des liens directs entre deux zones corrélées, plutôt que l'activation simultanée en réponse à une commande partagé [VAN DEN HEUVEL et HULSHOFF POL, 2010; POWER et collab., 2011; REIJNEVELD et collab., 2007]. Le concept de réseau fonctionnel s'applique aussi dans d'autres mesures d'activité étendues sur le cerveau comme l'électroencéphalographie (EEG) ou sur des coupes de cerveau et des cultures neuronales dans les données de matrices d'électrodes (qui seront discutées dans le chapitre 6. Il est important de remarquer que ces réseaux fonctionnels ne correspondent pas directement au réseau neuronal biophysique des connections synaptiques entre neurones subjacents. Les corrélations qu'ils manifestent devraient en être cependant l'expression et refléter une partie des propriétés du réseau physique.

Depuis une dizaine d'années le concept de réseau complexe et les méthodes de théorie des graphes sont ainsi devenues une des approches les plus importantes d'analyse des données de neurosciences fonctionnelle [BULLMORE et SPORNS, 2009] également à des fins médicales pour la mise en évidence de caractéristiques de désordres et de diagnostic [VAN DEN HEUVEL et HULSHOFF POL, 2010].

2.7 Conclusion du chapitre

Biologie des systèmes, sciences des réseaux complexes et applications dans les neurosciences sont les champs disciplinaires de mes intérêts de recherche durant la vingtaine d'années que couvre la synthèse des travaux que je présente dans ce mémoire.

Dans ces domaines ma démarche a été celle du théoricien modélisateur et de la simulation numérique. Contribuant auprès de spécialistes des différentes questions abordées avec des méthodes venant de la modélisation physique, des systèmes dynamiques et de la physique numérique. Cette habilitation est aussi le prétexte pour moi d'un travail rétrospectif sur la démarche la démarche même de modélisation pour le vivant, objet du prochain chapitre.

2.8 Références

- ALBERT, R. et A. L. BARABÁSI. 2002, «Statistical mechanics of complex networks», *Reviews of Modern Physics*, vol. 74, n° 1, doi :10.1103/RevModPhys.74.47, p. 47–97, ISSN 00346861. 50, 56
- ALON, U. 2007, «An Introduction to Systems Biology : Design Principles of Biological Circuits», doi :citeulike-article-id:1314150. URL <http://www.loc.gov/catdir/enhancements/fy0654/2005056902-d.html>. 37
- ANTONELLI, P. 2014, «MoMA | Designing Life : Synthetic Biology and Design», URL <https://www.moma.org/explore/inside/out/2014/12/17/designing-life-synthetic-biology-and-design/>. 42
- ARDLIE, K. G., D. S. DELUCA, A. V. SEGRÈ et collaborateurs. 2015, «The Genotype-Tissue Expression (GTEx) pilot analysis : Multitissue gene regulation in humans», *Science*, vol. 348, n° 6235, doi :10.1126/science.1262110, p. 648–660, ISSN 10959203. 37
- BALL, P. 2013, «In retrospect : On growth and form», doi :10.1038/494032a. URL <http://www.nature.com/articles/494032a>. 25, 29

- BARABÁSI, A. L. et R. ALBERT. 1999, «Emergence of scaling in random networks», *Science*, vol. 286, n° 5439, doi :10.1126/science.286.5439.509, p. 509–512, ISSN 00368075. [56](#)
- BARABÁSI, A. L. et Z. N. OLTVAI. 2004, «Network biology : Understanding the cell's functional organization», *Nature Reviews Genetics*, vol. 5, n° 2, doi :10.1038/nrg1272, p. 101–113, ISSN 14710056. [51](#)
- BARRAT, A., M. BARTHÉLEMY et A. VESPIGNANI. 2005, «The effects of spatial constraints on the evolution of weighted complex networks», *Journal of Statistical Mechanics : Theory and Experiment*, vol. 2005, n° 5, doi :10.1088/1742-5468/2005/05/P05003, p. 49–68, ISSN 17425468. [56](#), [57](#)
- BAUERLE, C., A. DEPASS, D. LYNN et collaborateurs. 2009, *VISION AND CHANGE IN UNDERGRADUATE BIOLOGY EDUCATION A CALL TO ACTION*, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE. [27](#)
- BERNOULLI, D. 1760, «Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, & des avantages de l'inoculation pour la prévenir», dans *Histoire et Mémoires de l'Académie Royale des Sciences de Paris*, p. 1-45, 1760 (1766), die werke éd., Birkhäuser, p. 235–267. [30](#)
- BOCCALETTI, S., V. LATORA, Y. MORENO, M. CHAVEZ et D. U. HWANG. 2006, «Complex networks : Structure and dynamics», *Physics Reports*, vol. 424, n° 4-5, doi :10.1016/j.physrep.2005.10.009, p. 175–308, ISSN 03701573. [52](#), [56](#)
- BRODER, A., R. KUMAR, F. MAGHOUL et collaborateurs. 2000, «Graph structure in the Web», *Computer Networks*, vol. 33, n° 1, doi :10.1016/S1389-1286(00)00083-9, p. 309–320, ISSN 13891286. [50](#)
- BULLMORE, E. et O. SPORNS. 2009, «Complex brain networks : Graph theoretical analysis of structural and functional systems», *Nature Reviews Neuroscience*, vol. 10, n° 3, doi :10.1038/nrn2575, p. 186–198, ISSN 1471003X. [56](#), [57](#), [58](#)
- CALDARELLI, G. 2007, *Scale-free networks : complex webs in nature and technology*, June, Oxford University Press, Oxford, ISBN 9780199211517, 19–20 p.. [50](#)
- CHANCE, B. 1999, «The kinetics of the enzyme-substrate compound of peroxidase. 1943.», *Advances in enzymology and related areas of molecular biology*, vol. 73, n° 8, p. 3–23, ISSN 0065-258X. [32](#)
- CHAZOTTES, J.-R. et M. MONTICELLI. 2013, «Sur les modèles proie-prédateur en écologie», URL <http://www.espace-turing.fr/Sur-les-modeles-proie-predateur-en.html>. [32](#)
- CHEN, Q., H. CHANG, R. GOVINDAN et collaborateurs. 2002, «The origin of power laws in Internet topologies revisited», *Proceedings - IEEE INFOCOM*, vol. 2, doi :10.1109/INFOCOM.2002.1019306, p. 608–617, ISSN 0743166X. [50](#)
- CLAUSET, A., C. R. SHALIZI et M. E. J. NEWMAN. 2009, «Power-law distributions in empirical data», *SIAM review*, vol. 51, n° 4, doi :10.1137/070710111, p. 661–703, ISSN 1063-6897. [50](#)
- COMMITTEE ON UNDERGRADUATE BIOLOGY EDUCATION. 2003, *BIO2010 : Transforming Undergraduate Education for Future Research Biologists*, ISBN 0-309-50338-8. [27](#)
- CORNISH-BOWDEN, A. 2013, «The origins of enzyme kinetics», doi :10.1016/j.febslet.2013.06.009. URL <http://doi.wiley.com/10.1016/j.febslet.2013.06.009>. [32](#)
- CRUCITTI, P., V. LATORA et S. PORTA. 2006, «Centrality measures in spatial networks of urban streets», *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 73, n° 3, doi : 10.1103/PhysRevE.73.036125, p. 036 125, ISSN 15393755. [56](#), [57](#)

- DAISY GINSBERG, A., J. CALVERT, P. SCHYFTER et collaborateurs. 2014, *Synthetic Aesthetics*, MIT Press. 42
- DOROGVTSEV, S. N. et J. F. MENDES. 2002, «Evolution of networks», *Advances in Physics*, vol. 51, n° 4, doi :10.1080/00018730110112519, p. 1079–1187, ISSN 00018732. 50
- DOROGVTSEV, S. N., J. F. F. MENDES et A. N. SAMUKHIN. 2000, «Structure of growing networks with preferential linking», *Physical Review Letters*, vol. 85, n° 21, doi :10.1103/PhysRevLett.85.4633, p. 4633–4636, ISSN 00319007. 56
- ERDOS, P. et A. RENYI. 1959, «On Random Graphs I», *Publ. Math. Debrecen*, vol. 6, p. 290–297. 54
- FESSARD, A., J. TAXI, P. GUYENET et M. HAMON. 2017, «SYNAPSES», URL <http://www.universalis-edu.com/encyclopedie/synapses/>). 57
- FRIGG, R. et S. HARTMANN. 2017, «Models in Science», URL <https://plato.stanford.edu/archives/spr2017/entries/models-science/>. 27
- GABAIX, X. 1999, «Zipf’s Law for Cities : An Explanation», *The Quarterly Journal of Economics*, vol. 114, n° 3, doi :10.1162/003355399556133, p. 739–767, ISSN 0033-5533. 25
- GABRIEL, J.-P. et P. DE LA HARPE. 2010, «Daniel Bernoulli, pionnier des modèles mathématiques en médecine», URL <http://images.math.cnrs.fr/Daniel-Bernoulli-pionnier-des.html{#}nb2>. 30
- GOBAS, F. A., C. E. MACKINTOSH, G. WEBSTER et collaborateurs. 2003, «Bioaccumulation of phthalate esters in aquatic food-webs», dans *Handbook of Environmental Chemistry*, vol. 3, ISBN 978-3-540-00992-4 978-3-540-36611-9 SV - 3Q, p. 201–225, doi :10.1007/b11467. 48
- GROSSMAN, J. W. et P. D. F. ION. 1995, «On a Portion of the Well-Known Collaboration Graph», *Congressus Numerantium*, vol. 108, p. 129–131, ISSN 03849864. 50
- GUILLIER, S., V. MUÏ½OZ, J. ROGAN, R. ZARAMA et J. A. VALDIVIA. 2017, «Optimization of spatial complex networks», *Physica A : Statistical Mechanics and its Applications*, vol. 467, doi :10.1016/j.physa.2016.09.011, p. 465–473, ISSN 03784371. 56, 57
- GUNAWARDENA, J. «Luria & Delbrück», URL <http://vcp.med.harvard.edu/timeline/luria-delbruck.html>. 26, 31
- GUNAWARDENA, J. 2010, «Models in systems biology : the parameter problem and the meanings of robustness», dans *Elements of Computational Systems Biology*, édité par S. H. Muggleton et H. M. Lodhi, chap. 2, John Wiley & Sons, ISBN ISBN : 9780470180938. 37
- GUNAWARDENA, J. 2014, «Models in biology : ‘accurate descriptions of our pathetic thinking’», *BMC biology*, vol. 12, doi :10.1186/1741-7007-12-29, p. 29, ISSN 1741-7007. 23, 29
- HARTWELL, L. H., J. J. HOPFIELD, S. LEIBLER et A. W. MURRAY. 1999, «From molecular to modular cell biology.», *Nature*, vol. 402, n° 6761 Suppl, doi :10.1038/35011540, p. C47—52. 52
- VAN DEN HEUVEL, M. P. et H. E. HULSHOFF POL. 2010, «Exploring the brain network : A review on resting-state fMRI functional connectivity», *European Neuropsychopharmacology*, vol. 20, n° 8, doi :10.1016/j.euroneuro.2010.03.008, p. 519–534, ISSN 0924977X. 58
- HODGKIN, A. L. et A. F. HUXLEY. 1990, «A quantitative description of membrane current and its application to conduction and excitation in nerve», *Bulletin of Mathematical Biology*, vol. 52, n° 1-2, doi :10.1007/BF02459568, p. 25–71, ISSN 00928240. 33
- HUTTER, P. 2000, «Entendrons-nous un jour les symphonies intégrales de gènes?», *Médecine/Sciences*, vol. 16, p. 135–40. 27

- JEONG, H., S. P. MASON, A. L. BARABASI et Z. N. OLTVAI. 2001, «Lethality and centrality in protein networks.», *Nature*, vol. 411, n° 6833, doi :10.1038/35075138, p. 41–42. [52](#)
- JEONG, H., Z. NÉDA et A. L. BARABÁSI. 2003, «Measuring preferential attachment in evolving networks», *Europhysics Letters*, vol. 61, n° 4, doi :10.1209/epl/i2003-00166-9, p. 567–572, ISSN 02955075. [50](#)
- JOHANNSEN, W. L. 1905, *Arvelighedslaerens elementer*, København og Kristiania :Gyldendalske Boghandel Nordisk Forlag, Copenhagen. [32](#)
- JOHANNSEN, W. L. 1909, *Elemente der Exakten Erblichkeitslehre*, Gustav Fischer, Jena. [32](#)
- JUNG, S., S. KIM et B. KAHNG. 2002, «Geometric fractal growth model for scale-free networks», *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 65, n° 5, doi :10.1103/PhysRevE.65.056101, p. 6, ISSN 1063651X. [56](#)
- J. WATTS, D. 2000, «The Dynamics of Networks Between Order and Randomness», *The American Mathematical Monthly*, vol. 107, n° 7, p. 664–668. [54](#)
- KERMACK, W. O. et A. G. MCKENDRICK. 1991a, «Contributions to the mathematical theory of epidemics-I 1927», *Bulletin of Mathematical Biology*, vol. 53, n° 1-2, doi :10.1007/BF02464423, p. 33–55, ISSN 00928240. [32](#)
- KERMACK, W. O. et A. G. MCKENDRICK. 1991b, «Contributions to the mathematical theory of epidemics-II. The problem of endemicity 1931», *Bulletin of Mathematical Biology*, vol. 53, n° 1-2, doi :10.1007/BF02464424, p. 57–87, ISSN 00928240. [32](#)
- KERMACK, W. O. et A. G. MCKENDRICK. 1991c, «Contributions to the mathematical theory of epidemics-III. Further studies of the problem of endemicity 1933», *Bulletin of Mathematical Biology*, vol. 53, n° 1-2, doi :10.1007/BF02464425, p. 89–118, ISSN 00928240. [32](#)
- KIRSCHNER, M. W. 2005, «The meaning of systems biology», *Cell*, vol. 121, n° 4, doi :10.1016/j.cell.2005.05.005, p. 503–504, ISSN 00928674. [37](#), [38](#)
- KITANO, H. 2002, «Systems biology : a brief overview.», *Science*, vol. 295, n° 5560, doi :10.1126/science.1069492, p. 1662–1664. [52](#)
- KLEINBERG, J. M., R. KUMAR, P. RAGHAVAN, S. RAJAGOPALAN et A. S. TOMKINS. 1999, «The web as a graph : Measurements, models, and methods», dans *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 1627, Springer, Berlin, Heidelberg, ISBN 3540662006, ISSN 16113349, p. 1–17, doi : 10.1007/3-540-48686-0_1. [56](#)
- KRESGE, N., R. D. SIMONI et R. L. HILL. 2004, «Britton Chance : Olympian and Developer of Stop-Flow Methods 100 Years of Biochemistry and Molecular Biology», *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, vol. 279, n° 50. [32](#)
- LOTKA, A. J. 1925, *Elements of Physical Biology*, Williams & Wilkins Company, 425 p.. [32](#)
- MALAKOFF, D. 1999, «Genomic nanotech centers open \$200 million push by Harvard.», *Science (New York, N.Y.)*, vol. 283, n° 5402, doi :10.1126/SCIENCE.283.5402.610, p. 610–1, ISSN 0036-8075. [27](#)
- MCKENDRICK, A. G. 1926, «Applications of mathematics to medical problems», *Proceedings of the Edinburgh Mathematical Society*, vol. 14, p. 98 – 130. [30](#)
- MENDEL, G. 1866, «Versuche über Pflanzen-Hybriden (Experiments in plant hybridization)», dans *Verhandlungen des Naturforschenden Vereins (Proceedings Natural History Society)*, Brünn. [32](#)

- MOODY, J. 2001, «Race, School Integration, and Friendship Segregation in America», *American Journal of Sociology*, vol. 107, n° 3, doi :10.1086/338954, p. 679–716, ISSN 0002-9602. [48](#)
- MURRAY, J. 1977, *Lectures on Nonlinear Differential Equation Models in Biology*, Oxford U.P., Oxford, 370pp p.. [33](#)
- MURRAY, J. D. «Why Are There No 3-Headed Monsters? Mathematical Modeling in Biology», doi : 10.1090/noti865. [30](#), [33](#)
- MURRAY, J. D. 2002, *Mathematical Biology*, Springer. [30](#)
- MURRAY, J. D. 2012, «Vignettes from the field of mathematical biology : the application of mathematics to biology and medicine.», *Interface focus*, vol. 2, n° 4, doi :10.1098/rsfs.2011.0102, p. 397–406, ISSN 2042-8901. [30](#), [32](#), [33](#)
- NEWMAN, M. E. et D. J. WATTS. 1999, «Scaling and percolation in the small-world network model.», *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics*, vol. 60, n° 6 Pt B, p. 7332–7342. [52](#)
- NEWMAN, M. E. J. 2003, «The structure and function of complex networks.», *SIAM Review*, p. 167–256. [48](#), [50](#), [52](#), [54](#)
- PORTER, M. A., J.-P. ONNELA et P. J. MUCHA. 2009, «Communities in Networks Introduction : Networks and Communities», *Notices of the American Mathematical Society*, vol. 56, n° 9, p. 19. [53](#)
- POTTERS, G. 2010, «Systems Biology of the Cell», *Nature Education*, vol. 3, n° 9, p. 33. [35](#), [47](#)
- POWER, J. D., A. L. COHEN, S. M. NELSON et collaborateurs. 2011, «Functional Network Organization of the Human Brain», *Neuron*, vol. 72, n° 4, doi :10.1016/j.neuron.2011.09.006, p. 665–678, ISSN 08966273. [58](#)
- PRICE, D. D. S. 1976, «A general theory of bibliometric and other cumulative advantage processes», *Journal of the American Society for Information Science*, vol. 27, n° 5, doi :10.1002/asi.4630270505, p. 292–306, ISSN 10974571. [56](#)
- PRIGOGINE, I., R. LEFEVER, A. GOLDBETER et M. HERSCHKOWITZ-KAUFMAN. 1969, «Symmetry breaking instabilities in biological systems.», *Nature*, vol. 223, n° 5209, p. 913–6, ISSN 0028-0836. [33](#)
- PRIGOGINE, I. et G. NICOLIS. 1967, «On Symmetry-Breaking Instabilities in Dissipative Systems», *The Journal of Chemical Physics*, vol. 46, n° 9, doi :10.1063/1.1841255, p. 3542–3550, ISSN 0021-9606. [33](#)
- RAMÓN Y CAJAL, S. 1899, «Reglas y consejos sobre investigación científica.», . [26](#)
- RAVASZ, E. et A.-L. BARABÁSI. 2003, «Hierarchical organization in complex networks.», *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 67, n° 2 Pt 2, doi :10.1103/PhysRevE.67.026112, p. 026 112, ISSN 1539-3755. [51](#), [52](#)
- RAVASZ, E., A. L. SOMERA, D. A. MONGRU, Z. N. OLTVAI et A. L. BARABÁSI. 2002, «Hierarchical organization of modularity in metabolic networks.», *Science (New York, N.Y.)*, vol. 297, n° 5586, doi :10.1126/science.1073374, p. 1551–5, ISSN 1095-9203. [51](#), [52](#)
- REDNER, S. 1998, «How popular is your paper? An empirical study of the citation distribution», *European Physical Journal B*, vol. 4, n° 2, doi :10.1007/s100510050359, p. 131–134, ISSN 14346028. [50](#)

- REIJNEVELD, J. C., S. C. PONTEN, H. W. BERENDSE et C. J. STAM. 2007, «The application of graph theoretical analysis to complex networks in the brain», *Clinical Neurophysiology*, vol. 118, n° 11, doi :10.1016/j.clinph.2007.08.010, p. 2317–2331, ISSN 13882457. 58
- RUBINOV, M. et O. SPORNS. 2010, «Complex network measures of brain connectivity : Uses and interpretations», *NeuroImage*, vol. 52, n° 3, doi :10.1016/j.neuroimage.2009.10.003, p. 1059–1069, ISSN 10538119. 54
- RUTHERFORD, S. 2011, «Toward a physical biology», *BioEssays*, vol. 33, n° 6, doi :10.1002/bies.201190023, p. 397–397, ISSN 02659247. 26
- SAINT-HILAIRE, I. G. 1832, *Histoire Générale et Particulière des Anomalies de l'Organisation Chez l'Homme et les Animaux, ou Traité de Tératologie*, Paris. 30, 31
- SERVICE, R. F. 1999a, «Berkeley puts all its eggs in two baskets», *Science*, vol. 286, n° 5438, doi : 10.1126/science.286.5438.226, p. 226–227, ISSN 00368075. 27
- SERVICE, R. F. 1999b, «COMPLEX SYSTEMS :Exploring the Systems of Life», *Science*, vol. 284, n° 5411, doi :10.1126/science.284.5411.80a, p. 80a–83, ISSN 00368075. 27
- SHENDURE, J., S. BALASUBRAMANIAN, G. M. CHURCH et collaborateurs. 2017, «DNA sequencing at 40 : Past, present and future», *Nature*, vol. 550, n° 7676, doi :10.1038/nature24286, p. 345–353, ISSN 14764687. 36
- SHINGLETON, A. 2010, «Allometry : The Study of Biological Scaling | Learn Science at Scitable», *Nature Education Knowledge*, vol. 3, n° 10, p. 2. 28
- SHOU, W., C. T. BERGSTROM, A. K. CHAKRABORTY et F. K. SKINNER. 2015, «Theory, models and biology», *eLife*, vol. 4, doi :10.7554/eLife.07158, p. e07 158, ISSN 2050-084X. 27, 30, 32
- SIMON, H. A. 1955, «Biometrika Trust [425] ON A CLASS OF SKEW DISTRIBUTION FUNCTIONS», *Source : Biometrika*, vol. 42, n° 34, doi :10.1093/biomet/42.3-4.425, p. 425–440, ISSN 0006-3444. 56
- SOLE, R. V., R. PASTOR-SATORRAS, E. SMITH et T. B. KEPLER. 2002, «A model of large-scale proteome evolution», *Advances in Complex Systems*, vol. 05, n° 01, doi :10.1142/S021952590200047X, p. 43–54, ISSN 0219-5259. 56
- SOLOMONOFF, R. et A. RAPOPORT. 1951, «Connectivity of random nets», *The Bulletin of Mathematical Biophysics*, vol. 13, n° 2, doi :10.1007/BF02478357, p. 107–117, ISSN 00074985. 54
- STROGATZ, S. H. 2001, «Exploring complex networks.», *Nature*, vol. 410, n° 6825, doi :10.1038/35065725, p. 268–276. 50, 54
- TURING, A. M. 1952, «The Chemical Basis of Morphogenesis», *Philosophical Transactions of the Royal Society B : Biological Sciences*, vol. 237, n° 641, doi :10.1098/rstb.1952.0012, p. 37–72, ISSN 0962-8436. 33
- VAZQUEZ, A., A. FLAMMINI, A. MARITAN et A. VESPIGNANI. 2001, «Modeling of protein interaction networks», *Complexus*, vol. 1, n° 1, doi :10.1159/000067642, p. 38–44, ISSN 1424-8506. 56
- VOLTERRA, V. 1926, «Fluctuations in the abundance of a species considered mathematically», doi : 10.1038/118558a0. URL <http://www.nature.com/doifinder/10.1038/118558a0>. 32
- WATTS, D. J. 2003, *Six Degrees : The Science of a Connected Age*, W. W. Norton. 52
- WATTS, D. J. et S. H. STROGATZ. 1998, «Collective dynamics of 'small-world' networks.», *Nature*, vol. 393, n° 6684, doi :10.1038/30918, p. 440–442. 50, 51, 52, 54, 55

WORTHINGTON. 2018, «Introduction (Introduction to Enzymes)», URL <http://www.worthington-biochem.com/introbiochem/>. 32

ZACHARY, W. W. 1977, «An Information Flow Model for Conflict and Fission in Small Groups», *Journal of Anthropological Research*, vol. 33, n° 4, doi :10.1086/jar.33.4.3629752, p. 452–473, ISSN 0091-7710. 53

Chapitre 3

Approches de modélisation

« The skill and innovation of experimental biologists has enabled them to get more and more information about their preparations. This presents a challenge to anyone who wishes to create a mathematical model or simulation of the given system. »

Bard Ermentrout in « Simplifying and Reducing Complex Models » [ERMENTROUT, 2001]

Sommaire

3.1 Modélisation statistique	68
3.2 Modélisation théoriques de mécanismes.	69
3.2.1 Topologies des interactions et dynamique des systèmes	69
3.2.2 Modélisation à grande échelle et simulation	70
3.2.3 Simulation en ingénierie physique versus sciences biologiques	76
3.2.4 Simplifications et modèles réduits	77
3.2.5 Approches de simplification	79
3.3 Inférence de paramètres	82
3.3.1 Détermination de paramètres de modèles biologiques	83
3.3.2 Modèles stationnaires	84
3.3.3 Identifiabilité paramétrique	87
3.3.4 Inférence Bayésienne	87
3.4 Conclusion du chapitre	89
3.5 Références	90

La modélisation mathématique est désormais incontournable dans certains courants de la biologie moléculaire, de génétique et de neurosciences. L'histoire de la place des mathématiques et de la théorie n'est cependant pas la même dans ces deux domaines de l'étude du Vivant, mais partage un même ensemble de concepts, d'outils et méthodes en lien aussi avec la physique et l'informatique. Les mêmes considérations épistémologiques et pratiques sur la nécessité, la nature et la portée des modèles construits s'appliquent à propos des mécanismes moléculaires et génétiques et concernant les populations de neurones.

La rédaction de ce mémoire m'a amené à réfléchir sur la démarche de modélisation théorique sur le vivant dans laquelle se sont inscrits mes travaux. Quels sont les types de modèles réalisables, leur sens et leur pertinence? Dans ce second chapitre de la partie d'introduction générale j'ai souhaité exposer ce cadre général de la modélisation en biologie dont la pratique est hétérogène dans la sphère universitaire des sciences de la vie et déjà aux marges des domaines de la physique et des mathématiques. Dans la carte conceptuelle 3.1 j'ai tenté de rassembler et relier les différents enjeux et notions concernant la démarche de modélisation scientifique et dont on discutera plusieurs aspects fondamentaux dans la suite de ce chapitre.

J'aborderai dans les autres chapitres de ce mémoire des questions méthodologiques plus spécialisées en lien avec les domaines d'applications, biologie des systèmes et neurosciences où se sont situés mes recherches. C'est-à-dire à chaque fois les manières particulières de mettre en forme mathématiquement les modèles correspondants aux situations scientifiques étudiées. Quel que soit le formalisme, la création d'un modèle mathématique implique de décrire la situation naturelle étudiée avec des objets et relations mathématiques, ou en tout cas formelles, permettant de calculer des comportements en fonction de paramètres ou de conditions biologiques.

Le type de modélisation dont il est question dans mon travail, est la création de modèles représentant les mécanismes de phénomènes et de processus. Des modèles mécaniques, ou plutôt des mécanismes, qu'il est nécessaire de distinguer des modèles statistiques dont la pratique est beaucoup plus consensuelle et usuelle.

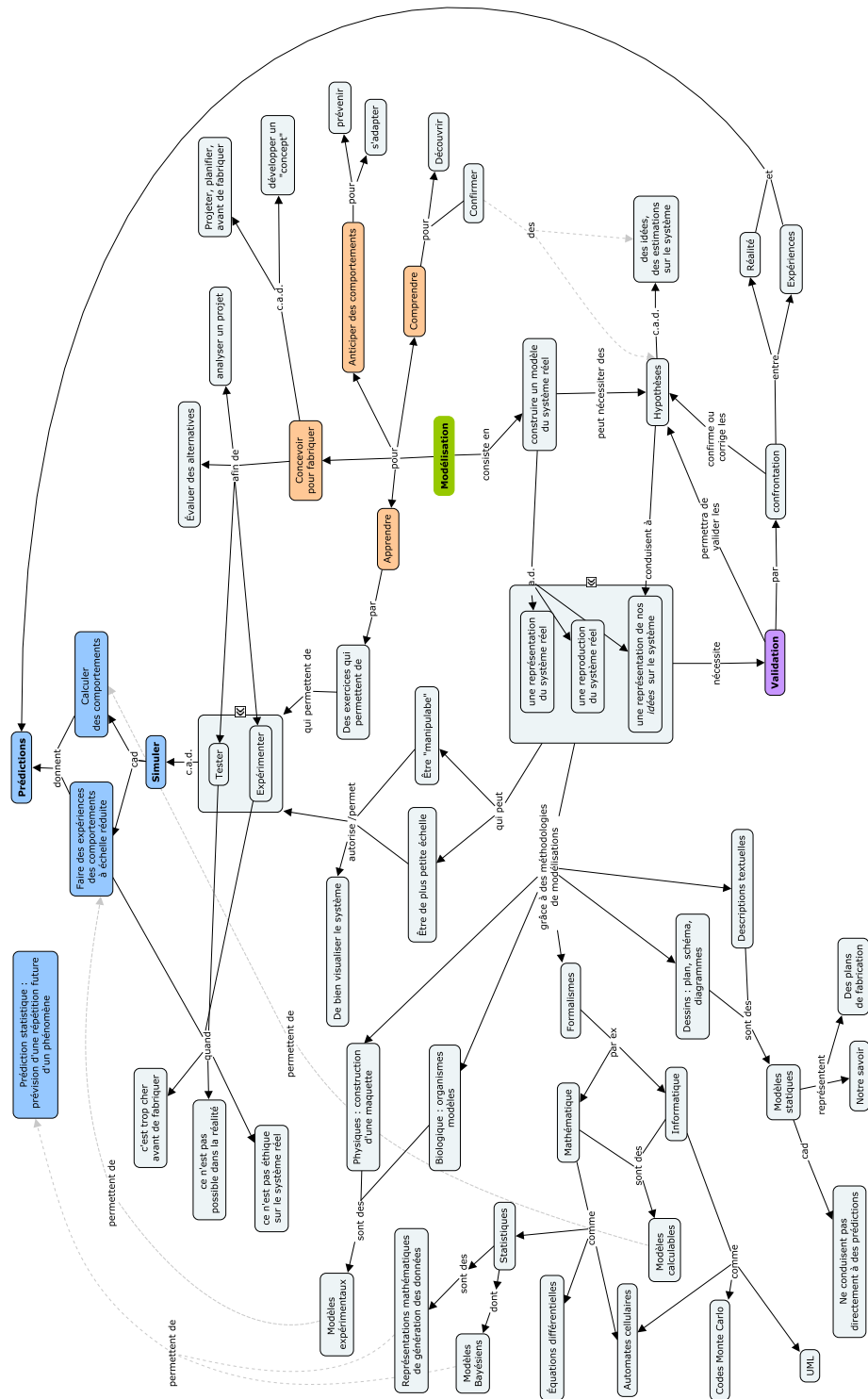


FIGURE 3.1 – Carte conceptuelle de la démarche de modélisation. Modélisation théorique à des fins de découverte, anticipation, conception ou apprentissage.

3.1 Modélisation statistique

La modélisation statistique est en effet centrale pour l'analyse de données d'expériences et d'observations. Elle vise à mettre en évidence et caractériser les tendances et les variations au sein de résultats d'observations. Les mesures de toute grandeur naturelle sont sujettes à des incertitudes. Ainsi l'accumulation de résultats issus de la répétition d'une expérience produit une distribution des mesures des grandeurs observées; une fonction qu'il s'agit de caractériser, décrivant comment se répartissent les valeurs, assimilée à la probabilité d'obtenir un résultat donné.

Prévisions statistiques Un *modèle statistique* représente mathématiquement les relations entre des grandeurs variables à chaque répétition des observations. Les objets mathématiques qui les forment, équations ou relations codées informatiquement, permettent par le calcul de reproduire la distribution des données observées ayant servi à les établir. Comme les autres modèles mathématiques de mécanismes dont on discutera plus bas, les modèles statistiques permettent une forme de prédiction : en supposant la persistance dans le temps des mêmes processus subjacents aux données, les modèles statistiques produisent des *prévisions* probabilistes des résultats possibles lors de répétitions d'un phénomène ou d'une expérience dans les mêmes conditions. Les modèles statistiques sont descriptifs. Ils capturent et représentent la variabilité et les relations de données obtenues dans des conditions précises. Supposant des conditions constantes ils donnent une estimation de la probabilité d'obtenir un certain résultat lors d'une nouvelle observation dans les mêmes conditions.

Les statistiques sont omni-présentes dans toute discipline produisant des données car leurs modèles permettent d'identifier les propriétés principales dans les observations et d'être utiles dans toutes les applications qui les exploitent pour effectuer des prévisions sur des résultats possibles, diagnostics en santé, effets probables de décisions, etc... Ces modèles ne fournissent pas cependant directement des explications des causes et mécanismes en jeu. Les corrélations qui peuvent apparaître entre des grandeurs ne sont pas des indications de liens causaux directs. L'absence de représentation des mécanismes se manifeste dans l'incapacité des modèles statistiques à établir des prédictions pour d'autres conditions que celles avec lesquelles ils-ont été construits.

Modèles statistiques et modèles de mécanismes Différemment des modèles statistiques, les modèles qui sont recherchés pour comprendre les processus en jeu dans un phénomène sont établis en premier lieu sur des hypothèses des mécanismes en jeu. Ces modèles de mécanismes visent à capturer les comportements produisant les données. Si le modèle représente de manière satisfaisante la physique, la biologie la chimie du système étudié, il doit être apte à reproduire des résultats dans des situations différentes de celles disponibles initialement et qui ont aidé à sa construction. La validation d'un tel modèle intervient vraiment lorsque ses prédictions auront été confirmées par de nouvelles observations dans des conditions nouvelles.

La modélisation statistique permet de guider la construction de modèles explicatifs des processus. Les relations qu'elle met en évidence dans les données et leur variabilités doivent naturellement être capturées par les modèles des processus physico-chimiques. Un point de convergence important entre les approches se retrouve dans l'évaluation des paramètres des modèles mécaniques. Les méthodes d'apprentissage statistiques qui visent à déduire, inférer, les caractéristiques de quantités inconnues de modèles à partir de données sont des outils très utiles pour déterminer les inconnues dans les modèles de processus. Réciproquement les relations et équations bien établies dans les modèles explicatifs peuvent servir à guider ou contraindre des choix de formes dans les modèles statistiques eux-mêmes.

À une époque d'explosion du nombre de données disponibles les concepts, outils et méthodes des statistiques connaissent un développement fulgurant s'affichant dans des terminologies à la mode « big data », « machine learning », « data science » etc... il s'agit d'extensions de l'analyse

de données tirant parti de la puissance de calcul moderne et des progrès du génie logiciel. La démarche de modélisation du physicien telle que je l'ai pratiquée et dans les laboratoires de recherche fondamentaux reste concentrée sur l'explication scientifique de mécanismes dans les systèmes étudiés et la modélisation des mécanismes plutôt que des données. Penchons-nous sur la nature de ce type de modélisation.

3.2 Modélisation théoriques de mécanismes.

Face à la variété des mécanismes en action la première étape d'une démarche de modélisation théorique consiste avant tout à préciser ses propres objectifs en l'entreprenant. Ces choix sont au cœur du travail de modélisation théorique et impliquent de nombreuses hypothèses, parfois arbitraires, sur les simplifications effectuées et les paramètres du système. La première étape du travail de modélisation est en effet de réaliser une abstraction du système naturel étudié. D'en réaliser une réduction dans des éléments plus restreints que l'on espère significatifs. C'est le cœur du travail théorique de faire la part entre les propriétés essentielles du phénomène étudié et les caractéristiques que l'on peut ignorer. Une fois la définition des processus dans le modèle effectuée, sa traduction en un système mathématique peut se faire de différentes façons en fonction là aussi de la question étudiée. L'approche mathématique déterminera notamment si l'on cherche à décrire un comportement moyen, qui serait captable par des méthodes déterministes, ou bien si l'on s'intéresse aux fluctuations des abondances biomoléculaires. Le choix de formalisme mathématique conditionnera également les méthodes de résolution et de calcul et la possibilité même d'obtenir d'une solution. La bonne conceptualisation permettra de capturer l'essentiel d'un phénomène dans un cadre qui permet de l'analyser et suffisamment flexible pour permettre des extensions.

Que mettre dans un modèle? Quels détails inclure? Quelles simplifications sont possibles? Le modèle peut rassembler quelques hypothèses, quelques idées simples sur des facteurs qui seraient en action. Si en développant le comportement du modèle, les mêmes propriétés que l'observation dans la nature sont observées, alors il aurait déjà la qualité de suggérer un scénario possible pour la réalité. Une suggestion d'explication peut-être insuffisamment étayée par une ample confrontation à des données expérimentales peut-être tout simplement non disponibles, mais dont la vraisemblance pourrait-être renforcée par son esthétique. Des hypothèses simples, peu nombreuses, reliées à des principes mathématiques ou physiques fondamentaux qui suffiraient à déduire les traits généraux des propriétés observées expérimentalement auront un sens interprétatif. Une valeur dont il importera d'être très lucide sur ses fondements théoriques plus qu'observationnels.

3.2.1 Topologies des interactions et dynamique des systèmes

Aussi bien la biologie des systèmes que les neurosciences sont en rapport avec des réseaux d'entités biologiques en interaction. Espèces biochimiques et structures biomoléculaires dans le premier cas, neurones et zones cérébrales communiquant les uns avec les autres dans l'autre cas. Rien de fondamentalement différent d'autre champs d'application de la théorie des systèmes dynamiques combinée à la théorie des graphes. Deux questions théoriques sont effectivement associées. D'une part la structure des réseaux, qui ne sont pas réguliers, et dont les propriétés topologiques vont affecter comment l'activité s'y déroule ainsi que la manière dont la structure elle-même peut évoluer. D'autre part les propriétés d'un système dynamique dont le comportement au cours du temps dépend de la nature des interactions entre les éléments qui le composent. La théorie des systèmes dynamique est déjà difficile sur de petits systèmes en y démontrant une très grande richesse de comportements. Un challenge de la théorie en biologie systémique ainsi que des neurosciences est bien d'étendre l'analyse à de grands ensembles construits suivant des graphes complexes.

L'objectif est bien dans le premier cas de construire des modèles dynamiques reproduisant

l'évolution temporelle d'abondances de protéines ou d'autres bio-molécules, d'analyser les régimes dynamiques, leurs transitions et sensibilité à des changements de paramètres. Et dans le cas des neurones de comprendre les types d'activité électrique en fonction des connectivités, des interactions synaptiques ainsi que des courants d'entrée et des modulations sur le système. Des questions fondamentalement anciennes, bien présentes même avant que l'on parle autant d'approches théoriques en biologie qui ont gagné assez récemment légitimité et importance avec les changements d'échelles d'investigation.

On peut tenter de traduire mathématiquement, ou informatiquement en code de simulation, toutes les connaissances disponibles sur le système biologique étudié. Chercher une représentation dans des équations de toutes les relations connues ou supposées, et avoir des variables pour décrire toutes les composantes identifiées pour être présentes et participer d'une certaine manière. Mais que donnera ce modèle? La description formelle n'est pas tout, pour être intéressante elle doit permettre le calcul de comportements et des prédictions. Encore faut-il pouvoir calculer un modèle. Et pouvoir encore le faire de manière à pouvoir apprendre des informations utiles sur les phénomènes en jeu. La manière de modéliser n'est pas unique, elle est affaire de gout, de choix d'objectif, de moyens et des connaissances scientifiques sur les domaines étudiés. De nos jours la question du type de modélisation et de leurs différentes possibilités et limitations se manifeste par deux grandes tendances répondant en réalité à des préoccupations différentes et complémentaires. Il s'agit d'une part de la modélisation pour la simulation haute performance reposante sur la puissance de calcul des ordinateurs et des algorithmes modernes, développée et appliquée dans un but essentiellement d'ingénierie. Et d'autre part, concernant plutôt la recherche scientifique fondamentale, la modélisation de systèmes réduits, hautement simplifiés, visant des approches mathématiques analytiques ou semi-analytiques en combinaison avec la simulation.

3.2.2 Modélisation à grande échelle et simulation

La simulation est devenue un élément essentiel de la méthode scientifique. Le calcul à haute performance et la simulation est dans ces secteurs un facteur clé du développement technologique. Pour la conception de produits, l'ingénierie de tous les secteurs concernés par les fluides, l'énergie, les structures, l'électromagnétisme et l'optique repose largement de nos jours sur l'usage de logiciels de modélisation et de simulations essentiellement commerciaux associés à des interfaces de Conception Assistée par Ordinateur CAO pour la spécification des systèmes à simuler. Ces environnements permettent des études dites « Multi-physique » combinant plusieurs domaines pour explorer des interactions telle qu'entre chaleur et propriétés mécaniques ou électromagnétiques, entre écoulements et structures et autres combinaisons [ANSYS; COMSOL; DIGICAD; "IDAHO NATIONAL LABORATORY "; MATLAB; SIEMENS], exemple figure 3.2. L'enjeu est d'étudier et de tester *in silico* de nouveaux dispositifs avant des réalisations expérimentales plus coûteuses, et de cette manière de réduire les coûts et durées de développement ou, comme en astrophysique, quand les expériences ne sont tout simplement pas possibles. Remarquons que les logiciels évoqués ici, comme l'exemple de la figure 3.3, sont à la fois des outils de modélisation et des outils de simulation, servant à construire et faire fonctionner les modèles. Ce sont toutes des ressources qui mettent en œuvre de la modélisation pour de la simulation. En physique, hors du champ de l'ingénierie la simulation est également devenue un outil standard de recherche quand les expériences ne sont pas possibles.

Cette tendance est également bien présente hors de l'ingénierie basée sur la physique et a été explorée pour le développement d'installations chimiques dès les années 1960 [YOULE, 1961]. Elle correspond à une offre de logiciels spécialisés (une liste étendue référencée dans [LIST OF CHEMICAL PROCESS SIMULATORS]) ou de modules de certains des environnements de modélisation multi-physique [COMSOL, 2017; SIEMENS DIGITAL FACTORY DIVISIO, 2016] permettant des études physico-chimiques telles que la combinaison de réactions et de phénomènes de transport de matière et thermiques. Pour le médicament et les biotechnologies la simulation est maintenant

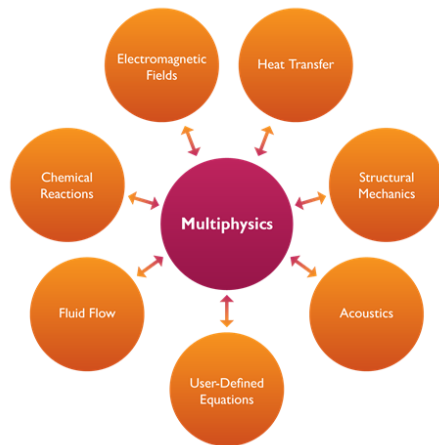


FIGURE 3.2 – **Les simulations multi-physiques.** Illustration des simulations multi-physiques d'après une présentation promotionnelle sur le site de la suite COMSOL [KAN, 2012]. Les systèmes multi-physiques visent à combiner tous les types de phénomènes naturels (fondamentalement physiques) dans un modèle unique possible à simuler.

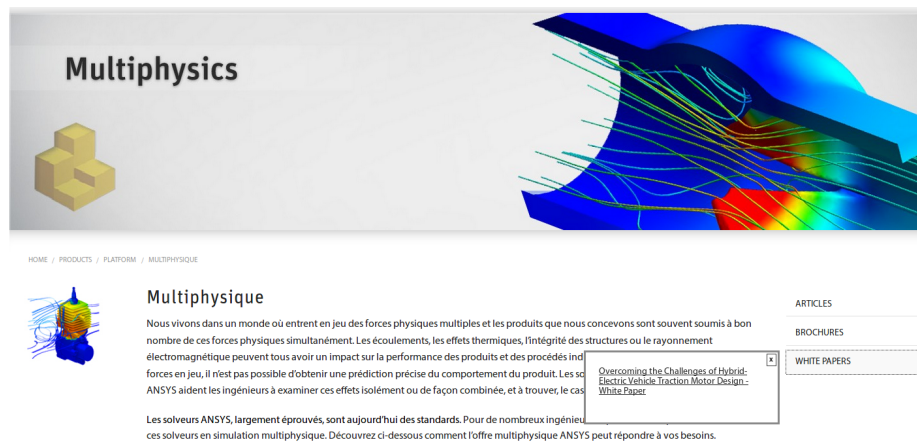


FIGURE 3.3 – **Exemple d'une suite logicielle de simulation multi-physique.** La suite Ansys [ANSYSIS, 2018], comme les autres systèmes cités dans le textes permettent de combiner design 3D des dispositifs étudiés avec la modélisation et la simulation de leurs propriétés physiques et chimiques.

centrale pour l'investigation d'affinités entre composés et la recherche et le design de nouvelles macro molécules [BIOVIA, a]. Ces secteurs appliquent maintenant industriellement les technologies de la génomique, et s'intéressent à explorer systématiquement sur de larges échelles les structures 3D et les propriétés de macromolécules biologiques telles que enzymes, protéines, anticorps, molécules d'ARN... La « dynamique moléculaire » est dans cet esprit une technique pour simuler les mouvements d'atomes et de molécules individuels utilisés pour comprendre les mouvements de structures biochimiques [LOUKATOU et collab.; VLACHAKIS et collab., 2014].

On touche déjà ici les ambitions de la biologie des systèmes et de la biologie synthétique, dont une des motivations de la communauté est justement la création d'une ingénierie du vivant à l'image de l'ingénierie physique. Cependant les applications de modélisation en biologie des systèmes affichées par les éditeurs de logiciels d'ingénierie comme Dassault Systèmes avec sa succursale BIOVIA [b] pour le biomédical et les biotechnologies restent encore largement au niveau des structures moléculaires et des interactions chimiques. L'ouvrage collectif « Computational Systems Biology » [GU et collab., 2014] publié en 2014 discute les très importants développements en cours vers des outils de modélisation pour la simulation intensive à grande échelle en biologie des systèmes. Plusieurs plateformes d'envergure disponibles à l'heure actuelle sont passées en revue dans le chapitre 8 « Standards, Platforms and Applications » [GU et collab., 2014]. Ainsi des logiciels en développement continu depuis plus de quinze ans comme E-Cell [TOMITA, 2001], VCell « The virtual Cell » [LOEW et SCHAFF, 2001], ou CellDesigner [FUNAHASHIA et collab., 2003; KITANO] illustrent cette ambition d'avoir à disposition un outil de modélisation intégré à l'échelle de la cellule toute entière alimentant une simulation de celle-ci. La figure 3.5 illustre le cycle de



FIGURE 3.4 – **Exemple d’environnement de simulation pour les biotechnologies et le pharmaceutique.** Par la société Dassault Systèmes, BIOVIA « Biologics unified research, development and manufacturing processes » se veut de proposer aux industries des biotechnologies et pharmaceutiques un ensemble d’outils de modélisation pour couvrir toutes les échelles du développement de produits [SYSTEMES]. La partie dédiée à la recherche du logiciel (partie nommée « découverte » dans le jargon industriel), le système fournit par exemple « un environnement de modélisation d’anticorps pour la prédiction des propriétés physiques et leur optimisation ».

modélisation d’un système cellulaire dans l’environnement VCell.

En neurosciences également, bien que l’aspect d’outil pour de l’ingénierie ne soit pas présent, existe la volonté de pouvoir réaliser des modélisations et des simulations de réseaux de neurones et de structures cérébrales macroscopiques. En neurosciences, la simulation est utilisée pour étudier les relations entre données anatomiques et physiologiques, pour explorer des modèles dynamiques non accessibles par des méthodes mathématiques analytiques, ainsi que par les théoriciens pour valider des approximations faites à partir de dérivations théoriques. Un rapport historique avec la modélisation existe dans ce domaine depuis l’explication ayant conduit au prix Nobel des potentiels d’action en 1952 par Alan Lloyd Hodgkin et Andrew Fielding Huxley en termes d’un modèle mathématique [HODGKIN et HUXLEY, 1952]. Pendant les décennies suivantes la modélisation de neurones uniques, ainsi que que des interactions synaptiques a été raffinée tant du point de vue mathématique, de l’identification de nouveaux canaux ioniques impliquées et de l’ajustement des paramètres sur des données expérimentales [CLAY, 2005].

Comme dans d’autres secteurs, le développement de la simulation en neurosciences été rendu possible grâce aux progrès réalisés du matériel et du logiciel informatique, ainsi que des algorithmes de simulation eux-mêmes pour des modèles couvrant un large spectre d’application depuis la dynamique moléculaire des canaux ioniques à des modèles détaillés par compartiments de cellules nerveuses individuelles, de réseaux neuronaux et de champs cérébraux. Aujourd’hui, il existe des codes de simulation pour tous ces niveaux, même si le degré d’utilisation par la communauté des chercheurs en neuroscience semble variée suivant les spécialités.

À des échelles d’observation microscopiques, des modèles détaillés étudient le comportement de neurones individuels, de compartiments particuliers de ceux-ci comme synapses et cônes de croissance, ou encore de constituants moléculaires comme des canaux ioniques, éventuellement de deux ou quelques neurones en interaction. D’autres modèles à une échelle mésoscopique impliquent des populations de centaines de milliers de neurones correspondant à des zones cervicales. Le sujet est alors principalement la coordination des activités collectives à partir des comportements de chaque unité. Enfin d’autres approches macroscopiques plus larges visent une simulation de cerveaux entiers pour approcher le fonctionnement de processus cognitifs [CESSAC et collab., 2010]. Pour les spécialités qui peuvent la mettre en œuvre la modélisation et la simulation est une démarche très utile pour valider, ou infirmer des hypothèses de recherche.

L’usage de la simulation dans les neurosciences surtout développée à partir du milieu des an-

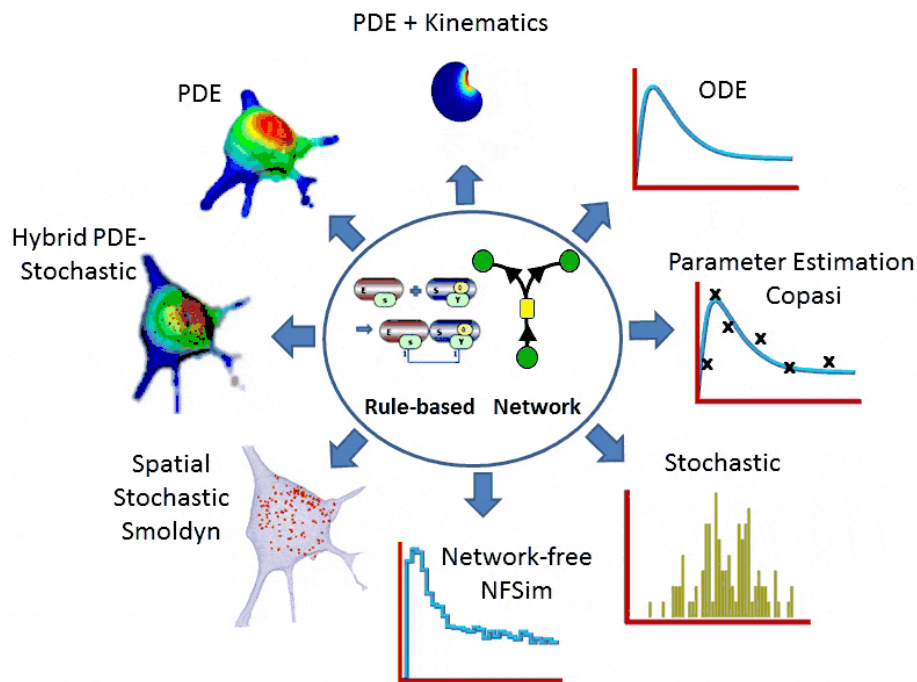


FIGURE 3.5 – **Exemple d’environnement modélisation de systèmes biologiques cellulaires.** VCell, (Virtual Cell) est un environnement de modélisation et simulation de systèmes biologiques cellulaires se basant sur des bases de données d’interactions et accessible en ligne [LOEW et SCHAFF., 2018].

nées 1980 quand de nombreuses initiatives ont vu le jour pour fournir des outils pratiques de simulation de l’activité de neurones individuels ou de réseaux [BRETTE et collab., 2007; DE SCHUTTER, 1992; MINGUS]. De nos jours plusieurs logiciels de modélisations gratuits, parfois opensource, sont disponibles, comme NEURON [HINES et CARNEVALE, 2001], GENESIS [BOWER et BEEMAN, 1988], BRIAN [GOODMAN et BRETTE, 2008], NEST [GEWALTIG et DIEMANN, 2007]. Ces environnements ont chacun des spécificités d’applications ou d’usage, la plupart activement maintenus et développés, et repris dans des interfaces pour faciliter la validation et l’exploitation des modèles de manières diverses [DAVISON et collab., 2008; TIKIDJI-HAMBURYAN et collab., 2017]. S’appuyant sur ces développements le début de ce siècle a aussi vu l’essor d’un courant d’envergure visant l’étude théorique non plus de neurones individuels ou de circuits de taille réduite, mais la modélisation et la simulation de structures cérébrales macroscopiques à l’échelle du cerveau pouvant se mettre au niveau des possibilités des expériences de l’imagerie neuronale actuelle. Un changement d’échelle de la modélisation jugé indispensable par le consensus dans la communauté neuroscience sur la nature collective impliquant de très nombreux neurones et différentes zones cérébrales aboutissant aux comportements cérébraux de haut niveau tels que ceux liés à la perception la motricité et la cognition.

Le Human Brain Project lancé en 2005 avec un financement d’un milliard d’euros par l’Union Européenne est la plus connue de ces actions. Son objectif annoncé est de construire une simulation opérationnelle du cerveau humain entier [MARKRAM et collab., 2015]. Des résultats intermédiaires sur le cortex somato-sensoriel du rat ont été publiés en 2015, les figures 3.6 et 3.7 présentent la démarche du projet de modélisation à grande échelle et très détaillée. La démarche implique d’abord la modélisation de l’organisation spatiale des cellules et de leurs connections, et ensuite du traitement de la dynamique d’activité électrique qui s’y déroule. Ces deux niveaux de modélisation de la structure et de modélisation de la dynamique se retrouvent dans les différentes mo-

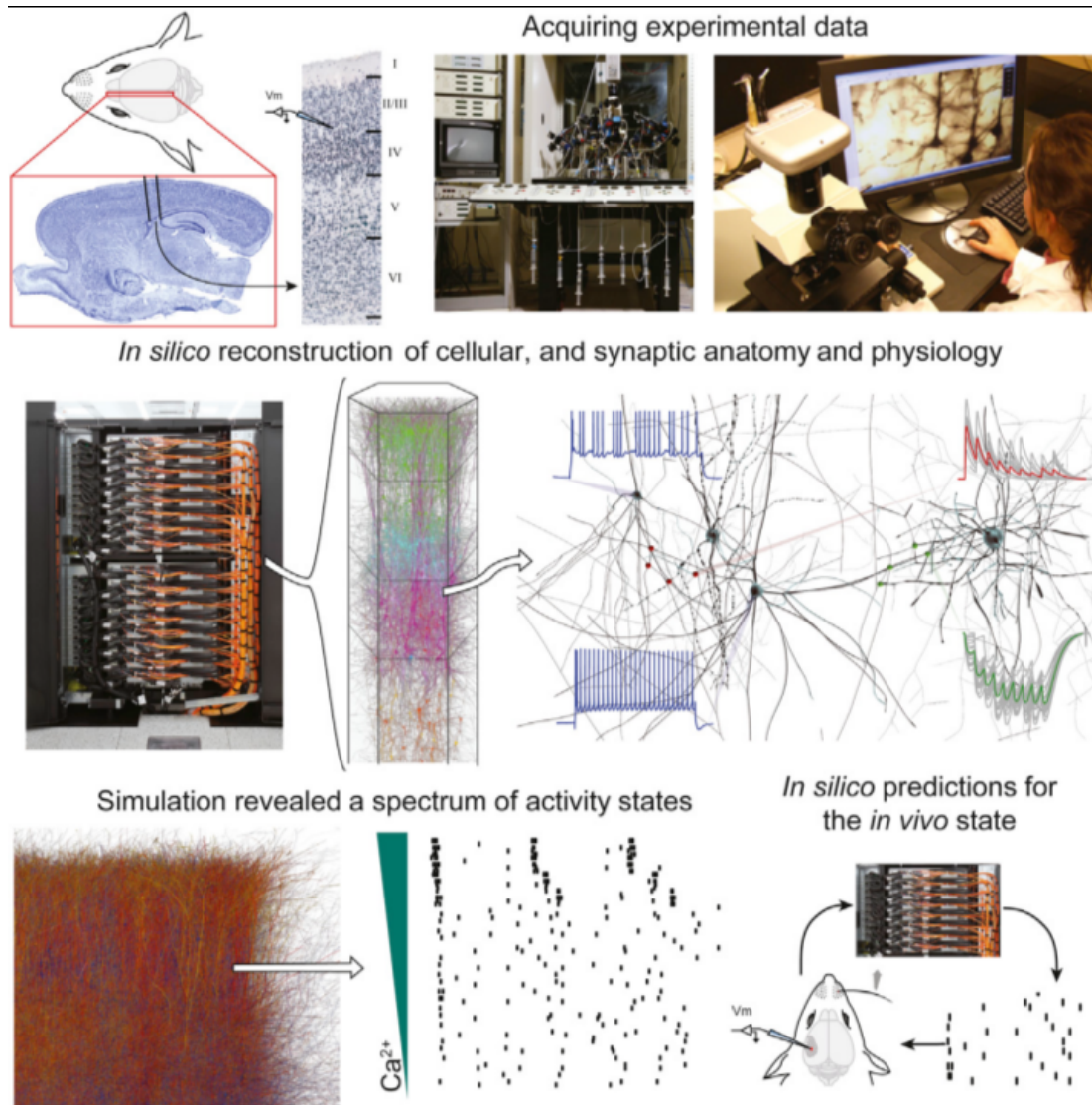


FIGURE 3.6 – **Résumé graphique de l'article de MARKRAM et collab. [2015]** Cet article rapporte en 2015 les premiers résultats du Human Brain Project qui illustre les ambitions de ce projet et des autres initiatives de simulations du cerveau à grande échelle. La figure présente la modélisation et simulation numérique de l'anatomie et la physiologie du cortex somato-sensoriel du rat juvénile entreprise dans l'étude qui reproduit un ensemble d'expériences *in vitro* et *in vivo* sans ajustements de paramètres. Les résultats des simulations suggèrent que les mécanismes cellulaires et synaptiques peuvent reconfigurer dynamiquement l'état du réseau pour permettre une diversité des stratégies de traitement de l'information.

délisations discutées dans ce mémoire, aussi bien dans le champ de la biologie des systèmes que dans les neurosciences.

Plusieurs autres projets ont également vu le jour et sont par exemple discutés dans les revues de De Garis [DE GARIS et collab., 2010] et Michael Breakspear [BREAKSPEAR, 2017]. L'importance des échelles impliquées, de la diversité et du nombre des éléments et des interactions demandent le développement de nouvelles méthodes informatiques de modélisation, paramétrage et exécution des simulations sur des hardware existants. Des défis sur les capacités numériques des superordinateurs qui attirent l'intérêt de l'industrie informatique pour le développement de processeurs spécialisés dit « neuromorphes » s'inspirant de la structure des cerveaux elle-même, comme en témoigne le projet SyNAPSE [DARPA, 2013] entre 2008 et 2016 de l'agence américaine de recherche pour la défense DARPA (Defense Advanced Research Projects Agency) et d'IBM. Au-delà des neurosciences le domaine du calcul neuro-morphique est devenu un secteur de recherche informatique important pour l'exploration de nouvelles architectures de calcul parallèle et de nou-

velles générations de processeurs qui pourraient également trouver des applications pour l'intelligence artificielle.

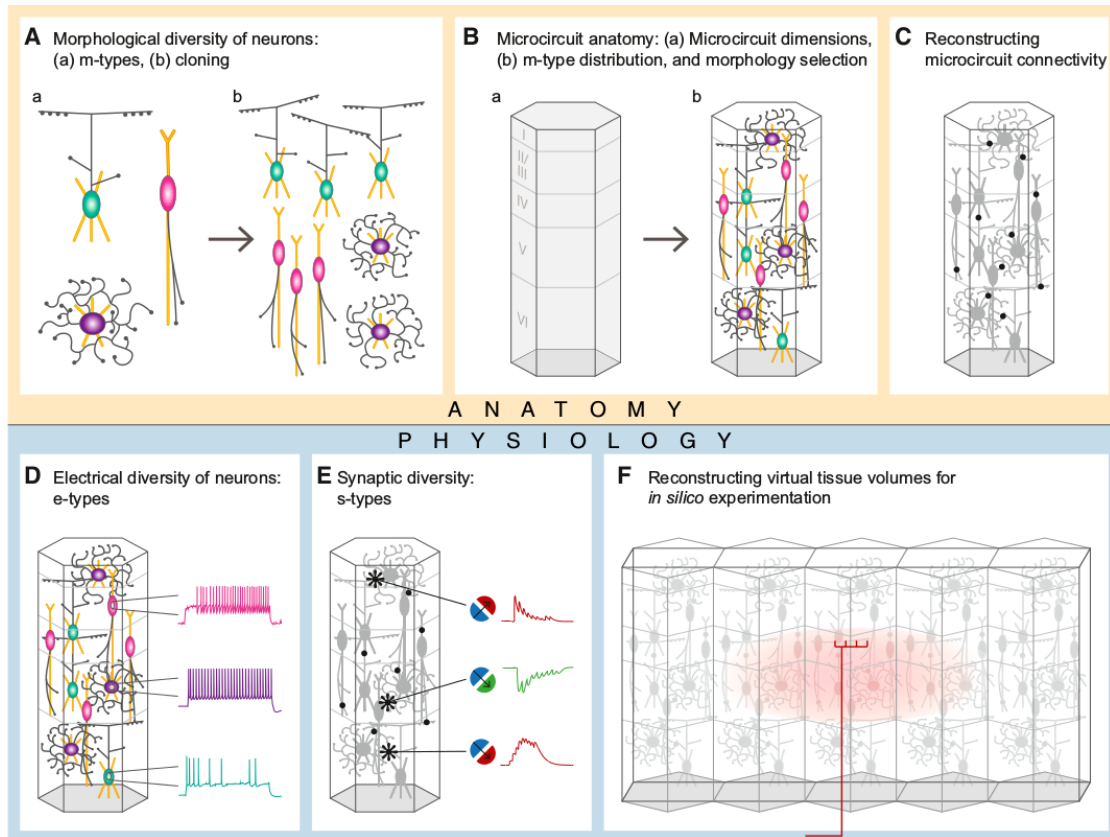


FIGURE 3.7 – Différentes étapes de la modélisation du Human Brain Project. [MARKRAM et collab., 2015]. Le premier niveau anatomique, consiste dans la modélisation précise de l'organisation spatiale et de la diversité des neurones et des circuits de connectivité dans les colonnes 3D du système somato-sensoriel. Il s'agit de l'établissement de la structure du réseau. Le second niveau consiste dans la modélisation de l'activité électrique des différents types de neurones impliqués et de la communication synaptique.

Probablement en raison des investissements en jeu, de la fascination générale autour des neurosciences et d'amalgames du grand public avec l'intelligence artificielle et sa généralisation actuelle, les projets de modélisation et simulation de cerveau sont très visibles et ont donné lieu à de nombreuses polémiques. Y compris au sein de la communauté scientifique il n'existe pas de consensus les manières de procéder ni même sur les objectifs. Des choix de modélisation très différents ont été faits suivant les projets et donné lieu à des polémiques publiques [ADEE, 2009a,b; SEUNG, 2012] et [DATTERI et LAUDISA, 2015] (p. 35). Comme le discute Chris Eliasmith, un de leader du modèle à grande-échelle SPAUN (Semantic Pointer Architecture Unified Network) [ELIASMITH et collab., 2012], dans un article d'opinion critique sur le domaine [ELIASMITH et TRUJILLO, 2014] on peut reconnaître deux approches pour la construction de modèles neuronaux. Du microscopique au macroscopique ou bien la démarche inverse du « haut » vers les « bas ». Le Humain Brain Project en rassemblant un très grand nombre de détails sur les neurones, les canaux ioniques, les types synaptiques, la morphologie spatiale et leurs propriétés physiologiques déduites d'expériences, poursuit clairement la première démarche. Les comportements intéressants devant émerger de la simulation rassemblant toutes les informations disponibles. De nombreuses autres initiatives, dont Spaun, incorporent un travail conceptuel de simplification des modèles des éléments en interaction. Elles utilisent des représentations effectives qui permettent d'avoir accès plus facilement à des simulations de très grande taille, mais également, et peut-être surtout, offrent plus de flexibilité pour étudier des comportements collectifs. On est donc confronté ici aussi à la question classique de la modélisation du niveau de détails nécessaire. On discutera un peu plus bas davantage de la simplification des modèles, mais on peut déjà remarquer qu'un

« bon niveau » de détails n'existe pas en absolu. Le modèle doit servir à un objectif et que c'est une construction du débat scientifique pour la construction d'un savoir. Le travail de simplification lui-même contribuant à rechercher la réponse aux questions posées.

D'usage commun en génie physique et chimique, partie essentielle de la méthode scientifique dans les sciences physiques et de l'univers, la modélisation et la simulation à grande échelle et détaillé s'est imposé dans ces domaines disposant d'un corpus théorique mathématique établi, avec des lois et des équations précises pour la dynamique des variables observables. Dans ces secteurs, un historique de pratique ancienne a fourni des méthodes validées pour maintenant mettre en œuvre avec confiance des modèles des systèmes composites de grandes tailles. Dans les applications les plus complexes de conditions extrêmes, propres par nature à des questions scientifiques, la simulation détaillée reste néanmoins difficile et une entreprise de recherche scientifique plus que d'application d'ingénierie. De nombreux biais de représentation, et sources d'erreurs numériques sont possibles [ARNETT, 2001]. La consistance des résultats doit être analysée de nombreuses manières. Des cas limités vérifiables analytiquement ou expérimentalement doivent être évalués. Bien entendu en définitive la comparaison avec des observations sur des prototypes ou dans des tests dans des situations expérimentales possibles.

3.2.3 Simulation en ingénierie physique versus sciences biologiques

En biologie des systèmes et en neurosciences les modélisations détaillées simulées à grande échelle n'ont pas encore acquis cette maturité. Pour des raisons qui ne sont pas entièrement comparables dans les deux cas. Plus de connaissances quantitatives avec une tradition de mesures électrophysiologiques précises existe en neurosciences, ainsi que des modèles précis et paramétrés de neurones individuels et de synapses. Pour la simulation des réseaux biochimiques d'une cellule par contre, on n'a pas de lois fondamentales mathématiques subjacentes de la biologie moléculaire ou de modèles effectifs parfaitement établis. Les objets en jeu sont une variété de macromolécules et des structures moléculaires sophistiquées dans des milieux complexes et mal définis. La science ne sait pas encore caractériser suffisamment quantitativement ces différents éléments, leurs comportements et les conditions de leur environnement. Les formes mathématiques des interactions sont-elles pertinentes dans les conditions d'études? Quels paramètres choisir? Ces questions se posent car la plupart du temps aucune caractérisation expérimentale n'existe. En dépit du socle quantitatif et théorique mieux défini ces difficultés se posent également en neurosciences. Dans les deux domaines se posent tous les défis des systèmes complexes, diversifiés, de grande taille et que partiellement connus. Qu'ils soient neuronaux ou intra-cellulaires, les systèmes vivants que l'on voudrait modéliser sont ouverts, recevant des signaux en entrée et produisant une sortie. Toutes les influences externes et les échanges avec l'environnement ne sont pas connues. L'écriture de grands modèles réseaux d'interactions, pour de la régulation génétique, pour de la transmission de signaux ou du fonctionnement du métabolisme est certes possible. Elle se fait cependant dans le cadre d'un formalisme encore incertain et toutes les pièces ne sont pas connues invitant à beaucoup de précautions sur les conclusions [HEISKE et collab., 2017; KAWAKAMI et collab., 2016; SCHÜTZHOLD et collab., 2016]. Ces interrogations valent également pour les modèles neuronaux, malgré les progrès récents très importants [KOCH et BUICE, 2015] la caractérisation des éléments en jeu pertinents est largement débattue ainsi que portée des résultats possibles.

Cette image est celle d'entreprises de recherche en cours. Les grandes initiatives scientifiques pour la modélisation et simulation grande échelle en biologie des systèmes et neuroscience dont je viens de faire état ont largement été alimentées par l'évolution technologique impressionnante de l'informatique, hardware et software. Mais elles n'auraient pas pu avoir lieu sans l'historique de recherches théoriques mathématiques dans ces disciplines durant les décennies précédentes qui ont donné les fondements pour les modèles à grande échelle. Des traditions d'études qui, peut-être sous la contrainte des limitations technologiques et de l'absence d'ordinateurs suffisants, ont

privilegé la réflexion sur les ingrédients et les méthodes pertinentes pour établir des modèles suffisamment simples pour être étudiés.

Bard Ermentrout, mathématicien des systèmes dynamiques, expert de longue date de modélisation biologique et auteur du très puissant logiciel de simulation des systèmes dynamiques et d'analyse de bifurcation XPPAUT écrit dans un excellent article sur la question de la complexité et la simplification des modèles « Simplifying and reducing complex models » [ERMENTROUT, 2001] :

« At what point does the model cease to have explanatory value having become too complex to do anything more than simulate it at a variety of parameter and initial conditions?... Large simulations produce a tremendous amount of output and much of it is likely to be useless for the particulars of a given experiment. Finally, for many biological systems, one can only guess at the mechanism. A simulation does not tell you how dependent the behavior is on the particular instance of the mechanism that you have chosen. Only a detailed analysis can tell you that and for complex models and simulations this is difficult at the very least usually impossible. »

Il est temps de discuter de la simplification et de la réduction des modèles de biologie et neurobiologie théorique.

3.2.4 Simplifications et modèles réduits

« Mother Nature will always have more resolution elements than any simulation. Even the most powerful computers are painfully finite with respect to some computation problems. But understanding should be a distillation of information, a compact representation in terms of essential concepts. The art of computing may lie in doing the essential calculations but no more. »

David Arnet, dans « Simulation of Nature By Computer » [ARNETT, 2001]

Limitations calculatoires des modèles détaillés Deux défis se posent chaque fois que l'on entreprend d'établir un modèle. D'une part on est confronté à toutes les inconnues sur le système étudié. Connait-on tous les facteurs internes et externes contribuant au phénomène, les valeurs des paramètres dynamiques et des quantités initiales que l'on va suivre? Dès lors que le système est complexe ce n'est généralement pas le cas, spécialement pour la biologie. D'autre part, on peut aussi faire face à une abondance d'informations, à nombre de détails expérimentaux, de situations et de données. Particulièrement en biologie où un foisonnement de méthodes et de recherches ingénieuses scrutent le vivant de multiples points de vues. La quantité de connaissance est un atout. Le modèle aura des faits sur lesquels s'appuyer et des valeurs pour fixer ses paramètres. La tendance à la simulation discutée dans la section précédente tire partie des progrès de l'informatique pour donner sens aux quantités de connaissances sur les systèmes biologiques et des données que l'on sait produire. Mais la quantité d'information peut se révéler aussi un leurre. Construire un modèle en incorporant tous les détails connus, peut le rendre tout simplement impraticable, impossible à simuler ou calculer avec les ordinateurs actuels. Une limitation que de nouveaux hardware et software pourront dans certains cas permettre de résoudre plus tard. Mais pas nécessairement car certains modèles et algorithmes ont des complexités calculatoires augmentant trop fortement avec le nombre des éléments simulés ne laissant pas envisager la faisabilité de simulations à grande échelle.

Limitations conceptuelles des modèles détaillés Plus fondamentalement, les modèles détaillés ont des résultats volumineux, compliqués, qui peuvent se révéler incapables de fournir une explication intelligible des phénomènes qu'ils abordent avec l'accumulation des faits qu'ils comportent et des données générées. En outre, plus le nombre de détails que l'on souhaite inclure

dans un modèle est élevé, plus grand est aussi le risque de ne pas disposer des paramètres nécessaires ou de les connaître avec une grande imprécision. Inclure tous les détails certes apparaît attrayant, mais comment tenir compte alors des inconnues et manques de fiabilités? Dans une section précédente on a parcouru la démarche de simulation intensive pour certains secteurs d'ingénierie physique. Les constantes physiques, et les caractéristiques des matériaux y sont généralement bien établies. Pour la simulation du vivant on est par contre est confrontés pleinement au problème du manque d'information. Y compris en cette époque de production de données massive en biologie; on le discutera dans le chapitre 4. Ce même problème est à l'origine aussi des tensions évoquées plus haut concernant les projets de simulation neuronale à l'échelle du cerveau entre tenants de modélisations hautement détaillées et partisans d'approches certes à large échelle mais mettant en œuvre une modélisation simplifiée.

Pour une simplification des modèles La démarche de simplification n'est pas qu'une contrainte résultant de l'incapacité à établir un modèle détaillé opérationnel. Même dans le contexte d'approches très détaillée elle est essentielle pour aider à l'interprétation des processus, pour identifier l'importance réelle des différents détails : les précisions supplémentaires dans un modèle sophistiqué sont-elles réellement significatives pour rendre compte de données? Ou bien la qualité d'un accord entre les prédictions du modèle et les données sont-elles dues à un surnombre de paramètres permettant de meilleurs ajustements numériques mais sans signification biologique réelle?

Construire un modèle simplifié consiste à ne sélectionner qu'un ensemble limité d'éléments supposés suffisants pour une représentation opérationnelle conduisant à des résultats et des conclusions pertinentes sur un phénomène étudié. Par l'étude des relations entre les grandeurs intervenant dans un système, l'analyse dimensionnelle dont je parlerai plus bas, fournit d'ailleurs un moyen vers des modèles simplifiés capturant les relations principales qui déterminent les comportements. Des choix de niveaux de détails et d'approximations nécessaires pour permettre l'écriture d'un modèle, mathématiquement sous forme analytique, ou sous une autre écriture formelle par exemple informatique, d'une telle manière qu'il puisse à la fois être résolu et que la solution soit compréhensible. C'est-à-dire que le modèle puisse prédire par le calcul un comportement et que cette prédiction soit interprétable avec un sens logique pour notre entendement. Si le modèle est interprétable, les simplifications et approximations peuvent certes limiter la précision du modèle, mais dès lors que ses prédictions approchent suffisamment la réalité, on aura gagné une intuition sur les mécanismes dominants.

À la recherche de modèles minimaux À l'extrême on peut chercher à simplifier le plus possible une situation réelle. À construire un modèle qui avec le moins d'ingrédients possibles, le moins d'équations, de variables, de paramètres reproduise les traits majeurs du comportement réel. L'objectif sera ici de tenter de capturer dans un cadre mathématique le plus simple possible les relations les plus significatives qui donnent le comportement observé du système modélisé. Ce modèle minimal établira des éléments qui *suffisent*, du point de vue mathématique, pour les propriétés observées. Il n'affirmera pas que ce sont les seuls possibles, et que ce sont bien exactement ceux en action dans la réalité. Le modèle simplifié, réduit ses ingrédients à des interactions effectives qui capturent plusieurs processus distincts de la réalité biologique mais concourant à un effet commun dans la dynamique du système. Le nombre des variables nécessaires correspondra au degré de liberté du système, la dimension de l'espace dit des phases dont chaque point est un état différent du système et dans laquelle son évolution forme une trajectoire. Déterminer la dimension d'un phénomène est important pour s'assurer qu'il n'y ait pas de sur-paramétrage causant artificiellement un bon accord du modèle.

Rechercher le modèle minimal compatible avec la richesse des comportements, le plus réduit possible en termes d'objets mathématiques permet d'identifier les fondements logiques de la dynamique mise en œuvre par le vivant. Le résultat est aussi souvent un modèle plus facile, rapide

à calculer, et pouvant être utilisé dans des applications à plus grande échelle. De tels modèles minimaux sont utiles en tant qu'approche réductionniste en modélisation. Si de cette manière on établit des mécanismes déjà suffisants, alors il sera plus aisé en raffinant les modèles d'identifier des propriétés supplémentaires qui peuvent intervenir et raffiner les résultats ou bien de proposer des explications alternatives.

Bien que la biologie moderne fournit des données en abondances dans de nombreuses situations, celles-ci sont rarement suffisamment précises. Et les mécanismes en jeu sont mal compris et seules des suppositions sont possibles. Le mieux que la modélisation pourra proposer sera alors des relations empiriques, basées sur des observations sans être déduites de mécanismes. Pour limiter l'inconnu, les modèles minimaux sont alors utiles dans ces situations nécessitant l'invention de descriptions efficaces pour reproduire la réalité, mais imparfaitement justifiées. Dans la section suivant je vais discuter de différentes approches pour la construction de modèles réduits.

3.2.5 Approches de simplification

À l'échelle microscopique les processus biologiques génétiques et neuronaux correspondent à des processus stochastiques. Aux échelles moléculaires la biologie est un système thermodynamique hors d'équilibre basé sur la chimie. Comme tous les systèmes impliquant des réactions chimiques, l'évolution temporelle dépendra des rencontres aléatoires entre les réactifs et de probabilités de réactions, ou de fluctuations et d'événements conduisant à l'ouverture ou la fermeture de canaux. Des phénomènes qui d'un point de vue physico-chimiques s'assimilent à des processus de réaction-diffusion, dans lesquelles des espèces chimiques sont sujettes à des mouvements aléatoires thermiques, dans des conditions plus ou moins proches ou éloignées de l'équilibre, et interagissent pour former de nouveaux composés ou changer d'état. Dans le contexte des matériaux la physique statistique a développé de nombreux outils théoriques pour étudier des systèmes à grand nombre de degrés de liberté et leurs propriétés collectives. Ces approches se retrouvent ou servent d'inspiration pour les modélisation dans le cas du vivant.

Modélisation et physique statistique Les systèmes biochimiques ainsi que les systèmes neuronaux font intervenir de nombreux éléments en interaction; on est dans le domaine de la physique statistique et de la construction de modèles à l'échelle de l'ensemble du système avec des variables macroscopiques qui représentent les macro-états capturant un grand nombre de situations microscopiques équivalentes. Les raisonnements et méthodes de la physique statistique s'appliqueront donc pour ces modélisations notamment concernant tous les processus spatiaux de diffusion et de transport. L'idée clé de la physique statistique pour l'établissement de ses équations macroscopiques, équations de diffusion, courants est de calculer des moyennes à partir du dénombrement des états microscopiques. Ses méthodes qui conduisent par exemple à l'équation de Fokker-Planck pour décrire l'évolution des densités de probabilités, ou à l'équation de Langevin sont basées sur l'idée de moyennes et de simplifications par les termes dominants. Le principe d'établir des grandeurs et des relations moyennes permet aussi dans certaines situations lorsque l'espace des processus est homogène d'en faire abstraction. C'est notamment le cas de réactions chimiques dans un volume bien mélangé; c'est-à-dire une répartition uniforme des composantes, qui seront décrits par des concentrations moyennes des espèces présentes. C'est aussi l'effet moyen recherché en modélisant les neurones de populations larges homogènes par des neurones ponctuels dont les potentiels membranaires intègrent les contributions des courants de toutes leurs synapses.

La construction de modèles réduits, de simplification est donc très souvent la recherche, guidée par l'inspiration et l'heuristique, de représentations moyennes et de détermination d'effets dominants qui s'opère sur une variété de formalisations et de méthodologies de la littérature scientifique sur des exemples d'applications divers et qu'il est nécessaire de s'approprier dans ses études ou ses recherches. Des pratiques également guidées par l'inspiration et des inventions

heuristiques de descriptions des situations observées. Selon Bard Ermentrout dans l'article déjà cité à la fin de la section précédente [ERMENTROUT, 2001] on retrouve principalement deux types de moyennes, et voir aussi [ERIKSSON et TEGNÉR, 2016]. L'une consiste dans le calcul d'un effet « moyen » d'une partie du système sur une autre, notamment quand des différences d'échelles importantes, comme des temps caractéristiques très différents, justifient un découplage entre les processus de chaque partie. L'autre démarche, connue comme « l'approximation champ moyen », s'appliquant dans des systèmes à grand nombre d'éléments et consiste à approcher l'effet du système entier sur un élément individuel par un seul effet moyen, permettant de passer d'un problème à de nombreux corps, à un problème à un seul corps. L'établissement de ce champ moyen consiste à négliger les corrélations entre les éléments en supposant que les interactions individuelles sont faibles, mais nombreuses et équivalentes permettant d'établir l'interaction effective moyenne.

Une fois des modèles réduits établis, exploitant des variables moyennes dans une ou quelques équations manipulables, l'analyse de leurs comportements fait appel à des méthodes analytiques dans le meilleurs des cas, numériques sinon, ou des combinaisons des deux suivant les régimes. Les méthodes mathématiques d'analyse des comportements de ces systèmes d'équations sont du champ de la théorie des systèmes dynamiques et de l'analyse de bifurcation qui permettent d'étudier les différents régimes possibles en fonction des paramètres. Celles-ci peuvent aussi permettre d'obtenir dans certains cas des réductions ultérieures des modèles aidant à identifier leurs propriétés mathématiques. Il s'agit de l'analyse dimensionnelle d'une part et d'autre part de la réduction des modèles à leur forme normale.

Analyse dimensionnelle

« I have often been impressed by the scanty attention paid even by original workers in physics to the great principle of similitude. It happens not infrequently that results in the form of « laws » are put forward as novelties on the basis of elaborate experiments, which might have been predicted a priori after a few minutes of consideration. »

Lord Rayleigh [RAYLEIGH., 1915]

Le théorème de Buckingham (aussi connu comme le « théorème-Pi ») établit en effet qu'il est toujours possibles transformer les équations qui décrivent un problème physique en une forme qui n'utilise que des paramètres sans dimension physique obtenues par combinaison des paramètres physiques originaux [MIRAGLIOTTA, 2011; SONIN, 2001]. Ces paramètres sans dimensions n'ont aucun lien entre eux, alors que les paramètres dimensionnés peuvent être reliés par certaines dimensions physiques fondamentales (masse, temps, longueur, charge,...) qu'ils auraient en commun. Une représentation sans dimensions qui minimise encore la description du système car seuls figurent des paramètres indépendants, qui seuls suffisent à déterminer fondamentalement la réponse dynamique. Leur nombre correspond à une mesure de la complexité du problème [BALAGUER, 2013].

Plus qu'uniquement une méthodologie de simplification l'analyse dimensionnelle fournit d'ailleurs une approche puissante pour la construction de modèles simplifiés. Ces approches sont très utiles pour des problèmes que l'on ne sait pas ou mal aborder exactement formellement. L'ouvrage de près de mille pages écrit par Szirtes en fait une démonstration éclatante [SZIRTES, 1998]. En se basant sur l'identification des paramètres d'un système, de leurs dimensions physiques et de l'homogénéité nécessaire des expressions mathématiques de modèles physique, l'analyse dimensionnelle elle permet d'établir des relations caractéristiques entre paramètres et variables. En biologie ces idées ont eu un usage déjà assez ancien notamment en lien avec l'allométrie pour établir des relations entre des propriétés physiologiques d'organismes à des échelles différentes. Leur caractère global, notamment dans les relations en liens avec le métabolisme en font certains des modèles les plus anciens dans l'esprit de la biologie des systèmes [SAVAGEAU, 1979; STAHL,

1961, 1962].

Dans les approches de simplification de modèles, l'analyse des dimensions physiques des variables d'un problème met en œuvre des principes profonds liant réalité physique et structure mathématique. La théorie des bifurcations, qui explique les changements qualitatifs de comportements des systèmes d'équations en fonction des variations de leurs paramètres fait intervenir également des relations mathématiques fondamentales qui peuvent servir à l'élaboration de modèles les plus simples pour reproduire certains types de comportements. Il s'agit en définitive de reformuler un modèle avec le type d'équation mathématique la plus simple connue dont les trajectoires dans l'espace des phases des variables partageant les mêmes symétries. Une des propriétés les plus intéressantes des bifurcations qui contribue à ces simplifications est l'existence d'un nombre restreint de classes différentes (pour un nombre donné de paramètres) [GUCKENHEIMER, 2007] correspondant chacune à certaines symétries. Chaque type de bifurcations est alors caractérisée par une forme normale, ou encore « forme canonique », qui correspond à l'équation la plus simple présentant une structure donnée de changements. Tout modèle d'un système naturel, quel que soit sa nature, sous forme de description détaillée ou réduit, est associé à des classes de bifurcation pour les changements qu'il présente en fonction de chacun de ses paramètres.

Une stratégie de réduction de modèles, consistera à établir des modèles minimaux utilisant les formes normales, en faisant abstraction du cadre réel (la situation physique ou biologique étudiée) mais avec une dynamique équivalente des modèles originaux car choisies avec les même types de bifurcations. Ce programme a notamment été appliqué à plusieurs reprises pour construire des modèles très simplifiés de la dynamique de neurones en réduisant les modèles physiologiquement fondés de type Hodgkin-Huxley à des formes canoniques [HOPPENSTEADT et IZHIKEVICH]. La démarche d'abstraction des propriétés mathématiques essentielles dans la dynamique du modèle détaillé des flux ioniques de sodium et potassium a été suivie juste quelques années après les publications de Hodgkin et Huxley par Richard FitzHugh [FITZHUGH, 1961] en 1961 et Jin-ichi Nagumo, Suguru Arimoto, and Shuji Yoshizawa [NAGUMO et collab., 1962] en 1962. Sur cette idée de nombreux autres modèles ont été introduits dans les décennies suivantes [HINDMARSH et ROSE, 1984; MORRIS et LECAR, 1981] dont les plus récents particulièrement simplifiés et très versatiles par Eugene Izhikevich [IZHIKEVICH, 2003a] et Romain Brette avec Wulfram Gerstner [BRETTE et GERSTNER, 2005] que je reprendrais dans le chapitre 6. Dans tous ces exemples, la réduction des modèles à des formes les plus simples possibles proches de formes canoniques de bifurcation, fournit les outils conceptuels pour vraiment comprendre sur des bases mathématiques les principes mis en œuvre par le vivant.

« **Toy-models** » La nature, la portée et la signification des modèles scientifiques est un sujet épistémologique classique [FRIGG et HARTMANN, 2017]. La notion de modélisation est très générale à la démarche scientifique et technologique avec des formes très variées dont la carte conceptuelle de la figure 3.1 propose une description. Les modèles scientifiques prennent de nombreuses formes, y compris ceux formalisés mathématiquement, et les démarches de constructions sont très variées. Avant de conclure cette discussion sur la simplification des modèles mathématiques il convient de commenter encore d'autres démarches. Certaines modélisations connues peuvent paraître faiblement correspondre aux descriptions précédentes mais néanmoins très importantes dans leur impact scientifique.

Notamment on peut trouver dans la littérature théorique de nombreux secteurs, notamment de physique théorique le concept de « toy-model », modèle jouet, invoqué des descriptions sommaires de phénomènes imaginées suffisantes. Il s'agit de règles empiriques, dépourvues de détails, inventées suivant une intuition de mécanismes abstraits qui pourraient expliquer les traits principaux des systèmes étudiés. Le modèle de Ising est certainement l'exemple le plus marquant de modélisation ad hoc, caricaturale au niveau microscopique, mais fondamental pour la compréhension du ferromagnétisme et de la physique statistique. Un domaine d'ailleurs très friand de toy-models. Sans forcément prétendre être rattachés « mécaniquement » à des situations réelles,

ce genre de modèle espère capturer des principes cachés subjacents à des comportements complexes observés. Le modèle de Kuramoto [KURAMOTO et YOSHIKI, 1975] de la synchronisation collective d'oscillateurs est un autre exemple significatif qui a inspiré la compréhension de systèmes autant physiques, chimiques, biologiques qu'en neurosciences [STROGATZ, 2000, 2003].

On voit sur ces exemples comment, loin d'un programme systématique de modélisation détaillée et de simplification, la création de modèles conceptuels, toy-models, guidée par l'imagination et l'intuition sur des processus mal définis ou encore mal connus peut jouer un rôle central dans le progrès scientifique. Un dernier exemple dans le contexte de ce mémoire, c'est le tout premier modèle connu d'activité neuronale et encore largement utilisé, le modèle « Integrate & Fire » de Louis Lapicque en 1907 [LAPICQUE, 1907] bien avant qu'une quelconque base biophysique des mécanismes de génération des potentiels d'action ne soit connue [ABBOTT, 1999].

3.3 Inférence de paramètres

« *With four parameters I can fit an elephant and with five I can make him wiggle his trunk.* »

Aurait été dit par Enrico Fermi à Freeman Dyson et attribué à John von Neumann [DYSON, 2004], cité dans [GUNAWARDENA, 2010].

Comment connaître les paramètres d'un modèle ? Déterminer les valeurs des paramètres, c'est-à-dire des variables indépendantes d'un modèle hors temps et espace, ainsi que de conditions initiales est indispensables pour l'application des modèles et fait pleine partie de la démarche de modélisation. La paramétrisation de modèles biologiques est un problème d'une part technique, pour la possibilité de mesurer expérimentalement les valeurs ou de les déterminer à partir de données. Mais plus fondamentalement aussi en lien avec l'exigence indispensable à la vie d'être robustes face aux nombreuses fluctuations et accidents possibles. La question de la robustesse biologique doit se vérifier également au niveau des modèles. C'est le sujet de différentes techniques d'analyse de sensibilité de paramètres qui jouent un rôle important pour l'évaluation des modèles. Les propriétés de robustesse et de sensibilité à la variation des paramètres ont des répercussions critiques aussi sur la possibilité même d'en déduire des valeurs fiables à partir de données expérimentales. Une question nommée l'*identifiabilité* des modèles que je discuterai tout à la fin de cette section.

Les modèles mathématiques de biologie sont complexes en raison aussi du nombre de leurs paramètres mal connus. Or il est bien connu que pour utiliser effectivement de tels modèles et calculer des prédictions, il faut disposer de valeurs pertinentes des paramètres, les variables indépendantes du système, ainsi que de connaître un point de départ du système, des valeurs initiales des variables dépendantes. Sans connaissance de ces valeurs, le modèle n'exprime que des relations possibles entre les objets du système. Il est alors une forme fonctionnelle complexe depuis l'espace de paramètres vers l'espace des trajectoires dynamiques possibles. Or suivant les paramètres un même modèle peut présenter des propriétés considérablement différentes. Par exemple, les mêmes deux équations différentielles du modèle réduit d'activité neuronale de Izhikevitch conduit à 8 types de neurones de dynamiques différents suivant les zones de valeurs de 4 paramètres libres, figure 3.8.

Certains paramètres des modèles peuvent être directement connus à partir d'expériences de mesures directes publiées dans la littérature ou réalisées au laboratoire. D'autres nécessitent d'être déterminés par ajustement des paramètres du modèle aux données expérimentales. C'est le champ des nombreux algorithmes d'optimisation, un champ de recherches de mathématiques appliquées et d'intelligence artificielle toujours très vivace. En effet si des méthodes classiques sont bien établies pour des modèles simples bien élevés, des algorithmes plus puissants sont indispensables pour ajuster des modèles complexes, non linéaires et comportent de nombreux paramètres [AMA-

RAN et collab., 2016; VENTER, 2010]. De nombreux travaux se sont ainsi intéressés à l'application de ces méthodes dans le contexte de la biologie des systèmes, dont également certaines approches d'apprentissage automatique auxquelles j'ai participé [PERRIN et collab., 2003; RALAIVOLA et collab., 2003] qui seront abordées dans le chapitre 5 section 5.2.1. On pourra se référer aux revues [ASHYRALIYEV et collab., 2009; BANGA, 2008], ainsi qu'aux travaux auxquels j'ai participé. Les applications sont également nombreuses dans les neurosciences computationnelles [ROSSANT et collab., 2011; VAN GEIT et collab., 2008] avec des développements très actuels [FRIEDRICH et collab., 2014; LYNCH et HOUGHTON, 2015; VAN GEIT et collab., 2016]. Cependant, comme on le verra un peu plus bas, même les algorithmes les plus sophistiqués ne pourront contre la non *identifiabilité* de modèles biologiques due à la structure des modèles et empêchant fondamentalement l'estimation fiable de valeurs. Dans ce cas la structure mathématique du modèle est telle que les observations ne permettent pas d'estimer tous les paramètres inconnus.

3.3.1 Détermination de paramètres de modèles biologiques

La détermination de paramètres pour des modèles biologiques fait face à des défis différents de la modélisation physique. La biologie est plus variés, plus fluctuante. C'est le propre de la vie d'être toujours en changement. Contrairement à la physique et la chimie on n'aura pas de constantes fondamentales propres à la biologie. Il n'existe pas d'équivalent du « Handbook of Chemistry and Physics » [RUMBLE, 2017] à sa 98^e édition en 2017. Seule initiative récente qui s'en approche peut-être la « base de données de nombres biologiques utiles » BIONUMBERS initiée en 2007 par Ron Milo [MILO et collab., 2009].

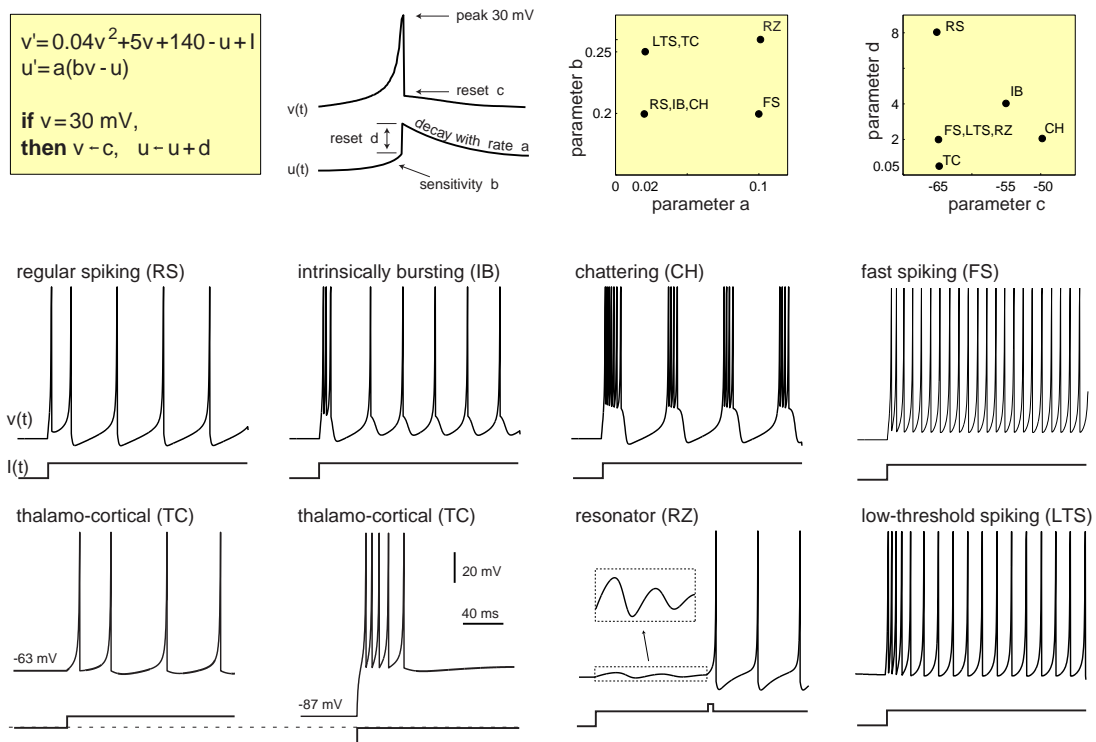


FIGURE 3.8 – **Comportements dynamiques des différents types de neurones.** Les comportements dynamiques des différents types de neurones correspondent à des valeurs différentes des paramètres a, b, c, d du modèle de Izhikevitch décrit par les équations différentielles dans l'encadré en haut à gauche. RS, IB, et CH sont des neurones corticaux excitateurs. FS et LTS sont des interneurons corticaux inhibiteurs. Chaque courbe représente la réponse en voltage d'un modèle de neurones à un pas de courant continu $I=10$ mV représenté sous la courbe. Source : [IZHIKEVICH, 2003b]

Le rapport aux paramètres est différent en biologie des systèmes par rapport aux neurosciences. L'électrophysiologie, que ce soit avec des enregistrements de potentiels par les techniques

de voltage, « current » ou « patch clamp » au niveau du neurone individuel, ou à l'échelle de cerveau avec des techniques d'électroencéphalographie est une discipline quantitative donnant des mesures de haute précision. Les données sont abondantes et de qualité pour faciliter la détermination des paramètres de modèles. Il est aussi possible d'avoir des données dans des conditions particulières pour multiplier les données dans des conditions différentes et permettre par exemple d'isoler les effets de certains paramètres via des drogues bloquant certaines canaux ioniques. C'est essentiellement la démarche suivie par Alan Lloyd Hodgkin and Andrew Fielding Huxley au début des années 50 dans la construction et le paramétrage du modèle des potentiels d'action de l'axone géant du calamar.

Les problèmes étudiés en biologie des systèmes sont moins facilement abordables d'un point de vue quantitatif de la détermination des paramètres. Ils concernent des ensembles de réactions chimiques mal ou partiellement définis se déroulant dans des environnements changeants et mal caractérisés. Des bases difficiles pour être précis! Une des difficultés majeure de modèles de réseaux biochimiques est de quantifier les concentrations moléculaires et les paramètres cinétiques. Les méthodes de biologie à haut débit n'aident pas à déterminer les taux des réactions, d'association et de dissociation ou les constantes d'équilibre nécessaires pour les calculs de dynamiques. Les études de modélisation et simulation les plus quantitatives sont encore focalisées sur des systèmes de taille relativement petite qui peuvent être caractérisés avec des expériences détaillées. Les modèles les plus prédictifs de biologie des systèmes l'ont été dans des systèmes classiques de biologie moléculaire étudiés très en détail comme, par exemple, le circuit de décision lyse-lysogénie du phage lambda [ACKERS et collab., 1982; PTASHNE, 2004; SHEA et ACKERS, 1985], la chimiotaxie bactérienne [BAKER et collab., 2006; BARKAI et LEIBLER, 1998; KOLLMANN et collab., 2005; TINDALL et collab., 2008; VERGASSOLA et collab., 2007], le cycle circadien [GOLDBETER, 2002; GONZE et collab., 2000; KRIETE et collab., 2014], le cycle cellulaire eucaryote [LI et collab., 2004; NOVAK et collab., 2001; TYSON et NOVAK, 2001], la réponse au stress hyper-osmotique [KLIPP et collab., 2005; KÜHN et collab., 2010; SAITO et POSAS, 2012; SCHABER et collab., 2010; ?]. Même dans ces exemples de nombreuses valeurs de paramètres doivent être estimées à partir d'autres systèmes ou ajustées numériquement pour reproduire les comportements observés. Le modèle du cycle cellulaire de NOVAK et collab. [2001] comprend par exemple 47 paramètres de cinétique moléculaire, figure 3.9.

En dépit des progrès de la biologie quantitative dont on a fait état précédemment les méthodes pour des mesurer à haut débit des paramètres biochimiques restent limitées [MAERKL et QUAKE, 2007]. En outre les valeurs peuvent dépendre fortement des conditions particulières les rendant incertaines dans d'autres contextes. Les conditions des réactions biochimiques *in vivo* sont bien différents de celles rencontrées dans des mesures *in vivo*; les premières ont lieu dans des environnement encombrées d'une diversité d'objets biologiques de tailles différentes; les secondes plutôt dans des solutions homogènes et diluées. Pour des modèles complexes, la surface de la fonction mesurant l'écart entre comportement prédit et les données expérimentales, n'est généralement pas régulière. Elle est parsemée de nombreux minima locaux qui piègent les algorithmes d'optimisation. La signification des paramètres ajustée est loin d'être certaine. Le modèle avec ses paramètres ajustés sur un jeu de données expérimentales ne sera fiable que si capable de reproduire des observations indépendantes non utilisées dans l'optimisation. Un cycle entre modélisation, ajustement de paramètres et expérimentation et comparaison est nécessaire.

3.3.2 Modèles stationnaires

Une autre stratégie est de se concentrer sur des modèles avec moins de paramètres, par exemple en ne regardant que les comportements stationnaires. C'est notamment une situation déjà suffisante pour une situation comme l'étude du métabolisme, où l'on s'intéresse plutôt aux propriétés énergétiques des cellules en régime permanent [FELL, 1997]. Dans ce cas on n'a plus besoin de connaître les taux cinétiques de chaque réaction et des rapports, constantes d'équilibre suf-

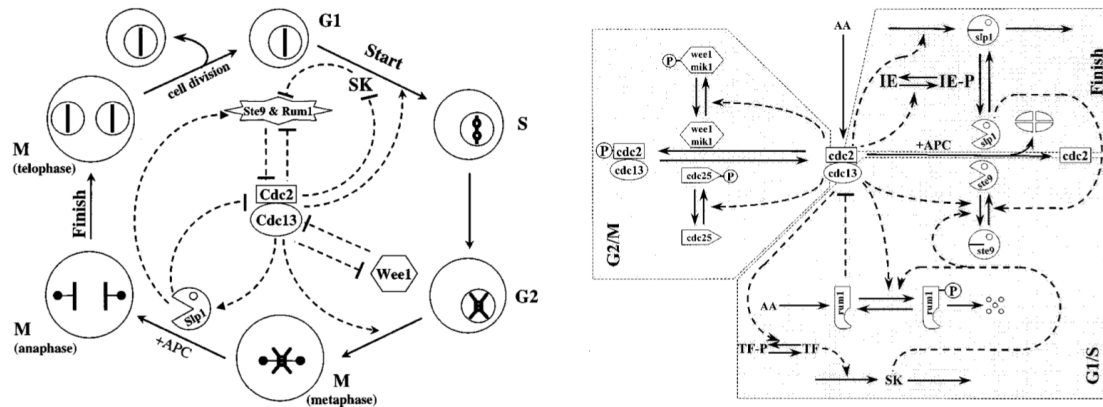


FIGURE 3.9 – Réseau et modèle du cycle cellulaire. *Gauche* : le cycle de division cellulaire eucaryote. Le cercle extérieur montre les étapes majeures de la synthèse de l'ADN et de la mitose. Le diagramme intérieur montre l'état d'avancement du processus. les relations entre les principaux composants moléculaires du circuit central du cycle cellulaire chez la levure. *Droite* : Schéma du circuit central de régulation du cycle cellulaire à levure de fission. Au cœur du cycle se trouve un hétérodimer Cdc2/Cdc13 formant une kinase avec une sous-unité catalytique (Cdc2 qui est une Cdk), et une sous-unité régulatrice Cdc3 (une cycline). Cette kinase est régulée par la protéolyse du composant Cdc13, la phosphorylation de la sous-unité Cdc2 et l'inhibition stœchiométrique du complexe. Ces processus sont organisés en fonction des transitions du cycle cellulaire dans lesquelles ils sont impliqués. Le modèle de [NOVAK et collab. \[2001\]](#) prend la forme d'un système de 9 équations différentielles, 3 équations de conservation et comprend 47 paramètres. Source : [\[NOVAK et collab., 2001\]](#)

fisent. Des méthodes comme l'analyse de flux métabolique (Flux Balance analysis) [\[ORTH et collab., 2010\]](#) et d'analyse de contrôle métabolique (Metabolic Control Analysis) [\[CORNISH-BOWDEN, 2012; FELL, 1992; MORENO-SÁNCHEZ et collab., 2008\]](#) permettent alors de calculer les rendements de voies biochimiques, leur robustesse et sensibilité à des variations. Ces méthodes sont brièvement introduites dans l'encadré 3.1.

La robustesse vis à vis à des changements transitoires et des fluctuations est une propriété reconnue essentielle des systèmes vivant qui doit se retrouver dans les modèles biologiques de fonctions physiologiques [\[KITANO, 2004\]](#). Elle se caractérise par la conservation d'un phénotype face à des perturbations environnementales ou également génétiques. Les types de variations et de fluctuations auxquels sont soumis les êtres vivants sont multiples et posent des contraintes de nature différentes. Les conditions de l'environnement peuvent varier, la dynamique du vivant doit rester stable et le système doit rester robuste aux perturbations des conditions initiales. Des concentrations d'enzyme, pour de structures des enzymes pouvant avoir lieu, le système doit résister à des variations de paramètres. Des variations de phénotypes sont possible autour d'une structure générique conservée, et sont reliés à des variations génotypiques. Dans tous ces cas la robustesse est une question de principe structurel de biologie qui suscite de nombreux débats [\[FÉLIX et BARKOULAS, 2015; GUNAWARDENA, 2010\]](#). Une qualité qui doit également être mise en regard avec l'évolution naturelle. Sur des échelles de temps beaucoup plus longues les systèmes doivent posséder aussi une propriété d'adaptabilité face à des changements à long terme. L'évolutivité des modèles est cependant généralement moins étudiée car hors du cadre de la plupart des modélisations de biologie des systèmes.

Pour les modèles de réseaux biochimiques la robustesse concerne la conservation immédiate d'un type de fonction en dépit de variations et de fluctuations de conditions qui se reflètent dans les paramètres et les conditions initiales. L'exigence de robustesse fournit une réponse commode aux incertitudes sur les paramètres : les systèmes doivent être tolérants aux incertitudes sur les valeurs. Un modèle dont la qualité des prédictions dépendrait de valeurs exactes de ses paramètres et conditions initiales n'est en toute vraisemblance pas robuste et ne saurait être biologiquement acceptable.

Flux Balance analysis (FBA)

Cette méthode mathématique de simulation du métabolisme permet de calculer les flux stationnaires de transformation de la biomasse le long des voies de réactions d'un réseau métabolique. C'est-à-dire les quantités de substrats transformés par seconde le long de chaque réaction du réseau. Elle caractérise les capacités de synthèse d'un réseau métabolique à l'état stationnaire ainsi qu'avec une contrainte d'un objectif. Celui-ci consiste généralement dans la maximisation de la biomasse finale produite.

La « Flux balance Analysis » est numériquement très efficace car elle considère le système opérant à l'état stationnaire, ce qui réduit le modèle à un ensemble d'équations linéaires faciles à résoudre. Comme le système d'équations est généralement sous-déterminé il existe des solutions multiples correspondant à des distributions faisables des flux dans le système [ORTH et collab., 2010; SAJITZ-HERMSTEIN et NIKOLOSKI, 2013].

Metabolic Control Analysis (MCA)

Cette méthode permet d'étudier comment le contrôle des flux et les concentrations intermédiaires des substrats de voies métaboliques est partagé entre les différentes enzymes impliqués dans le réseau [CORNISH-BOWDEN, 2012]. Elle consiste dans des analyses de sensibilité des paramètres de modèles cinétiques du réseau, par le biais de coefficients de contrôle qui quantifient les effets du changement de l'activité d'une enzyme sur le flux métabolique. Elle s'apparente à l'analyse de sensibilité et permet ainsi d'identifier les étapes clés de contrôle de voies de biosynthèse. Cependant son application dépend d'une modélisation cinétique et de la connaissance de valeurs de paramètres [COSTA et collab., 2011].

Encadré 3.1 – La caractérisation de l'influence de chaque réaction enzymatique participant dans un réseau métabolique de biosynthèse est nécessaire pour l'ingénierie rationnelle des systèmes biologiques. Les méthodes résumées ici sont parmi les principaux outils d'analyse.

On comprend alors l'importance particulière de l'application dans les modèles de biologie systémique des techniques d'*analyse de sensibilité* développées de longue date pour les systèmes dynamiques [HAMBY, 1994] et dont l'analyse de contrôle métabolique (MCA) est un cas particulier. Différentes méthodes permettent de tester la réponse du modèle à des changements quantitatifs sur les paramètres et les conditions en entrée. Éventuellement aussi les conséquences qualitatives de changements plus structuraux dans le modèle, comme l'altération de la connectivité de réseaux. Ces analyses aident à évaluer la fiabilité des modèles, mais peuvent aussi être utiles pour différentes questions. Notamment d'identifier les paramètres causant le plus de variations qui nécessitent une caractérisation indépendante, ceux qui ne contribuent que très peu et pourraient être simplifiés du modèle ou encore lesquels sont les plus corrélés entre eux et doivent donc être considérés de concert. Les applications de biologie des systèmes reprennent essentiellement les techniques classiques d'analyse de sensibilité dans le contexte de la biologie des systèmes, leur application dans ce contexte a été revue notamment par Z. Zi [Zi, 2011]. Comme l'indique cet auteur, les modèles mathématiques biologiques semblent être particulièrement robustes aux changements de paramètres. Une étude par GUTENKUNST et collab. [2007] a ainsi repris 17 modèles de biologie des systèmes publiés et étudié leurs variations à des changements de paramètres. De manière surprenante ces modèles se révèlent particulièrement insensibles à la plupart des changements de paramètres. Une caractéristique que ces auteurs nomment comme « sloppy sensitivity » (sensibilité négligée) et qu'ils suggèrent être une propriété partagée commune de la plupart des modèles non-linéaires multi-paramétriques, dont font partie les modèles mathématiques biologiques [GUTENKUNST et collab., 2007].

Cette caractéristique a la conséquence notable de rendre difficile la détermination de manière unique des paramètres d'un modèle par l'ajustement à des données expérimentales, mêmes de qualité et abondantes.

3.3.3 Identifiabilité paramétrique

« Any identification efforts aimed at estimating unidentifiable parameters will fail, leading to wrong estimates, waste of resources, and possibly misleading model predictions. »

VILLAVERDE et collab. [2016]

La propriété pour des paramètres de modèles de pouvoir être estimés, par une quelconque méthode ou algorithme, à partir de données expérimentales porte le nom d'*identifiabilité*. La possibilité de mettre en œuvre un tel programme pour déterminer si tous, ou une partie, des paramètres dépend de la structure du modèle [CHIS et collab., 2011]. Si celui-ci n'est pas identifiable, quelque soit l'algorithme d'optimisation numérique, toute estimation de paramètres qui serait néanmoins réalisée ne sera pas fiable [SACCOMANI, 2015]. D'un point de vue intuitif l'identifiabilité correspond à la propriété pour des valeurs différentes d'un paramètre de produire des valeurs différentes de résultats. Si tel est suffisamment le cas, une valeur optimale non ambiguë du paramètre est possible. Si la variation du paramètre ne produit pas de changement significatif, il n'est pas possible de lui assigner une valeur précise.

Alors que de nombreux efforts ont été dévolus à l'amélioration des algorithmes d'optimisation, la prise de conscience du problème de l'identifiabilité pour les modélisations biologiques est relativement récente [MIAO et collab., 2011; VILLAVERDE et BANGA, 2013], ces 10 dernières années, même si elle était bien présente auparavant dans d'autres domaines. Cependant dans la pratique l'analyse préalable de l'identifiabilité de paramètres est encore rarement effectuée en raison de la difficulté de mise en œuvre calculatoire qui devient rapidement prohibitive pour des systèmes de grande taille [VILLAVERDE et collab., 2016].

Cette situation a des conséquences sur des ambitions de modélisation en biologie. Une généralité de paramètres non identifiables fondamentalement contrarie l'usage de modèles avec des objectifs d'ingénierie inverse, pour déduire les caractéristiques de réseaux de régulation par l'ajustement des paramètres de modèles très génériques d'interaction. L'accomplissement de ce programme était par exemple l'espoir des approches d'inférence de réseaux par l'apprentissage automatique à partir de données à haut débit (séries temporelles de d'enregistrements de puces d'ADN par exemple) pour reconstruire un réseau de régulation qui reproduit les données (un sujet auquel je m'étais intéressé entre 2001 et 2003 et abordé dans la section 5.2.1 du chapitre 5). Sans aller jusqu'à la reconstruction d'un réseau complet, la problématique de l'identifiabilité est présente y compris à des échelles plus modestes de détermination des paramètres de modèles dynamiques ou d'ajustement des poids des interactions.

Il ne faut cependant pas perdre de vue que les modèles mathématiques sont des abstractions de la réalité. Les paramètres sont souvent des grandeurs efficaces capturant plusieurs processus biologiques et sans lien direct avec une grandeur biologique unique. Les valeurs obtenues par ajustement doivent ainsi être mise en regard avec la confiance qu'il est possible d'attribuer dans le modèle à l'interprétation biologique directe de chaque paramètre. La question de l'identifiabilité révèle les limites de la paramétrisation de modèles complexes. Elle incite également à prêter dans la démarche de modélisation une plus forte attention aux incertitudes sur les variables et les paramètres. Une attention qui se manifeste dans l'essor important depuis le tournant du millénaire de la modélisation et de l'inférence Bayésienne qui se diffuse de plus en plus dans la communauté scientifique.

3.3.4 Inférence Bayésienne

L'approche de modélisation statistique dite « Bayésienne » est une évolution majeure de plus en plus diffuse en statistiques de ces vingt dernières années. Elle propose un raisonnement et des

outils de modélisation en rupture conceptuelle avec la vision traditionnelle des probabilités statistiques basées sur la fréquence des événements. La question de l'identifiabilité que nous venons de discuter soulève la difficulté intrinsèque à l'apprentissage de paramètres de modèles. Le cadre de raisonnement Bayésien est complémentaire et donne des outils aux modélisateurs pour mieux prendre en compte et analyser les incertitudes sur les variables et paramètres des modèles.

L'approche Bayésienne traite les variables observées et les paramètres des modèles comme des grandeurs aléatoires caractérisées par des distributions de probabilités qui correspondent à des degrés de confiance *subjectifs* de trouver une grandeur à différents niveaux. À chaque grandeur inconnue on attribue une distribution préalable de la confiance de ses niveaux possibles, qui représente notre état de connaissance initial, avant expériences et observations. Contrairement à la définition classique des probabilités issues d'une répétition d'observations et de différents résultats, l'estimation de la confiance préalable, *a priori*, Bayésienne n'est pas forcément un résultat de fréquence de mesures, elle peut l'être, mais la prise en compte et l'évaluation de tout élément connu sur les grandeurs intervenant dans un modèle. Par exemple dans un modèle nécessitant une durée de demi-vie d'une protéine, juste en se basant sur notre connaissance « établie » dans le domaine, on attribuera une confiance négligeable à une valeur de demie vie de plusieurs mois, mais un degré de confiance plus élevé pour des durées de quelques minutes à quelques jours. Le travail de modélisation s'étendra à cette estimation, sans se baser nécessairement sur l'idée de répétitions d'expériences disponibles pour caractériser cette confiance [CHOW, 2010b].

On représente par α les paramètres d'un modèle que l'on cherche à caractériser en fonction de données disponibles représentées par D . On veut obtenir la distribution de confiance des paramètres tenant compte de l'information dans les données. C'est la probabilité conditionnelle *a posteriori* $P(\alpha|D)$. Le théorème de Bayes permet de relier cette distribution avec la distribution des connaissances *a priori* sur ces paramètres $P(\alpha)$, indépendamment des données, la probabilité des données elles mêmes $P(D)$ et la probabilité des données en fonction des paramètres $P(D|\alpha)$:

$$P(\alpha|D) = \frac{P(D|\alpha)P(\alpha)}{P(D)}.$$

Dans le cadre d'une inférence des paramètres d'un modèle, il s'agit de déterminer les paramètres α qui dans le cadre du modèle optimisent les données. Donc une maximisation de la fonction $P(D|\alpha)$ qui correspond à la vraisemblance classique des paramètres en fonction des données $P(D|\alpha) = L(\alpha, D)$.

Encadré 3.2 – Théorème de Bayes

Les valeurs des paramètres et leur confiance *a priori* établies, le modèle statistique Bayésien permettra de mettre à jour ces informations tenant compte des données des expériences et observations modélisées. Cette actualisation des distributions préalables se base sur le théorème des probabilités de Bayes et conduit à des distributions *a posteriori* des valeurs des grandeurs inconnues. On ne cherche pas à prédire des résultats exacts, une valeur ajustée d'un paramètre inconnu, mais des probabilités de valeurs, parmi toutes les valeurs possibles d'un paramètre lesquelles sont les plus compatibles avec les données.

La modélisation statistique Bayésienne prend donc naturellement en compte l'existence de multiples sources de variabilité et d'incertitude dans des systèmes complexes imparfaitement définis. Elle connaît un essor important également en raison des moyens de calculs accrus et de logiciels d'évaluation.

Ainsi le raisonnement Bayésien peut-être appliqué dans des méthodes d'inférence pour déduire les paramètres inconnus de modèles dynamiques de mécanismes [COELHO et collab., 2011; POOLE et RAFTERY, 2000]. Le point de départ est donc l'établissement pour tous les paramètres recherchés de distributions de confiance sur leurs valeurs possibles. On veut mettre à jour ces connaissances *a priori* sur les paramètres en fonction de données d'observations et des connaissances constituant le modèle dynamique. Comme pour les méthodes d'inférence classiques il

s'agit alors d'optimiser la fonction de vraisemblance des données en fonction des paramètres. Cependant, à la différence des algorithmes d'apprentissage d'optimisation classique, il est nécessaire dans le cadre Bayésien de tenir compte des probabilités *a priori* des paramètres et valeurs initiales du modèle. Les algorithmes nécessitent des adaptations numériquement intensives. L'évaluation des distributions de probabilités des caractéristiques inconnues tenant compte des données des données s'appuie sur des méthodes dites Monte-Carlo Markov Chain (MCMC) pour simuler le système en échantillonnant les variables suivant les distributions *a priori*, propager l'évolution dans le modèle et acquérir une statistique des distributions *a posteriori* des paramètres inconnus [CHOW, 2010a]. La distribution stationnaire de ces simulations Monte Carlo correspond à la distribution de probabilité *a posteriori* recherchée des variables et paramètres inconnus. De l'histogramme des valeurs on pourra extraire moyennes ou médianes comme estimations des paramètres et l'écart-type, ou tout autre largeur comme estimation de l'erreur.

Le résultat clé de l'inférence Bayésienne par rapport à des approches classique est cependant bien la production de distributions *a posteriori* des paramètres inférés, dont la forme renseigne sur la plage de certitude des valeurs apprises tenant compte du modèle et des connaissances préalables.

3.4 Conclusion du chapitre

Dans ce chapitre j'ai discuté longuement la notion de modélisation de processus en sciences et technologie, particulièrement dans les applications au vivant, à savoir la création de représentations calculables des mécanismes en jeu dans un système étudié. Une modélisation qui se distingue de celle très importante également de l'analyse statistique. Les modèles qui nous concernent visent à capturer le fonctionnement de systèmes naturels et technologiques. Les modèles statistiques représentent les caractéristiques de collections de données sur des systèmes fluctuants. Plutôt que des prédictions ces modèles statistiques permettent des *prévisions* qui portent sur les résultats de répétitions *des mêmes expériences*. Les modèles des mécanismes, objets des démarches de modélisation physiques, visent à proposer des explications des comportements et doivent s'ils sont efficaces prédire au-delà des jeux de données ou des informations ayant servi à les établir.

Dans les sciences du vivant la modélisation mathématique de théories des processus a une histoire ancienne mais est toutefois restée une pratique relativement marginale. Depuis le tournant du millénaire, avec le développement des champs de recherche de la biologie des systèmes et de la biologie synthétique la modélisation de processus cellulaires, biochimiques et génétiques connaît un développement important. Ces recherches sont stimulées par la nécessité de rendre compte de l'organisation et de la dynamique complexe des ensembles de constituants en interaction qui composent les réseaux biologiques issus des données de plus en plus disponibles. Les mêmes transformations et nécessités se manifestent également dans le champ des neurosciences, soutenues également par l'accessibilité facile à des moyens de calculs puissants.

La complexité du vivant en termes de nombre et d'hétérogénéité des composants en interaction, ainsi que l'action permanente de l'évolution naturelle source d'adaptation et de diversité posent de nombreux enjeux à la démarche théorique de modélisation. Un des premiers est déjà celui du langage et des connaissances interdisciplinaires nécessaires entre les méthodes et résultats spécialisés de chaque champ de biologie et les compétences mathématiques, informatiques et physiques pour la biologie.

C'est dans ce contexte que se sont situées mes recherches de théorie et modélisation sur des questions assez variées en sciences de la vie dans un premier temps essentiellement dans le champ de la biologie des systèmes et l'expression génétique, et depuis 2012 aussi dans un cadre très particulier des neurosciences portant sur les cultures neuronales. Les deux parties suivantes vont traiter de ces deux champs de recherche. Dans chaque partie un premier chapitre présente les fonde-

ments biologiques et les concepts méthodologiques pertinents à chaque domaine. Ceux-ci sont les bases des travaux de recherche particuliers dans chaque champ et qui sont présentés dans le second chapitre des deux parties.

3.5 Références

- ABBOTT, L. 1999, «Lapicque's introduction of the integrate-and-fire model neuron (1907)», *Brain Research Bulletin*, vol. 50, n° 5-6, doi :10.1016/S0361-9230(99)00161-6, p. 303–304, ISSN 03619230. [82](#)
- ACKERS, G. K., A. D. JOHNSON et M. A. SHEA. 1982, «Quantitative model for gene regulation by lambda phage repressor.», *Proc Natl Acad Sci U S A*, vol. 79, n° 4, p. 1129–1133. [84](#)
- ADEE, S. 2009a, «Cat Fight Brews Over Cat Brain», *IEEE Spectrum*, vol. Posted 23. [75](#)
- ADEE, S. 2009b, «The Cat Brain Cliff Notes», *IEEE Spectrum*. [75](#)
- AMARAN, S., N. V. SAHINIDIS, B. SHARDA et S. J. BURY. 2016, «Simulation optimization : a review of algorithms and applications», *Annals of Operations Research*, vol. 240, n° 1, doi : 10.1007/s10479-015-2019-x, p. 351–380, ISSN 15729338. [82](#)
- ANSYS. «ANSYS Pervasive Engineering Solutions», URL <http://www.ansys.com/>. [70](#)
- ANSYSIS. 2018, «Multiphysics», URL <https://www.ansys.com/fr-fr/products/platform/multiphysics-simulation>. [71](#)
- ARNETT, D. 2001, «Simulation of Nature By Computer», Springer, Dordrecht, p. 11–20, doi :10.1007/978-94-010-0864-8_2. [76, 77](#)
- ASHYRALIYEV, M., Y. FOMEKONG-NANFACK, J. A. KAANDORP et J. G. BLOM. 2009, «Systems biology : Parameter estimation for biochemical models», *FEBS Journal*, vol. 276, n° 4, doi :10.1111/j.1742-4658.2008.06844.x, p. 886–902, ISSN 1742464X. [83](#)
- BAKER, M. D., P. M. WOLANIN et J. B. STOCK. 2006, «Systems biology of bacterial chemotaxis», *Current Opinion in Microbiology*, vol. 9, n° 2, doi :10.1016/j.mib.2006.02.007, p. 187–192, ISSN 13695274. [84](#)
- BALAGUER, P. 2013, «Dynamical systems : dimensionless representation», dans *Application of Dimensional Analysis in Systems Modeling and Control Design*, Institution of Engineering and Technology, p. 23–39, doi :10.1049/PBCE090E_ch3. [80](#)
- BANGA, J. R. 2008, «Optimization in computational systems biology», *BMC Systems Biology*, vol. 2, n° 1, doi :10.1186/1752-0509-2-47, p. 47, ISSN 1752-0509. [83](#)
- BARKAI, N. et S. LEIBLER. 1998, «Bacterial chemotaxis. {U}nited we sense...», *Nature*, vol. 393, n° 6680, doi :10.1038/29881. [84](#)
- BIOVIA. a, «BIOVIA Discovery Studio», URL <http://accelrys.com/products/collaborative-science/biovia-discovery-studio/>. [71](#)
- BIOVIA. b, «SYSTEMS BIOLOGY & INTEGRATIVE THERAPEUTICS», URL <https://www.3ds.com/products-services/biovia/disciplines/systems-biology-integrative-therapeutics/>. [71](#)
- BOWER, J. M. et D. BEEMAN. 1988, «GENESIS», URL <http://genesis-sim.org/>. [73](#)
- BREAKSPEAR, M. 2017, «Dynamic models of large-scale brain activity», *Nature Neuroscience*, vol. 20, n° 3, doi :10.1038/nn.4497, p. 340–352, ISSN 1097-6256. [74](#)

- BRETTE, R. et W. GERSTNER. 2005, «Adaptive exponential integrate-and-fire model as an effective description of neuronal activity.», *Journal of neurophysiology*, vol. 94, n° 5, p. 3637–3642. 81
- BRETTE, R., M. RUDOLPH, T. CARNEVALE et collaborateurs. 2007, «Simulation of networks of spiking neurons : A review of tools and strategies», *Journal of Computational Neuroscience*, vol. 23, n° 3, doi :10.1007/s10827-007-0038-6, p. 349–398, ISSN 1573-6873. 73
- CESSAC, B., H. PAUGAM-MOISY et T. VIÉVILLE. 2010, «Overview of facts and issues about neural coding by spikes», *Journal of Physiology Paris*, vol. 104, n° 1-2, doi :10.1016/j.jphysparis.2009.11.002, p. 5–18, ISSN 09284257. 72
- CHIS, O. T., J. R. BANGA et E. BALSACANTO. 2011, «Structural identifiability of systems biology models : A critical comparison of methods», *PLoS ONE*, vol. 6, n° 11, doi :10.1371/journal.pone.0027755, p. e27755, ISSN 19326203. 87
- CHOW, C. 2010a, «Bayesian parameter estimation | Scientific Clearing House», URL <https://sciencehouse.wordpress.com/2010/11/11/bayesian-parameter-estimation/>. 89
- CHOW, C. 2010b, «MCMC and fitting models to data | Scientific Clearing House», URL <https://sciencehouse.wordpress.com/2010/06/23/mcmc-and-fitting-models-to-data/>. 88
- CLAY, J. R. 2005, «Axonal excitability revisited», *Progress in Biophysics and Molecular Biology*, vol. 88, n° 1, doi :10.1016/j.pbiomolbio.2003.12.004, p. 59–90, ISSN 00796107. 72
- COELHO, F., C. CODEÇO et G. MGM. 2011, «A Bayesian Framework for Parameter Estimation in Dynamical Models.», *PLoS ONE*, vol. 6, n° 5, doi :doi:10.1371/journal.pone.0019616, p. e19616. 88
- COMSOL. «COMSOL Multiphysics», URL <https://www.comsol.com/comsol-multiphysics>. 70
- COMSOL. 2017, «COMSOL Chemical Reaction Engineering Module», URL <https://www.comsol.com/chemical-reaction-engineering-module>. 70
- CORNISH-BOWDEN, A. 2012, «What is Metabolic Control Analysis», URL <http://bip.cnrs-mrs.fr/bip10/mcafaq.htm>. 85, 86
- COSTA, R. S., D. MACHADO, I. ROCHA et E. C. FERREIRA. 2011, «Critical perspective on the consequences of the limited availability of kinetic data in metabolic dynamic modelling.», *IET systems biology*, vol. 5, n° 3, doi :10.1049/iet-syb.2009.0058, p. 157–63, ISSN 1751-8849. 86
- DARPA. 2013, «DARPA SyNAPSE Program», URL <http://www.artificialbrains.com/darpa-synapse-program>. 74
- DATTERI, E. et F. LAUDISA. 2015, «Large-scale simulations of brain mechanisms : beyond the synthetic method», *Paradigmi*, vol. 33, p. 23–46. 75
- DAVISON, A. P., D. BRÜDERLE, J. EPPLER et collaborateurs. 2008, «PyNN : A Common Interface for Neuronal Network Simulators.», *Frontiers in neuroinformatics*, vol. 2, doi :10.3389/neuro.11.011.2008, p. 11, ISSN 1662-5196. 73
- DE SCHUTTER, E. 1992, «A consumer guide to neuronal modeling software», *Trends in Neurosciences*, vol. 15, n° 11, doi :10.1016/0166-2236(92)90011-V, p. 462–464, ISSN 01662236. 73
- DIGICAD. «ANSYS AIM Calcul de Structure, fluide, thermique et électromagnétisme», URL <http://www.digicad.fr/>. 70
- DYSON, F. 2004, «A meeting with Enrico Fermi.», *Nature*, vol. 427, p. 297. 82

- ELIASMITH, C., T. C. STEWART, X. CHOO et collaborateurs. 2012, «A Large-Scale Model of the Functioning Brain», *Science*, vol. 338, n° 6111, doi :10.1126/science.1225266, p. 1202–1205, ISSN 0036-8075. [75](#)
- ELIASMITH, C. et O. TRUJILLO. 2014, «The use and abuse of large-scale brain models», *Current Opinion in Neurobiology*, vol. 25, doi :10.1016/j.conb.2013.09.009, p. 1–6. [75](#)
- ERIKSSON, O. et J. TEGNÉR. 2016, «Modeling and Model Simplification to Facilitate Biological Insights and Predictions», Springer, Cham, p. 301–325, doi :10.1007/978-3-319-21296-8_12. [80](#)
- ERMENTROUT, B. 2001, «Simplifying and Reducing Complex Models», dans *From Computational Modeling of Genetic and Biochemical Networks*, édité par J. M. Bower et H. Bolouri, chap. 11, MIT Press, Cambridge, p. 307–323. [65](#), [77](#), [80](#)
- FÉLIX, M. A. et M. BARKOULAS. 2015, «Pervasive robustness in biological systems», *Nature Reviews Genetics*, vol. 16, n° 8, doi :10.1038/nrg3949, p. 483–496, ISSN 14710064. [85](#)
- FELL, D. 1997, «Understanding the Control of Metabolism Elements of Metabolic Control Analysis (MCA)», , p. 301. [84](#)
- FELL, D. A. 1992, «Metabolic control analysis : a survey of its theoretical and experimental development.», *The Biochemical journal*, vol. 286 (Pt 2, p. 313–30, ISSN 0264-6021. [85](#)
- FITZHUGH, R. 1961, «Impulses and Physiological States in Theoretical Models of Nerve Membrane.», *Biophysical journal*, vol. 1, n° 6, p. 445–66, ISSN 0006-3495. [81](#)
- FRIEDRICH, P., M. VELLA, A. I. GULYAJS, T. F. FREUND et S. KALI. 2014, «A flexible, interactive software tool for fitting the parameters of neuronal models», *Frontiers in Neuroinformatics*, vol. 8, doi :10.3389/fninf.2014.00063, p. 63, ISSN 1662-5196. [83](#)
- FRIGG, R. et S. HARTMANN. 2017, «Models in Science», URL <https://plato.stanford.edu/archives/spr2017/entries/models-science/>. [81](#)
- FUNAHASHIA, A., M. MOROHASHI, H. KITANO et N. TANIMURA. 2003, «CellDesigner : a process diagram editor for gene-regulatory and biochemical networks», *BIOSILICO*, vol. 1, n° 5, doi : 10.1016/S1478-5382(03)02370-9, p. 159–162, ISSN 1478-5382. [71](#)
- DE GARIS, H., C. SHUO, B. GOERTZEL et L. RUITING. 2010, «A world survey of artificial brain projects, Part I : Large-scale brain simulations», *Neurocomputing*, vol. 74, n° 1-3, doi :10.1016/j.neucom.2010.08.004, p. 3–29, ISSN 09252312. [74](#)
- GEWALTIG, M. et M. DIEMANN. 2007, «NEST :NEural Simulation Tool», *Scholarpedia*, vol. 2, p. 1430. [73](#)
- GOLDBETER, A. 2002, «Computational approaches to cellular rhythms.», *Nature*, vol. 420, n° 6912, doi :10.1038/nature01259, p. 238–45, ISSN 0028-0836. [84](#)
- GONZE, D., J. C. LELOUP et A. GOLDBETER. 2000, «Theoretical models for circadian rhythms in {N}eurospora and {D}rosophila.», *C R Acad Sci III*, vol. 323, n° 1, p. 57–67. [84](#)
- GOODMAN, D. et R. BRETTE. 2008, «Brian : a simulator for spiking neural networks in python.», *Frontiers in neuroinformatics*, vol. 2, doi :10.3389/neuro.11.005.2008, p. 5, ISSN 1662-5196. [73](#)
- GU, S., H. SAURO, S. GU et H. SAURO. 2014, «Standards, Platforms, and Applications», dans *Computational Systems Biology*, Elsevier, ISBN 9780124059269, p. 133–167, doi :10.1016/B978-0-12-405926-9.00008-3. [71](#)
- GUCKENHEIMER, J. 2007, «Bifurcations», doi :doi:10.4249/scholarpedia.1517. URL <http://www.scholarpedia.org/article/Bifurcation>. [81](#)

- GUNAWARDENA, J. 2010, «Models in systems biology : the parameter problem and the meanings of robustness», dans *Elements of Computational Systems Biology*, édité par S. H. Muggleton et H. M. Lodhi, chap. 2, John Wiley & Sons, ISBN ISBN : 9780470180938. 82, 85
- GUTENKUNST, R. N., J. J. WATERFALL, F. P. CASEY et collaborateurs. 2007, «Universally sloppy parameter sensitivities in systems biology models», *PLoS Computational Biology*, vol. 3, n° 10, doi :10.1371/journal.pcbi.0030189, p. 1871–1878, ISSN 1553734X. 86
- HAMBY, D. M. 1994, «A REVIEW OF TECHNIQUES FOR PARAMETER SENSITIVITY ANALYSIS OF ENVIRONMENTAL-MODELS», *Environmental Monitoring and Assessment*, vol. 32, n° 2, p. 135–154, ISSN 0167-6369. 86
- HEISKE, M., T. LETELLIER et E. KLIPP. 2017, «Comprehensive mathematical model of oxidative phosphorylation valid for physiological and pathological conditions», *The FEBS Journal*, vol. 284, n° 17, doi :10.1111/febs.14151, p. 2802–2828, ISSN 1742464X. 76
- HINDMARSH, J. L. et R. M. ROSE. 1984, «A Model of Neuronal Bursting Using Three Coupled First Order Differential Equations», *Proceedings of the Royal Society B : Biological Sciences*, vol. 221, n° 1222, doi :10.1098/rspb.1984.0024, p. 87–102, ISSN 0962-8452. 81
- HINES, M. et N. CARNEVALE. 2001, «NEURON : a tool for neuroscientists.», *The Neuroscientist*, vol. 7, p. 123–135. 73
- HODGKIN, A. L. et A. F. HUXLEY. 1952, «A quantitative description of membrane current and its application to conduction and excitation in nerve», *The Journal of Physiology*, vol. 117, n° 4, doi :10.1113/jphysiol.1952.sp004764, p. 500–544, ISSN 00223751. 72
- HOPPENSTEADT, F. C. et E. M. IZHIKEVICH. «Canonical Neural Models», dans *Brain Theory and Neural Networks.*, édité par M. Arbib, second edi éd., chap. Second Edi, The MIT press; Cambridge;. 81
- "IDAHO NATIONAL LABORATORY ". «Multiphysics Object-Oriented Simulation Environment (MOOSE)», URL <http://mooseframework.com/>. 70
- IZHIKEVICH, E. M. 2003a, «Simple model of spiking neurons.», *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, vol. 14, n° 6, doi :10.1109/TNN.2003.820440, p. 1569–1572, ISSN 1045-9227. 81
- IZHIKEVICH, E. M. 2003b, «Simple model of spiking neurons.», *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, vol. 14, n° 6, doi :10.1109/TNN.2003.820440, p. 1569–72, ISSN 1045-9227. 83
- KAN, D. 2012, «What Is Multiphysics? | COMSOL Blog», URL <https://www.comsol.com/blogs/what-is-multiphysics/>. 71
- KAWAKAMI, E., V. K. SINGH, K. MATSUBARA et collaborateurs. 2016, «Network analyses based on comprehensive molecular interaction maps reveal robust control structures in yeast stress response pathways», *npj Systems Biology and Applications*, vol. 2, n° 1, doi :10.1038/npjbsa.2015.18, p. 15 018, ISSN 2056-7189. 76
- KITANO, H. «CellDesigner™ : A modeling tool of biochemical networks», URL <http://www.celldesigner.org>. 71
- KITANO, H. 2004, «Biological robustness», *Nature Reviews Genetics*, vol. 5, n° 11, doi :10.1038/nrg1471, p. 826–837, ISSN 14710056. 85
- KLIPP, E., B. NORDLANDER, R. KRÜGER, P. GENNEMARK et S. HOHMANN. 2005, «Integrative model of the response of yeast to osmotic shock», *Nature Biotechnology*, vol. 23, n° 8, doi :10.1038/nbt1114, p. 975–982, ISSN 1087-0156. 84

- KOCH, C. et M. A. BUICE. 2015, «A Biological Imitation Game», *Cell*, vol. 163, n° 2, doi :10.1016/j.cell.2015.09.045, p. 277–280, ISSN 10974172. 76
- KOLLMANN, M., L. LØVDOK, K. BARTHOLOMÉ, J. TIMMER et V. SOURJIK. 2005, «Design principles of a bacterial signalling network», *Nature*, vol. 438, n° 7067, doi :10.1038/nature04228, p. 504–507, ISSN 0028-0836. 84
- KRIETE, A., R. EILS, J.-C. LELOUP, D. GONZE et A. GOLDBETER. 2014, «Chapter 10 – Computational Models for Circadian Rhythms : Deterministic versus Stochastic Approaches», dans *Computational Systems Biology*, ISBN 9780124059269, p. 183–222, doi :10.1016/B978-0-12-405926-9.00010-1. 84
- KÜHN, C., K. V. S. PRASAD, E. KLIPP et P. GENNEMARK. 2010, «Formal representation of the high osmolarity glycerol pathway in yeast.», *Genome informatics. International Conference on Genome Informatics*, vol. 22, p. 69–83, ISSN 0919-9454. 84
- KURAMOTO, Y. et YOSHIKI. 1975, «Self-entrainment of a population of coupled non-linear oscillators», dans *International Symposium on Mathematical Problems in Theoretical Physics*, vol. 39, Springer-Verlag, Berlin/Heidelberg, p. 420–422, doi :10.1007/BFb0013365. 82
- LAPICQUE, L. 1907, «Recherches quantitatives sur l’excitation électrique des nerfs traitée comme une polarisation», *J. Physiol. Pathol. Gen.*, vol. 9, p. 620–635. 82
- LI, F., T. LONG, Y. LU, Q. OUYANG et C. TANG. 2004, «The yeast cell-cycle network is robustly designed.», *Proc Natl Acad Sci U S A*, vol. 101, n° 14, doi :10.1073/pnas.0305937101, p. 4781–4786. 84
- LOEW, L. M. et J. C. SCHAFF. 2001, «The Virtual Cell : a software environment for computational cell biology», *Trends in Biotechnology*, vol. 19, n° 10, doi :10.1016/S0167-7799(01)01740-1, p. 401–406, ISSN 01677799. 71
- LOEW, L. M. et J. C. SCHAFF. 2018, «VCell Modeling & Analysis Software», URL <http://vcell.org/>. 73
- LOUKATOU, S., L. PAPAGEORGIOU, P. FAKOURELIS et collaborateurs. «Molecular dynamics simulations through GPU video games technologies.», *Journal of molecular biochemistry*, vol. 3, n° 2, p. 64–71, ISSN 2241-0090. 71
- LYNCH, E. P. et C. J. HOUGHTON. 2015, «Parameter estimation of neuron models using in-vitro and in-vivo electrophysiological data», *Frontiers in Neuroinformatics*, vol. 9, doi :10.3389/fninf.2015.00010, p. 10, ISSN 1662-5196. 83
- MAERKL, S. J. et S. R. QUAKE. 2007, «A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors», *Science*, vol. 315, n° 5809, doi :10.1126/science.1131007, p. 233–237, ISSN 0036-8075. 84
- MARKRAM, H., E. MULLER, S. RAMASWAMY et E. AL. 2015, «Reconstruction and Simulation of Neocortical Microcircuitry», *Cell*, vol. 163, n° 2, doi :10.1016/j.cell.2015.09.029, p. 456–492, ISSN 1097-4172. 73, 74, 75
- MATLAB. «FEATool Multiphysics», URL <https://www.featool.com/>. 70
- MIAO, H., X. XIA, A. S. PERELSON et H. WU. 2011, «On Identifiability of Nonlinear ODE Models and Applications in Viral Dynamics», *SIAM Review*, vol. 53, n° 1, doi :10.1137/090757009, p. 3–39, ISSN 0036-1445. 87
- MILO, R., P. JORGENSEN, U. MORAN, G. WEBER et M. SPRINGER. 2009, «BioNumbers The database of key numbers in molecular and cell biology», *Nucleic Acids Research*, vol. 38, n° SUPPL.1, doi : 10.1093/nar/gkp889, p. D750–D753, ISSN 03051048. 83

- MINGUS, B. «Comparison of Neural Network Simulators», URL https://grey.colorado.edu/emergent/index.php/Comparison_{ }of_{ }Neural_{ }Network_{ }Simulators. 73
- MIRAGLIOTTA, G. 2011, «The power of dimensional analysis in production systems design», *International Journal of Production Economics*, vol. 131, n° 1, doi :10.1016/j.ijpe.2010.08.009, p. 175–182, ISSN 09255273. 80
- MORENO-SÁNCHEZ, R., E. SAAVEDRA, S. RODRÍGUEZ-ENRÍQUEZ et V. OLÍN-SANDOVAL. 2008, «Metabolic control analysis : a tool for designing strategies to manipulate metabolic pathways.», *Journal of biomedicine & biotechnology*, vol. 2008, doi :10.1155/2008/597913, p. 597 913, ISSN 1110-7251. 85
- MORRIS, C. et H. LECAR. 1981, «Voltage oscillations in the barnacle giant muscle fiber», *Biophysical Journal*, vol. 35, n° 1, doi :10.1016/S0006-3495(81)84782-0, p. 193–213, ISSN 00063495. 81
- NAGUMO, J., S. ARIMOTO et S. YOSHIKAWA. 1962, «An active pulse transmission line simulating nerve axon», *Proc IRE.*, vol. 50, p. 2061–2070. 81
- NOVAK, B., Z. PATAKI, A. CILIBERTO et J. J. TYSON. 2001, «Mathematical model of the cell division cycle of fission yeast», *Chaos : An Interdisciplinary Journal of Nonlinear Science*, vol. 11, n° 1, doi :10.1063/1.1345725, p. 277, ISSN 10541500. 84, 85
- ORTH, J. D., I. THIELE et B. Ø. PALSSON. 2010, «What is flux balance analysis?», *Nature biotechnology*, vol. 28, n° 3, doi :10.1038/nbt.1614, p. 245–8, ISSN 1546-1696. 85, 86
- PERRIN, B.-E., L. RALAIVOLA, A. MAZURIE et collaborateurs. 2003, «Gene networks inference using dynamic Bayesian networks.», *Bioinformatics (Oxford, England)*, vol. 19 Suppl 2, n° 1, doi :10.1093/bioinformatics/btg1071, p. ii138–i148, ISSN 1367-4811. 83
- POOLE, D. et A. E. RAFTERY. 2000, «Inference for Deterministic Simulation Models : The Bayesian Melding Approach.», *Journal of the American Statistical Association.*, vol. 95, n° 452, doi :doi:10.2307/2669764, p. 1244–1255. 88
- PTASHNE, M. 2004, *A genetic switch : phage lambda revisited*, Cold Spring Harbor Laboratory Press, ISBN 0879697164, 154 p.. 84
- RALAIVOLA, L., B. E. PERRIN, D. F. BUC, A. MAZURIE et BOTTANI. 2003, «Inférence de réseaux d'interactions génique à l'aide de réseaux bayésiens dynamiques.», Journées Post-Génomiques de la DOUA, JPGD'03, La Doua, France. 83
- RAYLEIGH. 1915, «The principle of», *Nature*, vol. 95, n° 2368, doi :10.1038/095066c0, p. 2, ISSN 0028-0836. 80
- ROSSANT, C., D. F. M. GOODMAN, B. FONTAINE et collaborateurs. 2011, «Fitting Neuron Models to Spike Trains», *Frontiers in Neuroscience*, vol. 5, n° February, doi :10.3389/fnins.2011.00009, p. 1–8, ISSN 1662-4548. 83
- RUMBLE, J. R. 2017, «CRC Handbook of Chemistry and Physics, 98th Edition», URL <http://hbcponline.com/faces/contents/ContentsSearch.xhtml;jsessionid=66BEB556E64042684369458F2EBE4842>. 83
- SACCOMANI, M. P. 2015, «Identifiability of Nonlinear ODE Models in Systems Biology : Results from Both Structural and Data-Based Methods», dans *Bioinformatics and Biomedical Engineering*, Springer, Cham, ISBN 9783319164823, p. 650–658, doi :10.1007/978-3-319-16483-0_63. 87
- SAITO, H. et F. POSAS. 2012, «Response to Hyperosmotic Stress», *Genetics*, vol. 192, n° 2, doi :10.1534/genetics.112.140863, p. 289–318, ISSN 0016-6731. 84

- SAJITZ-HERMSTEIN, M. et Z. NIKOLOSKI. 2013, «Structural Control of Metabolic Flux», *PLoS Computational Biology*, vol. 9, n° 12, doi :10.1371/journal.pcbi.1003368, p. e1003 368, ISSN 1553-7358. [86](#)
- SAVAGEAU, M. A. 1979, «Allometric morphogenesis of complex systems : Derivation of the basic equations from first principles.», *Proceedings of the National Academy of Sciences of the United States of America*, vol. 76, n° 12, p. 6023–5, ISSN 0027-8424. [80](#)
- SCHABER, J., M. Å. ADROVER, E. ERIKSSON et collaborateurs. 2010, «Biophysical properties of *Saccharomyces cerevisiae* and their relationship with HOG pathway activation», *European Biophysics Journal*, vol. 39, n° 11, doi :10.1007/s00249-010-0612-0, p. 1547–1556, ISSN 0175-7571. [84](#)
- SCHÜTZHOLD, V., J. HAHN, K. TUMMLER et E. KLIPP. 2016, «Computational Modeling of Lipid Metabolism in Yeast», *Frontiers in Molecular Biosciences*, vol. 3, doi :10.3389/fmolb.2016.00057, p. 57, ISSN 2296-889X. [76](#)
- SEUNG, S. 2012, «Another Perspective on Massive Brain Simulations», *Scientific American*. [75](#)
- SHEA, M. A. et G. K. ACKERS. 1985, «The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation.», *J Mol Biol*, vol. 181, n° 2, p. 211–230. [84](#)
- SIEMENS. «STAR-CCM+», URL <https://mdx.plm.automation.siemens.com/star-ccm-plus>. [70](#)
- SIEMENS DIGITAL FACTORY DIVISIO, S. P. 2016, «Chemical Reaction White Paper», cahier de recherche, Siemens Digital Factory Division., URL <https://mdx2.plm.automation.siemens.com/brochure/chemical-reaction-white-paper>. [70](#)
- LIST OF CHEMICAL PROCESS SIMULATORS, W. «List of chemical process simulators», URL https://en.wikipedia.org/wiki/List_of_chemical_process_simulators. [70](#)
- SONIN, A. A. 2001, «The Physical Basis of DIMENSIONAL ANALYSIS», cahier de recherche, MIT, Cambridge, MA 02139, Department of Mechanical Engineering. URL http://web.mit.edu/2.25/www/pdf/DA_unified.pdf. [80](#)
- STAHL, W. R. 1961, «Dimensional analysis in mathematical biology I. General discussion», *The Bulletin of Mathematical Biophysics*, vol. 23, n° 4, doi :10.1007/BF02476492, p. 355–376, ISSN 0007-4985. [80](#)
- STAHL, W. R. 1962, «Dimensional analysis in mathematical biology. II», *The Bulletin of Mathematical Biophysics*, vol. 24, n° 1, doi :10.1007/BF02477867, p. 81–108, ISSN 0007-4985. [81](#)
- STROGATZ, S. H. 2000, «From Kuramoto to Crawford : exploring the onset of synchronization in populations of coupled oscillators», *Physica D : Nonlinear Phenomena*, vol. 143, n° 1-4, doi : 10.1016/S0167-2789(00)00094-4, p. 1–20, ISSN 0167-2789. [82](#)
- STROGATZ, S. H. S. H. 2003, *Sync : the emerging science of spontaneous order*, 1^{re} éd., Hyperion, New York, ISBN 9780786868445, 338 p.. [82](#)
- SYSTEMES, D. «BIOVIA Biologics – An Innovative Biologics Suite of Process Experiences that Supports the Discovery, Development and Manufacturing of Biotherapeutics», URL <http://accelrys.com/solutions/domains/biologics/>. [72](#)
- SZIRTES, T. 1998, *Applied dimensional analysis and modeling*, December, Elsevier/Butterworth-Heinemann, ISBN 0-07-062811-4, 820 p., doi :<http://dx.doi.org/10.1016/B978-012370620-1.50000-9>. [80](#)

- TIKIDJI-HAMBURYAN, R. A., V. NARAYANA, Z. BOZKUS et T. A. EL-GHAZAWI. 2017, «Software for Brain Network Simulations : A Comparative Study», *Frontiers in Neuroinformatics*, vol. 11, doi : 10.3389/fninf.2017.00046, p. 46, ISSN 1662-5196. [73](#)
- TINDALL, M. J., S. L. PORTER, P. K. MAINI, G. GAGLIA et J. P. ARMITAGE. 2008, «Overview of Mathematical Approaches Used to Model Bacterial Chemotaxis I : The Single Cell», *Bulletin of Mathematical Biology*, vol. 70, n° 6, doi :10.1007/s11538-008-9321-6, p. 1525–1569, ISSN 0092-8240. [84](#)
- TOMITA, M. 2001, «Whole-cell simulation : A grand challenge of the 21st century», *Trends in Biotechnology*, vol. 19, n° 6, doi :10.1016/S0167-7799(01)01636-5, p. 205–210, ISSN 01677799. [71](#)
- TYSON, J. J. et B. NOVAK. 2001, «Regulation of the eukaryotic cell cycle : molecular antagonism, hysteresis, and irreversible transitions.», *J Theor Biol*, vol. 210, n° 2, doi :10.1006/jtbi.2001.2293, p. 249–263. [84](#)
- VAN GEIT, W., E. DE SCHUTTER et P. ACHARD. 2008, «Automated neuron model optimization techniques : A review», *Biological Cybernetics*, vol. 99, n° 4-5, doi :10.1007/s00422-008-0257-6, p. 241–251, ISSN 03401200. [83](#)
- VAN GEIT, W., M. GEVAERT, G. CHINDEMI et collaborateurs. 2016, «BluePyOpt : Leveraging Open Source Software and Cloud Infrastructure to Optimise Model Parameters in Neuroscience», *Frontiers in Neuroinformatics*, vol. 10, doi :10.3389/fninf.2016.00017, ISSN 1662-5196. [83](#)
- VENTER, G. 2010, «Review of Optimization Techniques», dans *Encyclopedia of Aerospace Engineering*, John Wiley & Sons, Ltd, Chichester, UK, ISBN 9780470686652, doi :10.1002/9780470686652.eae495. [83](#)
- VERGASSOLA, M., E. VILLERMAUX et B. I. SHRAIMAN. 2007, «‘Infotaxis’ as a strategy for searching without gradients», *Nature*, vol. 445, n° 7126, doi :10.1038/nature05464, p. 406–409, ISSN 0028-0836. [84](#)
- VILLAVERDE, A. F. et J. R. BANGA. 2013, «Reverse engineering and identification in systems biology : strategies, perspectives and challenges», *Journal of The Royal Society Interface*, vol. 11, n° 91, doi :10.1098/rsif.2013.0505, p. 20130 505–20130 505, ISSN 1742-5689. [87](#)
- VILLAVERDE, A. F., A. BARREIRO et A. PAPACHRISTODOULOU. 2016, «Structural Identifiability of Dynamic Systems Biology Models», *PLoS Computational Biology*, vol. 12, n° 10, doi :10.1371/journal.pcbi.1005153, p. e1005 153, ISSN 15537358. [87](#)
- VLACHAKIS, D., E. BENCUROVA, N. PAPANGELOPOULOS et S. KOSSIDA. 2014, «Current State-of-the-Art Molecular Dynamics Methods and Applications», dans *Advances in protein chemistry and structural biology*, vol. 94, p. 269–313, doi :10.1016/B978-0-12-800168-4.00007-X. [71](#)
- YOULE, P. 1961, «H3. Simulation techniques in chemical reaction engineering», *Chemical Engineering Science*, vol. 14, n° 1, doi :10.1016/0009-2509(61)85083-5, p. 252–257, ISSN 0009-2509. [70](#)
- ZI, Z. 2011, «Sensitivity analysis approaches applied to systems biology models», *IET Systems Biology*, vol. 5, n° 6, doi :10.1049/iet-syb.2011.0015, p. 336–346, ISSN 1751-8849. [86](#)

Deuxième partie

Biologie des Systèmes

Chapitre 4

Systemes de r gulation g n tiques

« Systems biology offers an opportunity to study how the phenotype is generated from the genotype and with it a glimpse of how evolution has crafted the phenotyp »

Marc Kirschner, « The meaning of systems biology », Cell 121 :503-4 2005.

Sommaire

4.1 Introduction au chapitre	102
4.1.1 Organisation et auto-organisation	102
4.1.2 Organisation des systemes de r�gulation	107
4.2 R�seaux biochimiques	109
4.2.1 Diff�rents niveaux de r�seaux biochimiques	111
4.3 Mod�lisation math�matique biophysique de l'expression g�n�tique	114
4.4 Approches math�matiques et r�ductions de mod�les	119
4.4.1 Systemes d'�quations diff�rentielles ordinaires	120
4.4.2 Mod�lisations stochastiques	120
4.4.3 Cin�tique chimique pour les r�seaux de r�gulation g�n�tiques	124
4.5 Conclusion mod�lisation en biologie des systemes	127
4.6 R�f�rences	127

Présentation de la partie

Cette partie a trait à mes travaux à propos de l'organisation et l'activité de gènes et de réseaux génétiques. Ces études se situent dans le contexte des approches théoriques en génomique et plus particulièrement de biologie des systèmes. Ces travaux concernent des études sur l'organisation et le fonctionnement des réseaux génétiques qui sustentent la physiologie et l'évolution des organismes vivants.

Mon regard sur les questions de l'organisation et de la dynamique du vivant est naturellement influencé par mon parcours de physicien et ses approches. Dans le domaine de la modélisation pour la génomique mes travaux se sont ainsi répartis entre des études assez formelles de systèmes minimaux abstraits d'oscillations génétiques, des modélisations et des simulations pour l'exploration d'hypothèses biologiques sur les effets de déséquilibre de nombres de gènes ou de chromosomes, une implication dans des études expérimentales de biologie systémique sur des voies de signalisation cellulaires et des analyses statistiques et modélisation dans le cadre d'une étude expérimentale d'une relation génotype/phénotype.

Les réseaux d'interactions génétiques sont des systèmes complexes dans le sens d'objets comportant de nombreux éléments pouvant être différents, interagissant entre eux de différentes manières et aboutissent à une dynamique spatiale et temporelle variée, pas toujours prédictible directement. Leur étude implique plusieurs étapes. En premier lieu l'établissement des cartes d'interactions. C'est-à-dire la reconstitution informatique des réseaux à partir de données génétiques et de génomique fonctionnelle. Une fois cette information structurale établie, on peut ensuite analyser du point de vue de la théorie des graphes les particularités de l'ensemble des connexions et les caractériser comme un objet biologique virtuel. On peut alors chercher à anticiper les comportements dynamiques soutenus sur ces systèmes via la modélisation et la simulations des dynamiques, ainsi que de chercher à comprendre leur évolution possible.

Mes travaux sur des sujets assez divers en relation avec les thèmes exposés ci-dessus ont été menés dans le cadre de plusieurs collaborations interdisciplinaires avec des publications aussi bien dans des revues de physique que de sciences de la vie. On trouvera une synthèse de ces différents travaux dans le second chapitre de cette partie, chap. 5, présentés en mettant surtout l'accent sur les apports de modélisation auxquels j'ai contribué. Auparavant, le chapitre 4 est une synthèse du cadre scientifique de la biologie des systèmes dans lequel se situe cette partie de mes activités.

4.1 Introduction au chapitre

« It is an important question whether animal morphogenesis consists of a series of stop-and-go genetic instructions mediated by arbitrary biochemical reactions, or whether it is a phenomenon constrained by fundamental physical laws. »

FLEURY et collab. [2016]

4.1.1 Organisation et auto-organisation

La physique statistique, les systèmes dynamiques ont conduit à la découverte de nombreux phénomènes physiques générant des formes et des comportements présentant des variétés de propriétés et de détails à de nombreuses échelles d'observations, aussi bien spatiales que temporelles. Ces phénomènes résultent souvent de la répétition de processus élémentaires généralement simples, dont le déroulement temporel amène cependant à une manifestation macroscopique hautement structurée, communément qualifiée de « *complexe* ». Ce que l'on exprime aussi en parlant d'« *émergence* » de formes et variations sophistiquées, issues du déroulement naturel de règles mathématiques de quelques interactions physiques et chimiques en action. L'organisation observée est spontanée, on parle d'« *auto-organisation* ».

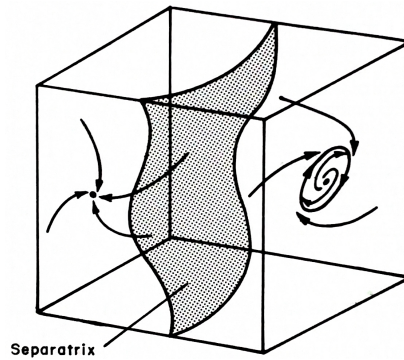


FIGURE 4.1 – **Attracteurs de la dynamique.** Espace des phases à trois dimension avec deux bassins d'attraction séparés par une délimitation, nommée la séparatrice. Au sein de chaque bassin d'attraction, les trajectoires évoluent de manière déterministe vers l'attracteur et restent sur l'attracteur sauf si une perturbation extérieure se produit. L'attracteur à gauche de la séparatrice est un seul point correspondant à un point stable. L'attracteur dans le bassin à droite est une trajectoire qui correspond à un cycle limite oscillant que le système parcourt à répétition. Les attracteurs d'un système dynamique représentant l'évolution temporelle de grandeurs biologiques peuvent correspondre à différents types physiologiques du système, comme des types cellulaires différents. Chaque type cellulaire serait un état stable différent du patron d'expression de plusieurs gènes en réseau. D'après [KAUFFMAN, 1993]p. 176

Ce concept est utilisé pour décrire la formation de structures spatiales ou temporelles, ou les deux, à partir du simple jeu propre d'interactions entre des éléments simples, sans référence à un plan ou à un schéma de construction préalable. Différents principes semblent à l'origine de propriétés d'auto-organisation complexes : boucles de rétro-contrôles, non-linéarité, ruptures de symétrie, dynamique hors-d'équilibres sont autant de facteurs retrouvés dans des systèmes auto-organisés. Pour les physiciens, la biologie est un domaine majeur d'exemples d'auto-organisation. L'étude de la formation des formes biologiques, la morphogénèse, en est le domaine particulier. Pour les physiciens, les formes biologiques sont la manifestation de phénomènes physiques dans le contexte du vivant. Les recherches sur la genèse des formes biologiques se situent donc à l'interface des lois physiques générales, valables pour toute substance et assemblage inanimé ou non, et de la nature propre du vivant. La physique est alors confrontée à la reproduction des organismes et leur développement, à leur robustesse face aux aléas de l'environnement et à l'évolution naturelle de génération en génération. Les formes biologiques que l'on veut expliquer par des principes physiques sont complexes. Elles déterminent parfois les fonctions physiologiques. Et en fin de compte déterminent également les valeurs adaptatives des espèces.

Les formes des physiciens sont des états stables, ou quasiment stables des processus qui les génèrent dynamiquement. Elles correspondent à un état stationnaire de distribution spatiale de la matière dont les caractéristiques sont déterminées par les propriétés thermodynamiques des interactions physico-chimiques propres à chaque système, ainsi que les flux globaux des tissus et leurs contraintes mécaniques. Pour le vivant, dont la nature est intrinsèquement d'être animé, formes et propriétés temporelles sont les attracteurs de la dynamique du vivant. Les propriétés physiologiques sont elles-mêmes aussi des attracteurs dans l'espace des fonctions vers lesquelles convergent spontanément les dynamiques des activités moléculaires et génétique durant le développement des organismes, figure 4.1.

Pour ne citer que quelques exemples de recherches suivant cette vision, dans mon laboratoire Matière et Systèmes Complexes, plusieurs collègues comme Yves Couder, Stéphane Douady, Vincent Fleury, et bien d'autres, ont magnifiquement expliqué la genèse de formes vivantes aussi bien dans le règne végétal qu'animal à partir de mécanismes physiques et thermodynamiques en action. Les approches de mes collègues sont largement fondées sur l'application en biophysique de l'hydrodynamique et la mécanique et de la physique des systèmes non-linéaires et d'un sens aigü de l'observation et de l'intuition.

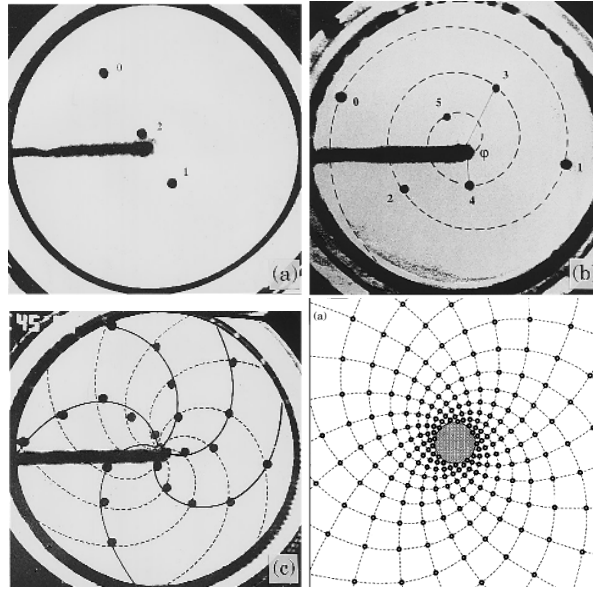


FIGURE 4.2 – **Modèle physique de la phyllotaxie de la fleur de Tournesol** par Yves Couder et Stéphane Douady. Des gouttelettes de ferrofluide sont déposées à une fréquence donnée au centre d'un disque couvert d'huile de silicone dans un champ magnétique constant perpendiculaire (photos issues de [DOUADY et COUDER, 1996]). Les gouttelettes se repoussent mutuellement et sont attirées par le bord du disque. Dès qu'elle est rajoutée une gouttelette s'ajuste dans le champ des répulsions des précédentes et se déplace radialement. Si le taux d'ajout est élevé (figure du bas par rapport à celle du haut) chacune est influencée par plusieurs autres et un schéma organisé en double spirales apparaît. Le processus est capturé dans un modèle dynamique dont la simulation reproduit les patrons typiques sur la fleur de tournesol (bas à droite).

La formation des doubles spirales constituant la fleur de tournesol (figure 4.2) à partir de quelques règles mécaniques de croissance et de contraintes mécaniques est ainsi une des illustrations les plus notoires de la phyllotaxie sous l'angle physique, à savoir le développement des plantes tel que déterminé par des principes physiques [DOUADY et COUDER, 1992, 1996; MALYGIN, 2006]. En effet, par le biais de la résistance des parois cellulaires à la croissance de la pression interne des cellules en développement, la mécanique est le facteur déterminant des formes et des tissus [CORSON et collab., 2009]. Sans exhaustivité on peut encore penser à d'autres exemples, comme la formation des poumons à partir de règles très simples de croissance visco-élastique [CLÉMENT et collab., 2012a,b], figure 4.3. Cet organe d'une complexité extrême, constitué d'une hiérarchie des structures imbriquées depuis l'échelle microscopique à celle macroscopique du thorax, développe spontanément sa forme mathématique fractale sous l'effet d'une poussée des cellules en croissance et de propriétés hydrodynamiques. Grâce aux progrès formidables de l'imagerie, un rôle clé de la physique de la matière, des milieux déformables et des fluides a même été mis en évidence dans le développement d'un organisme dans son ensemble, l'embryogenèse.

Au travers de nombreuses recherches combinant physique et biologie du développement, la morphologie des tissus apparaît de plus en plus comme le résultat de rétro-actions entre un niveau macroscopique et un niveau cellulaire. Au plan macroscopique il peut s'agir de la présence des champs intra- et extra-cellulaires de nature mécanique de contraintes et de tensions, ou bien de gradients chimiques de « morphogènes ». À l'échelle microscopique, au niveau intracellulaire et moléculaire, ce sera par exemple l'organisation coordonnée et liée à la génétique des fibres de microtubules constituant le cytosquelette des cellules et qui détermine les propriétés mécaniques de chaque unité. Toujours au laboratoire MSC, l'équipe dirigée par Vincent Fleury au laboratoire MSC a mis en évidence le rôle central du déroulé d'actions biomécaniques simples et robustes mis en mouvement par la prolifération cellulaire et guidant la formation de la structure fondamentale embryonnaire. Observée sur un modèle aviaire avantageux expérimentalement de l'œuf de poulet, la différenciation cellulaire se déroule simultanément aux mouvements hydrody-

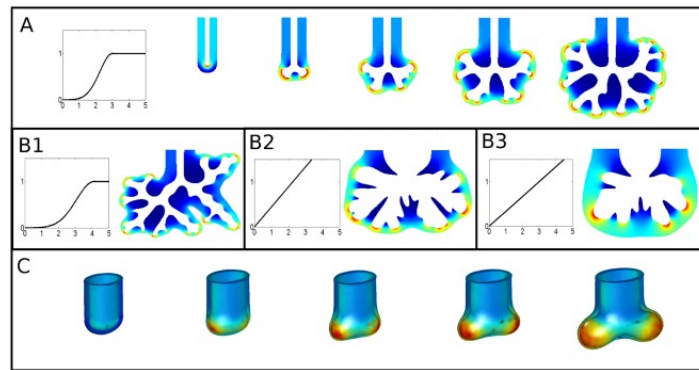


FIGURE 4.3 – **Modèle physique de morphogénèse du poumon.** La structure arborescente des bronches de mammifères est le résultat de branchements successifs de tubes d'épithélium durant le développement crée une structure ramifiée de branches qui s'auto-évitent. Dans le modèle de développement de l'organe de [CLÉMENT et collab. \[2012a\]](#) la croissance de la membrane des bronches, l'épithélium, dépend de la concentration locale dans le mésenchyme, la poche délimitée par le mésothélium où se développent les bronches, d'une molécule de signalisation, la protéine FGF10 sécrétée par les cellules épithéliales aux extrémités des bronches. Le modèle couple la production et diffusion d'un flux de FGF10 et croissance couplée de la membrane. A) Simulation au cours du temps de croissance avec différentes courbes de couplage concentration FGF10 et croissance locale de la membrane. B) Résultats de 3 simulations avec leurs courbes respectives de croissance. C) Séquence de croissance en 3D sur le même modèle. Quand le flux de FGF10 se localise aux pointes des bourgeons il y a création d'un branchement. Source : [\[CLÉMENT et collab., 2012a\]](#)

namiques des cellules dans le tissu induit par les tensions et pressions causées par la prolifération des cellules. Une cascade d'événements physiques de flux de repliement de la masse circulaire presque plate de cellules qui forme chez ces animaux l'embryon initial, la blastula, amène à la forme tri-dimensionnelle structurée de l'embryon qui se poursuivra vers le développement des différents organes (schéma de développement en haut [\[FLEURY et collab., 2016\]](#)). Ces processus sont conduits par une véritable hydrodynamique des cellules dont l'écoulement sous contrainte génèrent des lignes courbes et des tourbillons à l'origine des structures des organes [\[FLEURY et collab., 2015\]](#).

L'embryogenèse doit ainsi combiner des phénomènes de nature différente : les mécanismes biomoléculaires de différenciation cellulaire qui suivent un programme codé dans le génome; l'activation d'étapes de ce programme génétique, allumage ou extinction de certains gènes, en réponse à la présence dans le milieu de molécules facteurs de croissance, connues ou supposées, être présentes selon des champs de concentration dirigés dans l'embryon; et enfin, imprégnant ces processus génétiques, les forces mécaniques s'imposant à toute la matière.

Physique et génétique La contribution des physiciens à la compréhension du développement des organismes a en particulier l'importance de replacer dans le débat leur nature corporelle. Les propriétés d'un corps et les exigences pour le construire dépendent de la matière, de ses propriétés, sa mécanique, ses déformations et écoulements. Ceci contraste avec une vision commune de la morphogénèse, issue du triomphe de la biologie moléculaire, centrée sur les interactions d'interrupteurs génétiques de développement d'activés dans les cellules différemment suivant leur position dans un environnement macroscopique inhomogène de molécules chimiques activatrices. Ces idées fondamentales, remontant à Alan Turing [\[TURING, 1952\]](#), font également intervenir des mécanismes d'émergence de structures chimiques et cellulaires ordonnées. Des principes d'auto-organisation, dont le cœur est l'information plutôt que la mécanique. L'information génétique qui code pour les circuits de contrôle du développement, dont dans cette vision, dépend aussi l'information de position des cellules les unes par rapport aux autres par le biais des champs de molécules chimiques libérés sous contrôle par les cellules. Des approches qui, la plupart du temps, ne considèrent pas la matière.

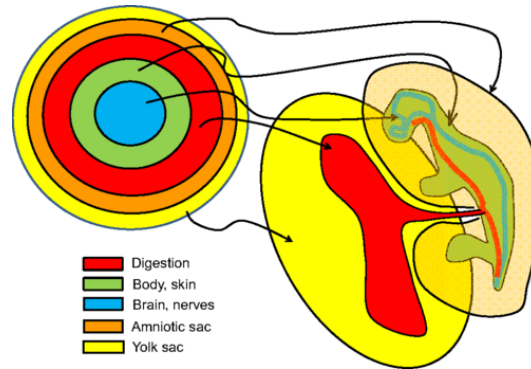


FIGURE 4.4 – **Physique du développement du poulet!** Les flux de cellules dans les tissus en développement ont été observés en détail durant le développement précoce d'embryons de poulet et de crocodile, représentatifs de la classe des vertébrés des amniotes, caractérisés par la présence d'un sac amniotique, protégeant l'embryon dans l'œuf. L'embryon initial est une masse de cellules à la forme d'une galette faites de couches de cellules aux propriétés physiques différentes. Les pressions et tensions s'établissant avec les divisions cellulaires vont conduire à des flux physiques de déplacements cellulaires dépliant les différentes strates vers des structures 3D primitives des organes. Source [FLEURY et collab., 2016].

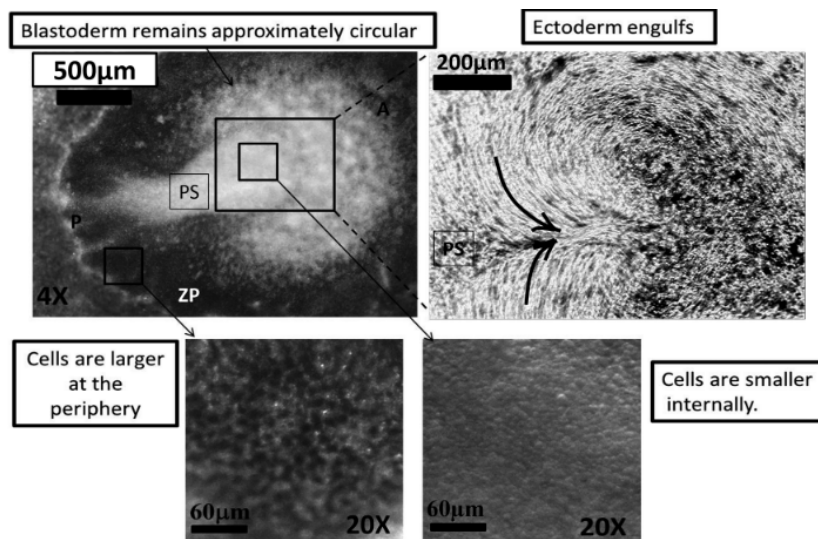


FIGURE 4.5 – **Tourbillons de la vie : flux cellulaires durant l'embryogenèse.** L'image en haut à gauche est une vue du haut de la zone centrale (blastoderm) de la blastula de poulet au bout du premier jour quand l'embryon initialement quasi circulaire démarre une involution. L'illustration en haut à droite combinant une succession d'images met en évidence les déplacements des cellules et l'apparition de flux circulaires comme des vortex de part et d'autre d'un axe de symétrie. La surface périphérique, l'ectoderme, subit un repliement sans que ses cellules soient mélangées avec les autres couches. Les mouvements sont contraints par les propriétés physiques différentes des différentes couches cellulaires (petites cellules au centre du blastoderme, grosses cellules à la périphérie).

Source [FLEURY et collab., 2015].

Les travaux sur la morphogénèse durant la dernière décennie dont je viens de faire état, et bien d'autres, montrent comment le développement n'est pas qu'une affaire d'activité génétique et est la conséquence également du déroulé de mécanismes physiques hydrodynamiques et mécaniques en action. L'expression génétique est pourtant cruciale aux activités cellulaires et au cœur de la vie. Elle l'est notamment dans le développement des organismes durant lequel l'activité des gènes varie avec les changements locaux d'environnement de chaque cellule durant la croissance

et le cycle de vie. Un des grands enjeux actuels pour la compréhension de la formation des organismes est bien celui de comprendre comment ces différents facteurs agissent et se combinent. Si les gènes et leurs interactions sont les agents fondamentaux pour la synthèse des unités de l'organisme; et si la mécanique contraint des flux globaux de la matière vivante en train de se façonner alors toute théorie du développement biologique et du cycle cellulaire devra intégrer génétique et physico-chimie de tout le milieu.

Plusieurs axes de recherches sont dès lors possibles et se sont développés durant la dernière quinzaine d'années dans ce contexte holistique. Des physiciens déjà mentionnés se sont intéressés en premier lieu aux mécanismes physiques d'organisation macroscopique de structures biologiques, sans forcément rechercher la connexion avec la composante génétique des organismes vivants impliqués. D'autres chercheurs venant de la physique comme Emmanuel Farge sont allés chercher à caractériser les mécanismes physiques dans la biologie des cellules et des molécules qui matérialisent un lien entre la contrainte physique et l'activité génétique [FERNANDEZ-SANCHEZ et collab., 2015]. C'est également le sens du travail d'autres collègues physiciens également dans des laboratoires de biologie comme Benoit Ladoux à l'institut Jacques Monod qui montrent comment la vie cellulaire dépend de son entourage et son ancrage mécanique et des interactions dans une véritable société de cellules qui l'entourent [CARSON-PARKER et RICHAR, 2017; LADOUX et MÈGE, 2017]. Enfin on peut se concentrer davantage aux niveaux biochimiques et génétiques et à la manière dont les gènes répondent à des signaux. C'est par exemple le thème de l'équipe au laboratoire MSC de Pascal Hersen qui a développé des systèmes de culture permettant de précisément contrôler et suivre l'activité génétique de cellules en réponse à des changements environnementaux [BISARIA et collab., 2014]. Si je cite ici de nombreux collègues proches, c'est que la biophysique parisienne et spécialement le laboratoire Matières et Systèmes Complexe auquel j'ai la chance d'appartenir est un endroit qui a été très propice à ces idées.

Mes activités dans ces thématiques se sont concentrées sur le niveau des interactions génétiques et des signaux y parvenant. Avant de présenter mes contributions dans le chapitre 5 qui suit, je discuterai auparavant ci dessous le concept de réseau biologique, plus particulièrement de réseau génétique, et les principales approches de modélisation de l'activité génétique s'y développant.

4.1.2 Organisation des systèmes de régulation

Les exemples précédents illustrent les principes d'auto-organisation dans le vivant dans le contexte propre de la morphogénèse, de l'explication des formes des organes et organismes. Mon orientation concernant l'auto-organisation biologique est plus abstraite et focalisée sur le niveau génétique. J'ai parlé avec admiration plus haut des travaux de morphogénèse par la physique de mes collègues du laboratoire. Je n'ai en cela pas suivi le même courant. Mon questionnement a plutôt porté sur l'organisation du système d'information. Dans un premier temps à propos de la structure des systèmes de régulation moléculaires et génétiques, c'est l'objet de la présente partie II de ce mémoire, ensuite à propos des structures neuronales, objet de la partie III finale.

Mécanique et hydrodynamique contrôlent incontestablement aux yeux des physiciens les formes biologiques et leur développement. Mais la biologie incontestablement aussi, a révélé les mécanismes moléculaires et génétiques fondamentaux du métabolisme et de la reproduction dans tout le vivant. Le développement du vivant, la morphogénèse, et l'évolution des organismes est physique et génétique. Le zygote, nouvelle cellule reproduite fertilisée qui se développera en un organisme multicellulaire, n'est pas fait que d'information génétique. Le zygote est code, ordinateur et imprimante 3D! L'ovocyte fertilisé porte l'information génétique de ses progéniteurs mais matérialise également les structures et les machineries pour l'interpréter, synthétiser les assemblages moléculaires et gérer les interactions avec l'environnement. Les principes de la morphogénèse mis en évidence par mes collègues doivent se combiner avec les activations en cascade d'interrupteurs génétiques du développement suivant un programme génétique en interaction

avec un environnement chimique que les cellules contribuent à façonner dans les tissus en développement.

Plutôt qu'être des structures matérielles avec des formes et des propriétés de solides, les objets d'étude génétiques auxquels je me suis initialement intéressé sont des ensembles abstraits de relations biochimiques potentielles, déterminant les processus et transformations de la vie au niveau biomoléculaire. Ce sont les ensembles des interactions moléculaires soutenant la vie cellulaire pilotées et en mises en œuvre par l'information génétique. Des interactions de mieux en mieux connues grâce aux progrès technologiques, tabulées dans des publications et bases de données. Considérées dans leur ensemble elles constituent un réseau indiquant tous les éléments connus pour pouvoir agir les uns sur les autres durant tout le cycle de vie. Ces cartes de réactions biomoléculaires possibles, manifestent la structure de l'organisme au niveau de son fonctionnement biomoléculaire. Elles impliquent les gènes et toutes les biomolécules participant au vivant. Une fois identifiés les composants et leurs possibles interactions, on peut alors regarder ces réseaux comme un objet biologique en soi, avec une forme –abstraite– propre, des propriétés et une évolution liée à la vie des espèces. Différemment des formes physiques précédentes, c'est la morphogenèse et les propriétés de ces formes virtuelles faites de l'ensemble théorique des interactions possibles, qui vont nous concerner ici.

L'étude des réseaux génétiques est au cœur de l'approche de la biologie des systèmes qui généralise la génétique à l'étude d'ensemble de gènes et traite fondamentalement de la relation entre l'information héréditaire, sa matérialisation moléculaire et la dynamique de celle-ci dans chimie du vivant. L'activité des réseaux génétiques est un des versants du développement des organismes qui contribue à déterminer les caractéristiques physico-chimiques des cellules et tissus et à les mettre en branle dans un organisme vivant capable de se reproduire et évoluer. Pour les travaux que je présente plus loin, j'étais fondamentalement motivé par comprendre les formes des réseaux biologiques et les liens entre leur structure et les comportements et l'évolution du vivant. Une problématique interrogeant donc le lien entre le génotype et le phénotype.

Développement du phénotype

Le phénotype correspond aux caractéristiques morphologiques, fonctionnelles, comportementales d'un être vivant à l'échelle de son organisme, à une échelle macroscopique par rapport au niveau moléculaire¹. Le phénotype d'un organisme est alors le résultat stable de l'ensemble des processus biochimiques moléculaires et biophysiques qui se déroulent en son sein, suivant le schéma propre à l'espèce, et modulé individuellement, codé dans l'information génétique et épigénétique. La relation entre les gènes dans un individu et leurs conséquences à grande échelle sur sa nature et ses propriétés peut se concevoir comme une fonction complexe entre génotype et phénotype. L'ensemble de l'information portée au niveau des gènes, le génotype, est alors l'argument d'entrée déterminant les nombreux processus complexes qui mécaniquement conduisent à un individu vivant d'une espèce donnée et doté de caractéristiques personnelles. On a discuté plus haut du rôle central de la physique, de l'action constitutive des forces, des déformations et des déplacements qu'elle induit, et auxquelles évidemment toute la matière biologique est soumise.

L'essentiel des efforts de la génétique et de la biologie moléculaire visent à caractériser dans leurs détails des éléments et mécanismes impliqués dans cette relation. La biologie des systèmes propose la recherche d'une compréhension synthétique de la vie biomoléculaires. Visant au moins une compréhension d'un processus de haut niveau complet au sein d'un organisme, à défaut de l'ambition hors de portée de la compréhension d'un organisme complet au sein d'un même modèle. La génétique et la biologie moléculaire ont relevé en effet d'une démarche réductionniste, d'identification des composants élémentaires et interactions de fonctions plus globales de la vie cellulaire comme le métabolisme énergétique, le cycle de reproduction, la maintenance du soi

1. Quoiqu'il est possible aussi de définir un phénotype moléculaire par rapport aux structures des macromolécules, protéines, associées à des séquences génétiques.

et l'adaptation à des stress extérieurs. Il s'agit alors de décrire au sein d'un modèle intelligible comment ces fonctions touchant l'intégralité de l'organisme opèrent en combinant de manière coordonnée un grand nombre de composants moléculaires et génétiques.

Relation génotype-phénotype Les relations génotype-phénotype (G→P) relient des différences génétiques au sein d'une espèce aux propriétés observées des organismes comme leur forme et leurs comportements. Ces relations sont réalisées par le processus de développement qui traduit l'information génétique héréditaire dans un phénotype lors de la formation de l'organisme. Au cœur des relations G→P se trouvent les réseaux d'interaction biomoléculaires et génétiques dont l'activité qu'ils soutiennent correspond au développement et à la manifestation d'un phénotype. Étudier les réseaux avec le point de vue de biologie systémique permet une approche des relations G→P et de leur relation avec l'évolution naturelle.

L'essentiel des travaux présentés dans le chapitre 5 porte sur des études des réseaux génétiques, plus précisément des réseaux de régulation de l'expression des gènes par les protéines dites facteurs de transcription. Le terme de « gène » sera utilisé ici pour simplement signifier une séquence d'ADN porteuse d'information pour la production de molécules d'ARN messager et conduisant à la synthèse d'une protéine. Par le biais de la machine moléculaire sophistiquée qu'est la polymérase, la séquence codante sur l'ADN est transcrite en la molécule d'ARN messager, qui à son tour sert dans le processus de translation mettant en œuvre l'autre machine remarquable qu'est le ribosome, va permettre la synthèse de la protéine. Pour déterminer si un gène est transcrit ou non des facteurs, dits de transcription, vont intervenir sur une région de l'ADN en amont de la séquence codante et réguler l'expression du gène. Ces facteurs sont des protéines, des produits donc de l'expression de gènes, qui interagissent avec des gènes cibles en se fixant dans les régions de régulation en amont, dites cis-régulatoires, et permettre, faciliter ou inhiber l'action des polymérases sur les gènes qu'ils contrôlent. Donc l'expression de certains gènes peut affecter l'expression d'autres gènes. L'ensemble des relations entre les gènes d'un organisme dont l'activité (l'expression) influence celle d'autres constitue son réseau de régulation génétique.

4.2 Réseaux biochimiques

« I have always been fascinated by enzymes and biochemical pathways. At the start of my career I was interested in research that would allow us to develop diagnostic tools. The need for an overview of the scientific knowledge of the time led to the development of the first Biochemical Pathways. Our knowledge over the last 50 years has grown exponentially and I hope that Biochemical Pathways will continue to expand with our ever increasing knowledge. »

Dr. Gerhard Michal, Éditeur, curateur en chef depuis 50 ans du poster Roche Biochemical Pathways (ex Boehringer).

Avant même la dernière décennie la biologie moléculaire a été très efficace dans l'identification de composants de chaînes de réactions chimiques et de systèmes de régulation dans les cellules vivantes. Ces éléments étaient généralement rassemblés manuellement dans des schémas et diagrammes représentant la synthèse des connaissances du système étudié. En biochimie on connaît depuis longtemps des cartes de grande taille de toutes les réactions métaboliques connues. Le poster des « Biochemical Pathways » dessiné par Gerhard Michal de la société Boehringer Mannheim (aussi connu sous le nom de Boehringer chart) représente le métabolisme général de différentes espèces et organes est un classique de tout laboratoire de biochimie [COLLINS, 2010], la figure 4.6 en présente un extrait. Cette procédure a été étendue et rendue systématique en biologie systémique pour la représentation des informations dans les bases de données génomiques. Comparé à la carte générique de Boehringer, il est maintenant possible de générer des

graphes qui représentent spécifiquement le métabolisme d'un organisme d'intérêt [KANEHISA et collab., 2002; WIXON et KELL, 2000]). Sous le nom de Roche Biochemical Pathways depuis son achat de Boehringer la carte originale du métabolisme est d'ailleurs devenue électronique [MICHAL, 2018] et permet d'y effectuer des recherches avec des mots clef, de zoomer et activer des filtres pour mettre en évidence des éléments d'intérêt.

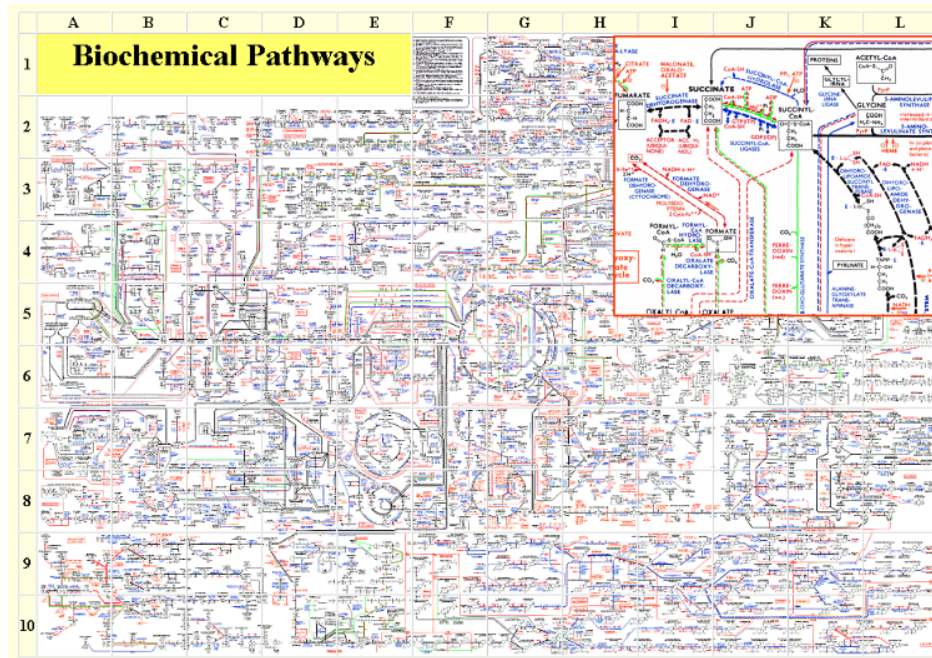


FIGURE 4.6 – **Carte du métabolisme cellulaire.** Le poster des voies biochimiques du métabolisme de la société Boehringer Mannheim dessiné par Gerhard Michal. En haut à droite un agrandissement d'une zone. Les flèches indiquent des réactions catalysées par une enzyme dont le nom est indiqué à côté. Les nœuds sont les substrats et produits successifs des réactions.

De nombreux types de propriétés ont été proposées et étudiées sur des réseaux de tout type. Les principales, concernant la distribution des connectivités, la transitivité des liaisons, le diamètre des réseaux ont été discutées dans le chapitre 2, section 2.6 ainsi que leur interprétations fonctionnelles. En tant que paradigme de complexité, la biologie a été un champ de recherche de prédilection pour ces recherches, ce qui a coïncidé et contribué à l'essor de la biologie des systèmes.

Ainsi plusieurs réseaux d'interaction biochimiques ont été étudiés à partir de données expérimentales. Cela a été le cas de réseaux métaboliques de plusieurs organismes [JEONG et collab., 2000; WAGNER et FELL, 2001], de réseaux des interactions entre protéines [JEONG et collab., 2001; SPIRIN et MIRNY, 2003], de réseaux de régulation de gènes [LOTEM et collab., 2004; SHAI et collab., 2002] dont mes propres travaux [GUELZIM et collab., 2002; MAZURIE et collab., 2005a], du système nerveux d'un animal modèle comme le vers *Caenorhabditis elegans* [WATTS, 1999].

Certaines applications clefs de la génomique visent à caractériser les gènes exprimés dans des types cellulaires, à certains instants et conditions déterminés. Il s'agit de déterminer par exemple comment l'expression génétique diffère d'une cellule à une autre dans des organismes multicellulaires. De préciser encore quelles protéines sont affectées quand un gène est muté ou inactivé. Les réseaux d'interactions biologiques sont les chaînes d'influence entre gènes et entre macromolécules qui déterminent les réponses à toutes ces questions.

Les représentations de ces réseaux doivent donc permettre d'aider à y répondre à condition de connaître le fonctionnement des interactions de disposer des outils d'analyse. Savoir que telle protéine aurait une influence sur telle autre et sur les abondances d'une liste de composés est une première étape. Prédire les répercussions d'un changement sur la première protéine nécessite

aussi de savoir calculer l'effet de chaque lien dans la chaîne de dépendance.

4.2.1 Différents niveaux de réseaux biochimiques

L'objet réseau biomoléculaire reste cependant une abstraction des représentations des tous les composants impliqués dans un processus et des relations possibles entre eux. Ces réseaux biochimiques ne sont pas des réseaux matériels, implémentés physiquement d'une certaine manière comme en technologie les câbles d'un circuits électroniques ou dans le vivant les veines et autres canaux des circuits de transports de fluides (systèmes vasculaires des animaux et des plantes). Les réseaux qui nous intéressent sont des concepts représentant la réalité des interactions au sein des espaces moléculaires auxquels ils se rapportent. Ils capturent notre connaissance des constituants du système étudié et de leurs potentielles interactions. Objets virtuels et basés sur des informations imparfaites les réseaux d'interactions peuvent être biaisés et leur interprétation refléter des idées préconçues sur l'organisation des circuits d'information du vivant. Bien que dans la littérature ce ne soit pas la règle, les conclusions des analyses théoriques doivent rester prudentes, comme l'illustre par exemple la ré-analyse de l'hypothèse des motifs génétiques à laquelle j'ai contribué qui sera discutée dans le chapitre 5, section 5.2.3.

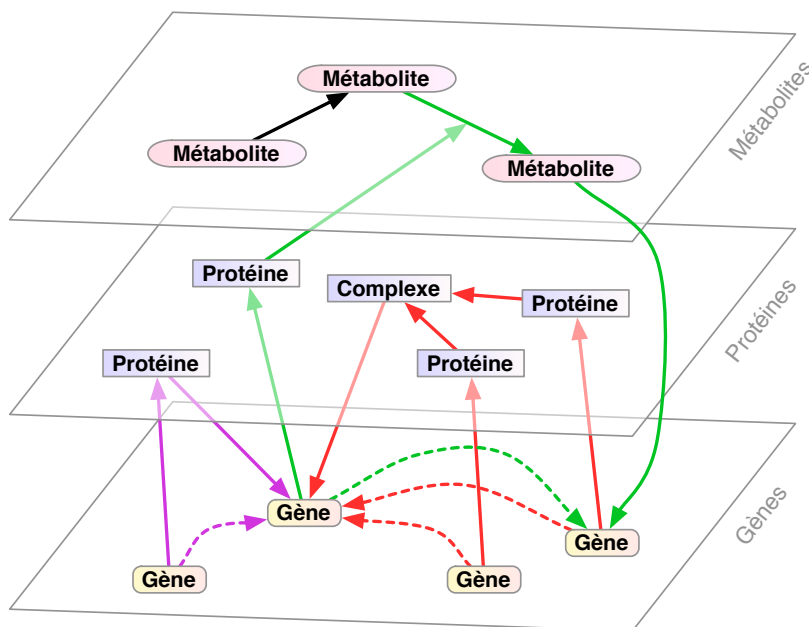


FIGURE 4.7 – **Carte du métabolisme cellulaire.** Les réactions biochimiques cellulaires se décomposent en trois espaces moléculaires principaux en interaction correspondant aux types moléculaires impliqués : réseau transcriptionnel des interactions entre protéines facteurs de transcription et gènes, réseau des interactions protéines-protéines de formation de complexes multi-protéiques et réseaux des interactions entre protéines enzymatiques. Chaque flèche est une interaction (réaction chimique ou enzymatique, ou altération d'une réaction), les pointillés symbolisant des interactions indirectes entre gènes.

Au niveau biomoléculaire on peut distinguer trois niveaux principaux de réseaux qui correspondent à différents espaces moléculaires. Les réseaux métaboliques représentent les métabolites et les réactions chimiques catalysées par les enzymes; un domaine d'étude parfois appelé « métabolome ». Les réseaux protéiques rassemblent les interactions entre protéines conduisant à l'association et dissociation de complexes multi-protéiques et les changements mutuels entre protéines; c'est le champ central du « protéome ». La figure 2.18 montre un exemple des interactions au sein du protéome de la levure. Les réseaux génétiques (aussi appelés réseaux de régulation génétiques ou encore réseaux transcriptionnels) décrivent quant à eux les actions, indirectes, des gènes entre eux. Ce niveau génétique est fondamental. Il est très abstrait aussi car les gènes sont des unités d'information qui n'ont pas d'action directe autrement que par le biais de leurs diffé-

rents produits, leurs transcrits ARN messagers et les traductions en protéines de ceux-ci. Un lien transcriptionnel entre deux gènes est la possibilité de l'action dirigée d'un premier gène régulateur sur l'activité d'un second gène cible. L'expression du gène régulateur synthétise des protéines dotées de la propriété d'affecter l'expression du gène cible. On nomme « facteurs de transcription » les protéines, et leurs gènes associés, qui ont une telle capacité car ils contrôlent la transcription des gènes ciblés. Le niveau des réseaux centré sur les gènes et les facteurs de transcription est au cœur de l'objet « génome ».

Ces trois niveaux de réseaux biochimiques s'entrecroisent intimement. Déjà les gènes n'agissent sur d'autres que par le biais de leurs protéines. En outre ce sont des complexes associant des protéines de types différents qui constituent parfois les unités de contrôle de la transcription d'une cible. Les protéines peuvent avoir besoin d'avoir été « activées » à leur tour par des molécules du métabolisme pour prendre une conformation adéquate leur permettant de se lier entre elles ou sur l'ADN et agir de régulateur.

Les réseaux biochimiques : produits de l'évolution Un réseau biologique, qui regroupe l'ensemble des composants et leurs interactions impliqués dans un processus du vivant, d'une fonction physiologique, peut être considéré comme un objet biologique en soi et résultat de l'évolution naturelle. Sa structure peut donc être questionnée à la recherche de principes biologiques sous-jacents. Si la vie est la manifestation de l'activité moléculaire et génétique dans les cellules, si cette activité est le résultat des réactions biochimiques et de régulation génétiques formant les interactions des réseaux, alors la structure de ceux-ci a été conditionnée et déterminée par l'évolution et l'adaptation aux contraintes de l'environnement. La structure des réseaux biologiques n'est donc probablement pas désordonnée. On s'attend plutôt qu'elle possède des propriétés particulières, résultat de l'adaptation. De ce point de vue les réseaux devraient posséder des qualités propres à la vie. Être robustes aux erreurs et accidents auxquels la reproduction et la vie cellulaire sont sujets en permanence et avoir la capacité de compenser ou de limiter les conséquences de défauts au niveau génétique. Faciliter des mécanismes de réparation essentiels à la survie ou bien des systèmes de blocage de la propagation des conséquences d'un défaut. Et finalement avoir une structure qui permette et facilite l'évolution naturelle elle-même. La sophistication d'un réseau ne doit pas se faire au prix du sacrifice de la créativité possible pour de nouvelles adaptations.

Des réseaux particulièrement complexes... Ces réseaux biochimiques sont en réalité encore plus complexes que dans la brève description ci-dessus. Les mécanismes d'interaction et les agents de régulation de l'activité génétique et biochimique dans les cellules sont plus nombreux et variés. Je n'ai pas mentionné ainsi les régulations au niveau de l'épissage des ARN messagers, ni les effets en lien avec la maturation des protéines. Également le rôle fondamental de l'accessibilité des séquences d'ADN des gènes déterminé par l'empaquetage des chromosomes, dont le rôle actif dans le développement et la vie des organismes a été de plus en plus reconnu ces vingt dernières années et constitué le domaine d'étude de l'épigénétique. Des découvertes fondamentales sont aussi venues se rajouter dans ce tableau depuis les années 2000 avec la mise en évidence des régulations par des micro-ARN exprimés à partir de séquences d'ADN non codantes pour des gènes. On pourrait d'ailleurs (on le devrait probablement) encore inclure d'autres effets non liés à un nouveau type de molécule de régulation mais à la localisation spatiale des gènes sur les chromosomes, et des chromosomes eux-mêmes au sein des cellules ou des noyaux. L'architecture des génomes joue un rôle. C'est le cas bien connu par exemple de la relation entre la localisation groupée des gènes de la famille Hox dans les chromosomes et leur expression dans l'organisme [OULION et collab., 2012]. Ces gènes intervenant dans le développement se retrouvent dans des groupes de plusieurs répétitions se suivant séquentiellement sur l'ADN. L'architecture des chromosomes au sein de la cellule pour les procaryotes ou du noyau pour les eucaryotes semble également jouer un rôle sur l'activité génétique. Celle-ci affecte dynamiquement la co-localisation de gènes dans des sous-régions pour rapprocher ou éloigner des gènes devant être associés ou non dans des processus

affectant l'expression [JUNIER et collab., 2010; NICODEMI et PRISCO, 2009; RAJAPAKSE et collab., 2009]. Ces différentes propriétés structurales ne sont pas capturées dans les réseaux génétiques qui peuvent ainsi être biaisés d'informations essentielles pour l'analyse des activités de leurs composants.

Diviser pour comprendre : à la recherche de modularité

Face à la variété des processus et le nombre des agents impliqués l'ambition de représenter et de donner un sens au réseaux de régulation biochimiques peut paraître illusoire. Face à cette impression la recherche en biologie des systèmes a souvent fait le choix implicite de l'ignorer et d'organiser des recherches là où c'est possible, en s'efforçant de limiter les systèmes à des sous-ensembles, espérés suffisamment homogènes pour qu'ils aient un sens en les isolant du reste. On parle de biologie des systèmes, pas de biologie du système. Faire de la recherche là où c'est possible. Dans le contexte de la biologie des systèmes cela a souvent consisté à restreindre artificiellement l'objet d'étude au sein de chaque espace moléculaire mentionné plus haut. Chacun de ces niveaux se rapportant à des outils expérimentaux et d'analyse spécifique à la nature des molécules qui le constituent.

Notre attendu, subjectif, d'une organisation modulaire est d'ailleurs peut-être dicté lui-même par la spécificité de l'outillage technologique à disposition. Chaque technique est spécialisée pour un type de molécules donnée. Notre regard est focalisé chaque fois sur un niveau de la réalité biologique. L'analyse des réseaux biochimiques devrait donc commencer par l'identification et la délimitation du système, ses éléments internes et ses entrées et sorties. On va supposer la fonction physiologique, ou le processus étudié comme isolable, formant un module en communication avec le reste de l'organisme et l'environnement via des canaux d'entrée et de sortie bien identifiés. Les nombreux cas d'application étudiés avec un certain succès semblent justifier de tels découpages. Nous devons cependant garder à l'esprit qu'il s'agit peut-être d'une ségrégation arbitraire dictée par notre besoin humain de délimitation des sujets pour chercher une compréhension. Nous apprécions ce que nous pouvons comprendre. Enfin, les systèmes d'études doivent être bien définis en fonction aussi du contexte. Dans un autre environnement, dans un autre stade du développement, les éléments du système considéré peuvent aussi intervenir dans d'autres processus et le réseau initialement délimité ne plus être pertinent tel quel.

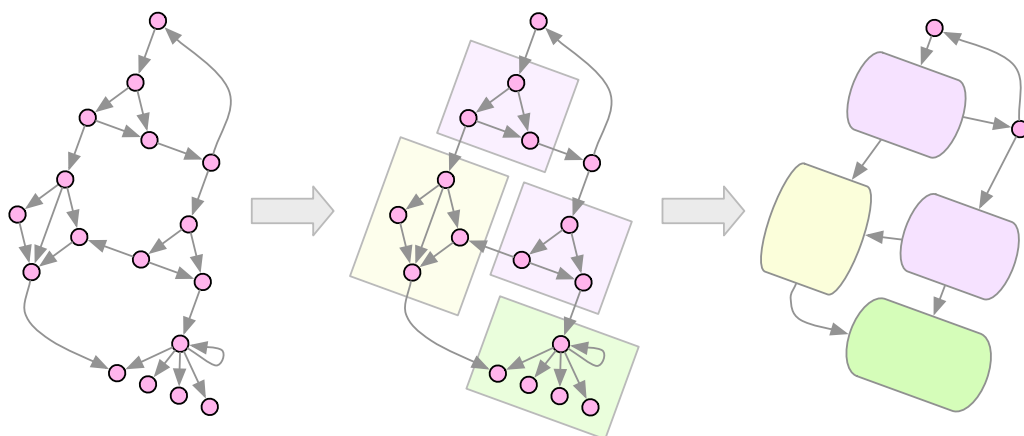


FIGURE 4.8 – À la recherche de modularité dans les réseaux. À gauche, exemple de réseau détaillé dans lequel des sous-structures répétées ou particulières sont recherchées (centre), aboutissant à la vision macroscopique modulaire du réseau (droite).

Plusieurs réseaux biochimiques semblent avoir un niveau de clustering particulièrement élevé : la régulation transcriptionnelle [GUELZIM et collab., 2002], les interactions protéines-protéines [HALL-SWAN et collab., 2018; PEREIRA-LEAL et collab., 2003], le réseau métabolique [RAVASZ et col-

lab., 2002]. Dans un article prospectif dans Nature qui a eu un impact important sur la biologie des systèmes et les études des réseaux biologiques [HARTWELL et collab.](#) ont suggéré que l'organisation modulaire que manifeste le degré de clustering important est nécessaire pour l'implémentation de fonctions de traitement d'information au niveau moléculaire pour la transmission et l'amplification de signaux, le filtrage de bruit, la réalisation de choix et l'organisation temporelle de processus. Cette vision d'une modularité nécessaire, qui serait justifiée par l'observation du degré de clustering, est fortement influencée par l'ingénierie où elle est commune. Elle a grandement inspiré le courant de la biologie synthétique. Cependant, malgré des mesures de clustering sur un certain nombre de réseaux biologiques, son association avec la modularité n'est pas une évidence. Il est en fait difficile de définir objectivement une notion de module biologique car l'étude détaillée d'exemples indique une pléiotropie importante de gènes mais aussi de circuits. Les sous réseaux séparés associés à des rôles fonctionnels propres semblent plutôt devoir être des exceptions [[MAZURIE et collab., 2005b](#); [PAYNE et WAGNER, 2015](#)].

Le réseau dans le temps : prédire l'activité Cependant la structure d'un réseau n'est pas forcément suffisante pour tout comprendre de la réalité qu'il décrit. Les réseaux sont en effet des représentations statiques qui manifestent des interactions possibles. Ces réseaux soutiennent une activité des nœuds qui se propage en leur sein et aboutit à des comportements dynamiques de tout l'ensemble. La vie du réseau dépendra de sa structure mais également des propriétés de chaque lien, de ses règles de transfert d'information, d'activité ou d'influence. Pour les réseaux biochimiques ceci signifie la construction de modèles décrivant les états des biomolécules mises en relation. Pour des gènes et des protéines en relation dans un réseau il s'agira par exemple de représenter les états d'activité des uns et les abondances des autres. D'établir donc des représentations mathématiques de la manière dont les différentes biomolécules d'une cellule interagissent et se contrôlent. Dans la section ci-dessous, je discuterai des principales approches de modélisation qui interviennent à ce niveau et que j'ai été amené à manipuler dans mes applications présentées dans le prochain chapitre.

4.3 Modélisation mathématique biophysique de l'expression génétique

« Chemical reactions are the lingua franca of biological modeling. They provide a unifying notation by which to express arbitrarily complex chemical processes, either qualitatively or quantitatively. Specifying chemical reactions is so fundamental that the same set of chemical reactions can lead to different computational models, e.g., a detailed differential equations model or a detailed stochastic model. In this sense, representing processes by chemical equations is more basic than using either differential equations or stochastic processes to run calculations to make predictions. »

Michael Gibson and Eric Mjolsness [[GIBSON et MJOLSNESS](#)]

Les principes de la démarche de modélisation scientifique ont été discutés dans le chapitre introductif 3, section 3.2.4. Je discuterai ici plus spécifiquement de modélisation mathématique pour la dynamique d'expression génétique, l'espace moléculaire auquel je me suis le plus consacré dans mon activité de recherche en biologie des systèmes. Il s'agit d'avoir à disposition un outillage théorique pour prédire l'évolution dans le temps des activités de gènes interconnectés, donc les effets de circuits formés par des facteurs de transcription.

La régulation et l'expression génétique font intervenir de nombreux facteurs et une machinerie moléculaire et des processus séquentiels particulièrement sophistiqués. Chaque étape peut agir sur le contrôle du niveau d'expression d'un gène, c'est-à-dire de la quantité de produit finalement synthétisé par unité de temps. Le niveau de protéine présent dans le noyau ou le cytoplasme dépendra également de sa dégradation et utilisation dans les processus cellulaires. La régulation

génétiq ue peut donc jouer sur chacun de ces niveaux. La littérature de modélisation de réseaux gé- nétiq ue a des origines déjà anciennes, bien avant la biologie synthétique. Des études théoriques de circuits génetiq ues ont été entreprises dès les années soixante dans la suite de la découverte des mécanismes de régulation génetiq ue. Une activité de modélisation mathématique s’est pour- suivie très localement dans certains cercles restreints tout le long des décennies suivantes avant de venir davantage sur le devant de la scène dans les années 2000. Les premiers théoriciens qui ont traité des systèmes de régulation comme [GOODWIN](#) en 1965 l’année du prix Nobel de Jacques Monod et Wolf, ainsi que [GRIFFITH](#) en 1968 et [THOMAS](#) en 1973, ont développé des approches mathématiques encore courantes actuellement. Ces auteurs ont abordé des circuits génetiq ues simples, souvent formels avec une vision d’exploration conceptuelle.

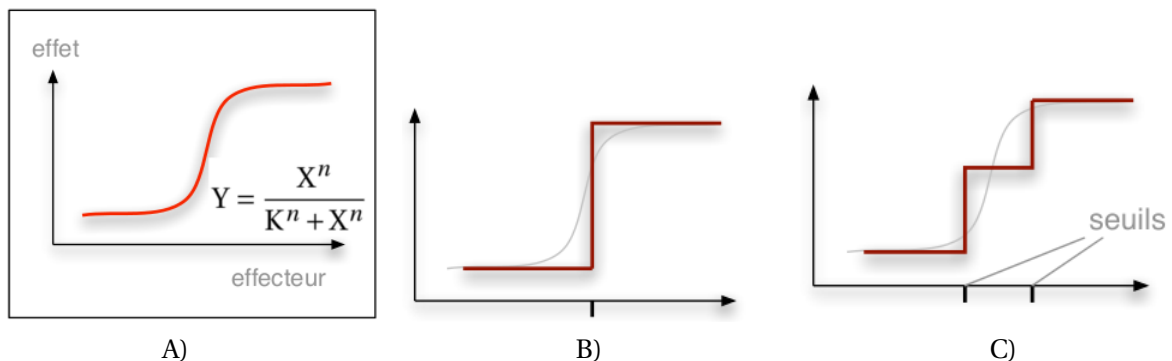


FIGURE 4.9 – **Différentes approches de représentation de la courbe de réponse de la régulation génétique.** A) fonction de réponse stéréotype de forme sigmoïde du niveau d’activité d’un promoteur génétique en fonction de la concentration d’un facteur de transcription (effecteur). Incrustation : fonction de Hill de la réponse Y en fonction de la concentration X du régulateur, n est le coefficient de Hill et K le niveau seuil à partir duquel l’activité est supérieure à la moitié de la maximale. B) Description logique Booléenne, la transcription est soit active ou éteinte. C) Approche logique étendue : une discrétisation à plusieurs niveaux de la fonction de réponse

De telles idées ont aussi été poursuivies par Stuart Kauffmann pour explorer l’organisation des organismes vivant comme les attracteurs d’états d’activité de gènes d’une dynamique génétique (il en fait une discussion extensive dans sa monumentale monographie « The Origin of Order » [[KAUFFMAN, 1992](#)]). Les différents attracteurs de réseaux de régulation génetiq ues peuvent ainsi s’interpréter comme des types cellulaires distincts dans les organismes. L’école de Bruxelles a également poursuivie ces réflexions dans la suite des idées de Prigogine [[PRIGOGINE et collab., 1969](#)]. Albert Goldbeter s’est ainsi particulièrement intéressé aux cycles périodiques d’activité biochimique et génetiq ue. Et René Thomas sur les états attracteurs possibles de circuits génetiq ues en fonction des conditions logiques des boucles présentes et des fonctions de régulation [[THOMAS et KAUFFMAN, 2001](#)]. Ces travaux durant les années 70 et 80 n’ont que partiellement été repris et cités durant la vague de la modélisation de la biologie des systèmes bien plus tardive. On peut toutefois se référer à plusieurs articles de revue de qualité [[BOLOURI et DAVIDSON, 2002](#); [ENDY et BRENT, 2001](#); [DE JONG, 2002](#); [SMOLEN et collab., 2000](#); [TYSON et collab., 2001](#)] publiés au tournant du millénaire qui ont jeté les jalons des approches possibles. On peut noter deux tendances principales pour la modélisation.

Modélisation empirique

Certains modélisateurs assument un point de départ d’emblée très formel de leur mise en équation d’un système génetiq ue, basé sur des hypothèses générales sur le fonctionnement de la régulation. C’est le choix élémentaire d’une fonction de régulation booléenne pour décider si tel gène est actif ou non. C’est l’écriture d’une fonction ad hoc, de Hill, pour le taux de transcription d’un gène en fonction de la concentration d’un facteur de transcription. Des choix inspirés par d’autres situations déjà étudiées et dont la justification viendrait a posteriori par le succès ou non du modèle à reproduire des comportements souhaités. La figure 4.9 montre une idéalisation de

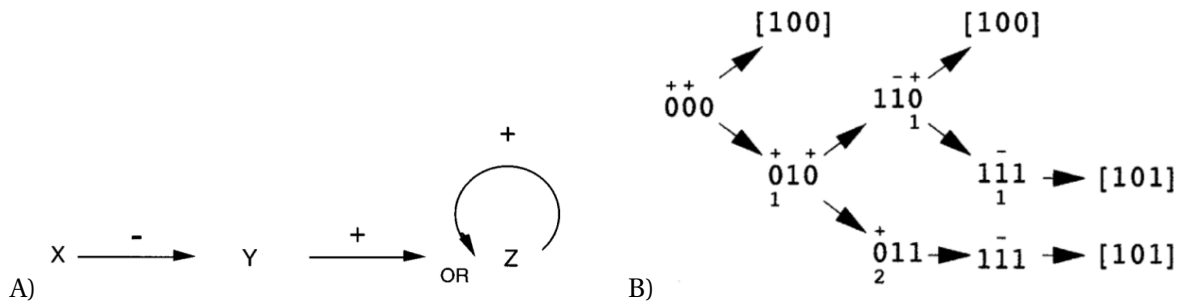


FIGURE 4.10 – **Exemple d'un modèle effectif logique : lyse et lysogénie de phage λ** Un exemple de la modélisation asynchrone logique de René Thomas inspiré de la réponse lyse ou lysogénie de bactéries infectées par un bacteriophage. Quand une population bactérienne est infectée certaines cellules vont donner lieu à une lyse, détruisant la cellule et libérant des copies du phage. D'autres effectueront une symbiose avec le virus qui s'incorpore dans leur génome. Toutes les cellules sont génétiquement identiques mais deux états sont possibles suite à l'invasion par des phages. Un modèle simplifié du système génétique de choix entre les deux voies fait intervenir le circuit de trois gènes X, Y et Z portés par le phage (correspondant aux gènes *cro*, *cII*, et *cI* dans le phage λ) est représenté en A). Le gène X est « constitutif », activé en permanence. Le gène Y est actif et produit sa protéine *y* si le produit *x* du gène X est absent. Le gène Z est actif si l'une des conditions est satisfaite : *y* est présent ou bien son produit *z* lui même. Dans la version la plus simple de l'approche de René Thomas, les états d'activités des gènes sont représentés par des variables booléennes, 0 inactif, 1 si le gène est actif. Activation et inactivation sont asynchrones, car des délais spécifiques à chaque gène sont nécessaires pour la transition, apparition ou disparition du produit du gène concerné. Immédiatement après l'infection l'état initial du système est 000 car aucune protéine du phage n'est encore produite. La séquence des transitions possibles est représentée en B). Les chemins suivis dépendent des délais t_G et $t_{\bar{G}}$ d'apparition et disparition des produits du gène G, avec $G \in \{X, Y, Z\}$. Les exposants « + » ou « - » indiquent les commandes d'activation ou inactivation en fonction de l'état des entrées sur un gène. Les états [100] et [101] sont les deux attracteurs possibles du circuit correspondant aux deux choix stables possibles. Le circuit complet fait intervenir des interactions supplémentaires mais peut être analysé de la même manière [THIEFFRY et THOMAS, 1995]. Source : d'après [THOMAS et KAUFMAN, 2001]

l'activation d'un gène sous forme d'une fonction ON-OFF et, ci celle-ci apparaît très caricaturale dans le système modélisé, sous la forme d'une fonction discrète. Donc c'est la construction de modèles de régulation génétiques effectifs en se basant sur un répertoire de fonctions et d'équations types déjà utilisées pour ce genre de modélisation. Le modèle est « efficace » dans le sens qu'il n'est pas déduit de la traduction mathématique de mécanismes moléculaires détaillés. Son succès à décrire un comportement est sa seule justification et non pas sa dérivation à partir des mécanismes. Un exemple simple d'une telle manière effective de construire des modèles est présenté dans la figure 4.10 qui illustre le principe d'un raisonnement logique discret asynchrone, suivant l'approche de René Thomas.

Construction de modèles de régulation à partir de principes microscopiques

Une tendance différente consiste à construire le modèle mathématique à partir de mécanismes biologiques. En tous cas d'une sélection d'un certain nombre de ceux-ci. Le point de départ d'une telle modélisation de réseau de régulation génétique est de penser le système étudié comme un ensemble de réactions chimiques en cascade dont les réactants sont les objets biologiques en interaction et les produits les résultats des processus. Il s'agit de représenter tous les éléments et les processus de la régulation et de l'expression génétique comme des réactions chimiques en solution.

Il conviendrait d'ailleurs de parler plutôt de « pseudo » -réactions chimiques étant donné que les acteurs ne sont pas nécessairement des biomolécules élémentaires, mais peuvent être toute structure cellulaire intervenant dans le processus biologique étudié. Par exemple on peut avoir choisi de représenter la synthèse d'une molécule de RNAm après l'initiation de la transcription

en une seule étape de production, faisant abstraction de la progression de la polymérase responsable de sa synthèse. L'intérêt d'une description des processus biologiques sous une forme de réactions chimiques ou de pseudo-réactions, sera de permettre d'une manière systématique l'écriture d'équations mathématiques traduisant le modèle. Le jeu consistera alors de faire usage de différentes formalisations existantes pour traiter mathématiquement des réactions chimiques.

L'utilisation d'un langage en termes de réactions chimiques, ou de pseudo-réactions, pour modéliser les processus des différents réseaux biomoléculaires dans les cellules qu'ils soient génétiques, protéiques, métaboliques et qu'ils impliquent des biomolécules simples ou des assemblages de sophistiquées, est en réalité un artifice pratique qui fournit un point de départ pour l'étude mathématique des systèmes représentés de cette manière. L'écriture du modèle sous forme de pseudo-réactions chimiques correspond à une description formelle de tous les composants inclus dans le modèle, de l'ensemble des interactions possibles et des résultats de chacune de ces interactions. Ces idées ont notamment été formalisées par Gibson et Mjølness [GIBSON et MJOLNESS] et un exemple est donné dans l'encadré (4.1).

1) *Description formelle par pseudo-réactions*

Si F désigne le facteur de transcription pris ici comme un activateur, DNA désigne le promoteur, F • DNA le complexe d'un facteur lié sur le promoteur, l'activité du gène peut se formaliser avec des pseudo-réactions suivantes.

- $F + \text{DNA} \leftrightarrow F \bullet \text{DNA}$ représente les événements d'association et dissociation du facteur de transcription sur le promoteur,
- $F \bullet \text{DNA} \rightarrow F \bullet \text{DNA} + \text{ARN}m$ représente la production d'ARN messenger tant que le promoteur est occupé par l'activateur.

2) *Mise en équation par la cinétique chimique* Chaque espèce du modèle est représentée par sa concentration, écrite avec son nom entre crochets. Les équations du modèle s'obtiennent par application directes des principes de la cinétique chimique

$$\frac{d[F]}{dt} = k_+ [F][\text{DNA}] - k_- [F \bullet \text{DNA}] \quad (4.1)$$

$$\frac{d[\text{DNA}]}{dt} = k_+ [F][\text{DNA}] - k_- [F \bullet \text{DNA}] \quad (4.2)$$

$$\frac{d[F \bullet \text{DNA}]}{dt} = -k_+ [F][\text{DNA}] + k_- [F \bullet \text{DNA}] \quad (4.3)$$

$$\frac{d[\text{ARN}m]}{dt} = k_p [F \bullet \text{DNA}] \quad (4.4)$$

Encadré 4.1 – Représentation schématique d'un régulation activatrice de l'activité d'un gène

Cette procédure de modélisation, formalisée en tant que système de « pseudo »-réactions chimiques a grandement inspiré des chercheurs en science informatique [DANOS et LANEVE, 2004] [REGEV et SHAPIRO, 2002]. Comment décrire ces interactions généralisées entre biomolécules et structures cellulaires intervenant dans les processus de régulation, transmission de signaux et métabolisme tout le long de la vie cellulaire? Quel formalisme permet de décrire précisément en toute généralité le réseau virtuel de toutes les interactions possibles? La notion de « pseudo »-réactions de ma description précédente est trop sommaire si on veut représenter de vrais systèmes : la nomenclature de tous les produits de réactions et assemblages possibles devenant rapidement impossible à suivre sans le support d'un formalisme suffisamment flexible. Interviennent alors des formalisations informatiques permettent de gérer le grand nombre de combinaisons et d'états dans un système biologique.

Ainsi durant les années 2000 plusieurs propositions de langages de modélisation informatique des systèmes biochimiques cellulaires ont été proposés [BANÂTRE et collab., 2005] pour toute sorte

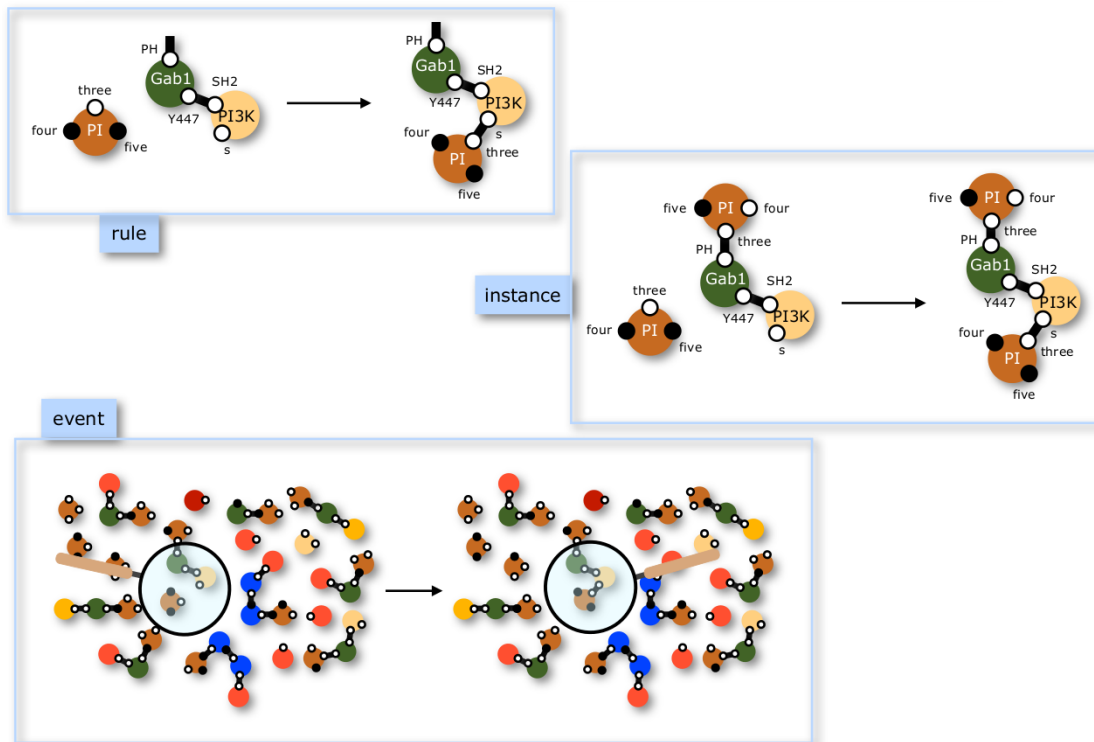


FIGURE 4.11 – **Illustration du langage Kappa.** Le langage Kappa représente chaque réactif chimique avec une désignation explicite des différents sites de réaction possible. Ici la molécule Pi possède les sites nommés *three*, *four* et *five*, qui s'écrivent selon $Pi(three, four, five)$. Un formalisme littéral, ou graphique sur une interface disponible du langage, permet de spécifier les règles de réaction possible, ici le site *three* de Pi peut se lier au site *s* de la molécule $PI3K(s, SH2)$, laquelle possède un autre site actif *SH2* (encadré règle). Cette réaction possible peut s'écrire avec un label sur les deux sites concernés $Pi(three^1, four, five), PI3K(s^1, SH2)$, ou bien dessinée graphiquement sur l'interface. Un simulateur stochastique permet d'évaluer la probabilité de réalisation de chaque réaction possible (encadré instance) et simuler l'évolution du système avec un grand nombre de molécules (encadré event). Source : Jean Krivine cours au master AIV, introduction au langage Kappa

d' applications dans des modèles de voies de signalisation cellulaires ou métaboliques et d'autres circuits de régulation biochimiques et génétiques. Ces langages se sont souvent inspirés ou ont repris des formalismes de l'informatique développés dans le contexte de recherches sur le calcul parallèle. L'implémentation chimique d'un réseau virtuel d'interactions est en effet un système où se déroulent un grand nombre de processus en parallèle de manière asynchrone. Ces formalismes portent souvent des noms exotiques, le π -calcul stochastique [PRIAMI, 1995] et continu [KWIATKOWSKI et STARK, 2008], les réseaux de Petri [HEINER et collab.] et les réseaux de Petri colorés [JENSEN, 1992], le système Kappa [CHIAVERINI et DANOS, 2003], Brane-calculus [CARDELLI, 2005], « Chemical Abstract Machine » (CHAM) et Biocham [CHABRIER-RIVIER et collab., 2005], Bio-NetGen, BlenX [DEMATTÉ et collab., 2008], PEPA [CALDER et collab., 2006; CIOCCHETTA et HILLSTON, 2008], le Systems Biology Graphical Notation SBGN [MOODIE et collab., 2015; NOVÈRE et collab., 2009] <http://sbgn.github.io/sbgn/>. Un système comme le langage Kappa [DANOS et LANEVE, 2004] permet de décrire des systèmes réactifs complexes et maîtrisant l'explosion combinatoire quand les molécules peuvent être impliquées dans des réactions de multiples manières, voir figure 4.11. Pour une revue récente et des illustrations des extensions d'actualité on pourra se reporter au « Language for Biological Systems » (LBS) [PEDERSEN et collab., 2015; PEDERSEN et PLOTKIN, 2010]. On peut considérer ces différents formalismes comme des généralisations de systèmes d'agents pouvant interagir selon des ensembles de règles conditionnelles. Dans le cadre de ce foisonnement de langages et du besoin d'outils de modélisation le « System Biology Markup Language » SBML est un développement complémentaire important visant à fournir un format

pour coder tout type de modèle de manière à être lisible et traitable automatiquement par un ordinateur [CHAOUIYA et collab., 2013; HUCKA et collab., 2003]. L'objectif est d'avoir une version spécialisée pour des modèles d'interactions biochimiques cellulaires du format général XML pour des bases de données.

L'application de ces langages s'appuie souvent sur des programmes permettant l'écriture des modèles de manière commode à l'aide d'interfaces graphiques. Inspirés des théories et des méthodes sur les langages de programmation informatiques, ces systèmes incluent généralement aussi des méthodes d'analyse de la cohérence logique intéressant permettant par exemple de détecter dans des modèles de grande taille l'absence de règles contradictoires et conduisant à des blocages, ainsi que la complétude des résultats simulés (toutes les solutions possibles ont-elles été trouvées.) Le site web du format SBML recense un grand nombre de ces outils avec lesquels il est généralement compatible [BERGMANN et collab.].

Ces langages de modélisation formels servent non seulement à la description des processus biologiques mais servent aussi aux analyses. Une fois le système étudié formalisé, il devient en effet possible de mettre en œuvre des méthodes différentes d'analyse mathématique la plupart du temps numériques. De nombreuses revues auxquelles je renvoie ont traité de ces approches [BOLOURI et DAVIDSON, 2002; ENDY et BRENT, 2001; DE JONG, 2002; SMOLEN et collab., 2000; TYSON et collab., 2001] ainsi que ma revue de 2007 [BOTTANI et MAZURIE, 2007]. Les approches numériques reprennent l'étendue des méthodes classiques des champs d'étude de systèmes dynamiques déterministes ou stochastiques et choisies pour un modèle donné en fonction de l'objectif de l'étude et de sa complexité. De nombreux travaux ont été consacrés aux algorithmes de résolution pour améliorer la vitesse de calcul et les capacités d'analyse. Ceci passe en particulier par une réflexion sur la simplification des modèles.

4.4 Approches mathématiques et réductions de modèles

Les différentes méthodes de modélisation mathématique et numériques qui ont pu être développées pour des systèmes dynamiques se retrouvent naturellement pour l'étude des dynamiques sur les circuits génétiques. L'activité de ces réseaux correspondant au déroulement de réactions chimiques, leur modélisation reprend donc naturellement aussi les approches connues pour le traitement de réactions chimiques.

À une échelle microscopique ces réactions sont des processus stochastiques, conduits par la diffusion des réactifs dans le milieu, l'occurrence de rencontres et des probabilités de réaction lorsque celles-ci ont lieu. Je discuterai plus en détail de la modélisation stochastique des système plus bas, section 4.4.2. On peut déjà anticiper le résultat classique selon lequel dans la limite des grands nombres de molécules, ces processus peuvent être bien approchés par des descriptions déterministes sous la forme d'équations différentielles de taux de réactions classiques de la cinétique chimique. De la même manière pour les modélisation de réseaux biochimiques, on pourra faire la part entre approches stochastiques et déterministes, généralement sous forme de systèmes d'équations différentielles (déterministes ou stochastiques) mais également dans des implémentations de nature informatique de type automates cellulaires ou d'approches Monte-Carlo diverses.

Dans le champ de la biologie des systèmes, les modélisations que je considérerai ici, comme la plupart de celles dans la littérature, commencent par négliger l'espace. À l'échelle la plus fine il faudrait tenir compte des positions et vitesses de toutes les molécules et suivre leurs interactions selon les lois de la physique. Il s'agirait de modèles très détaillés de dynamique moléculaire qui ne sont possibles que pour des systèmes limités, protéine, canaux ioniques, et une évolution temporelle très courte, inférieures à la microseconde [VLACHAKIS et collab., 2014] La plupart des modèles de dynamiques de réseaux biochimiques vont se placer à une échelle considérablement plus grande et représenter grossièrement les zones de l'espace dans lesquelles ont lieu les réactions. Ce

seront les compartiments cellulaires qui enferment les molécules et dans lesquels on s'intéressera à l'évolution des abondances respectives durant l'évolution des réactions. Dans chaque zone on suppose que les molécules diffusent librement entre les réactions formant un milieu bien mélangé dans lequel toutes les molécules sont réparties uniformément. Une hypothèse centrale à la plupart des approches bien que le milieu intra-cellulaire est loin de ressembler à un réacteur chimique [BERRY, 2002; VAN et collab., 2014]. Enfin, de manière implicite on considère aussi que le système global et ses compartiments sont à l'équilibre thermodynamique.

4.4.1 Systèmes d'équations différentielles ordinaires

Les approches basées sur des systèmes d'équations différentielles sont certainement les plus communes et de nombreux environnements de simulation de réseaux génétiques proposent des interfaces visant à faciliter leur mise en œuvre [ARKIN; FAGES et collab., 2004; MENDES et KUMMER; SAURO, 2000]. Les modèles s'écrivent en effet mathématiquement assez facilement, dérivant de la cinétique chimique classique. La forme des équations pour représenter les processus dans les circuits génétiques et biochimiques est issue la plupart du temps d'une démarche phénoménologique. Des choix empiriques d'expressions inspirées par une abstraction des processus de régulation et dont la justification viendrait seulement a posteriori par le comportement du modèle. Il en est ainsi de la vision simplifiée de la régulation génétique qui s'est diffusé à partir des années 60 où les gènes sont sous le contrôle de protéines agissant à l'instar d'interrupteurs permettant d'activer ou inhiber la transcription.

Cette notion d'interrupteur génétique conduit à la représentation de l'activation génétique sous forme de fonctions à seuils, plus ou moins fortement non linéaires, dépendantes de la concentrations des facteurs de transcription. Un raisonnement phénoménologique qui amène dès les premières modélisations dans les années 60 de tels systèmes [GOODWIN, 1965; GRIFFITH, 1968a,b] à l'usage de fonctions de type sigmoïde pour les taux d'expression de gènes en réponse à des concentrations différentes de régulateurs. Ou encore à des descriptions via des fonctions booléennes, comme dans l'approche logique développée par René Thomas [THOMAS, 1973, 1991] depuis le début des années 70 et ses généralisations ultérieures [THIEFFRY et SANCHEZ, 2003; THIEFFRY et THOMAS, 1998] dans des modèles effectifs prédisant des états d'expression discrets (figure 4.10). L'écriture empirique des modèles ne s'encombre pas de la dérivation des équations. Argumentant que ces modèles sont de toute façon le fruit de simplifications radicales de la complexité biologique, elle vise à écrire des systèmes d'équations efficaces. La justification de ces écritures vient de leur succès à reproduire des données existantes, permettant d'interpréter les expériences qui les ont produites dans le cadre d'une explication mécanique, ainsi qu' à permettre certaines prédictions.

Cependant, au-delà des usages et succès de la pratique empirique, il est satisfaisant de pouvoir relier les formes mathématiques qui y sont employées –équations différentielles cinétiques– et les termes correspondant aux formes d'interactions chimiques intervenant–fonctions sigmoïde, fonctions de Hill– à des modèles microscopiques des processus moléculaires. À un niveau fondamental les systèmes de réactions chimiques biomoléculaires sont gouvernés par la thermodynamique, soumis aux fluctuations chimiques et à des processus de diffusion pouvant amener à la rencontre, par hasard, des réactifs dans les bonnes conditions pour que la réaction ait lieu. Les réactions chimiques sont aléatoires : ce sont des processus stochastiques.

4.4.2 Modélisations stochastiques

Les réseaux génétiques stochastiques correspondent à des systèmes de réactions chimiques décrites en termes probabilistes où les transitions sont évaluées en termes de probabilité. Les successions d'approximations à des échelles différentes de systèmes de réactions chimiques a notamment été étudiée et discutée par Dan Gillespie avec ses travaux sur leur description stochastique [GILLESPIE, 2000, 2002]. En résumé, on sait que l'évolution probabiliste d'un système de réactions chimiques peut être suivi grâce à une équation différentielle ordinaire simple exprimant

les probabilités à chaque instant, pour chaque état possible du système de se transformer dans l'un des autres états. La description mathématique détaillée de tels systèmes s'exprime microscopiquement en terme d'une équation dite équation maîtresse exprimant l'évolution de la densité de probabilité instantanée des configurations de tous les réactifs [VAN KAMPEN, 2007]. Dans des systèmes bien mélangés, comme on suppose être le compartiment cellulaire modélisé, l'équation maîtresse chimique a une base microscopique solide et décrit précisément les effets aléatoires au niveau moléculaire.

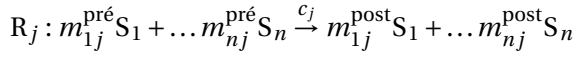
Malgré son écriture relativement simple, la dimension généralement énorme de cette équation, autant que le nombre d'états possibles, la rend insoluble directement en pratique. L'algorithme de GILLESPIE datant de 1977, qui l'a rendu célèbre, est un moyen numérique pour résoudre l'équation maîtresse par des simulations probabilistes (dites Monte-Carlo) d'histoires d'évolution possibles, voir l'encadré 4.2. Grâce à cette méthode on peut alors calculer l'activité d'un réseau biochimique et suivre l'évolution au cours du temps de la distribution des abondances de toutes les molécules impliquées. Il n'est même pas nécessaire de formuler explicitement l'expression de l'équation maîtresse pour le modèle étudié. Il suffit de la description des molécules et des réactions chimiques du modèle et des valeurs de leurs paramètres cinétiques, que l'on fournit à un programme dédié exécutant l'algorithme.

Facile à programmer, de nombreux programmes optimisés sont disponibles. Ceux-ci implémentent aussi plusieurs variantes de l'algorithme introduites tout le long des années 2000 le rendant plus rapide d'exécution sur des modèles avec des nombreuses réactions, il s'agit d'approximations [GIBSON et BRUCK, 2000] pertinentes quand de nombreuses réactions peuvent avoir lieu par unité de temps qui affectent faiblement les variables macroscopiques ainsi que des améliorations des implémentations informatiques [RAMASWAMY et SBALZARINI, 2010, 2011] (et pour des références récentes [THANH et PRIAMI, 2015]). Malgré ces outils et développements, les modèles sous la forme de l'équation maîtresse sont très coûteux en ressources numériques, sont laborieux d'utilisation et ne permettent aucune analytique. D'autres types de modélisations sont souhaitables quitte à s'éloigner de la précision microscopique.

Quand les nombres de molécules sont suffisamment grands les différences entre états discrets de l'équation maîtresse ne sont plus discernables et les trajectoires deviennent plus régulières. L'évolution du système devient continue et déterministe et peut se décrire par des équations différentielles ordinaires de la cinétique chimique classique. L'expérience montre que ces équations exprimant les taux de changement des abondances, moyennes, de chaque espèce présentes peuvent bien décrire phénoménologiquement la plupart des systèmes chimiques macroscopiques. Le passage d'une description microscopique détaillée en terme d'équation maîtresse vers des modèles effectifs continus dans la limite « thermodynamique » de populations moléculaires en pratique infinie est un sujet classique de la théorie des systèmes dynamique [KAMPEN, 2007], les principales approches sont résumées dans l'encadré 4.3. Cette transition prend la forme du passage à des descriptions continues sous forme d'équation Fokker-Planck et de Langevin, diminuant considérablement les dimensions du système et permettant des modèles plus manipulables. La première forme d'équations décrivant l'évolution temporelle de distributions de probabilités des abondances macroscopiques des molécules, la seconde permettant de suivre la trajectoire temporelle de ces abondances soumises aux fluctuations stochastiques du système.

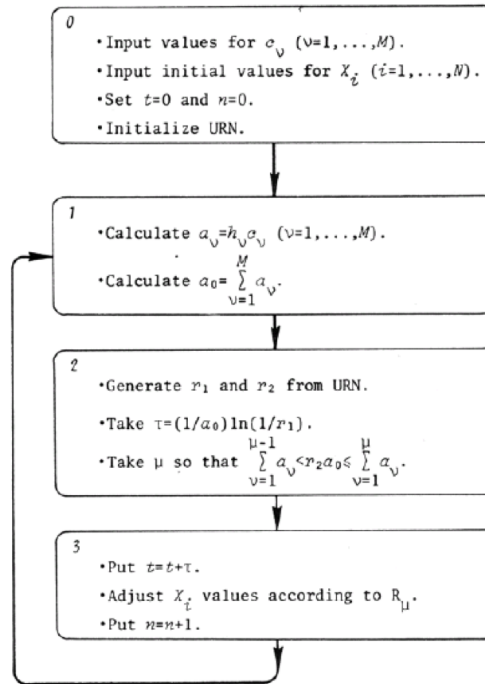
Dan Gillespie a étudié précisément dans le contexte particulier des réactions chimiques, les conditions sur les réactions chimiques modélisées sous lesquelles l'équation maîtresse chimique discrète pouvait être correctement approximées « *can be decently approximated* » par des équations continues de type Langevin ou bien de type Fokker-Planck [GILLESPIE, 2000, 2002]. Pour cela il faut à la fois que toutes les réactions aient des probabilités de réalisation qui restent stables sur des fractions très courtes de temps, et qu'elles soient également suffisamment grandes pour que chaque réaction ait lieu un grand nombre de fois durant le bref intervalle temporel. À cette double condition la dynamique stochastique détaillée des réactions chimiques décrite par l'équation maîtresse peut également être très bien approchée par les deux types de modélisations sto-

Le système chimique est composé de n molécules S_i présentes en x_i copies et de réactions $R_j, j = 1, \dots, m$:



les $m_{kl}^{\text{pré,post}}$ sont les coefficients stœchiométriques, c_j le taux élémentaire de la réaction.

- L'ensemble de micro-états $\bar{x}(t) = (x_1(t), \dots, x_n(t))$ caractérise toutes les configurations microscopiques possibles du système. $\bar{x}(t)$ change de manière stochastique en raison du caractère aléatoire de collisions moléculaires dues à l'agitation thermique.
- Chaque réaction chimique R_j change l'état moléculaire du vecteur $\bar{a}_j = (-m_{ij}^{\text{pré}} + m_{ij}^{\text{post}}), i \in [1 \dots n]$
- Pour chaque réaction R_j , le coefficient c_j décrit la probabilité par unité de temps d'avoir lieu par combinaison existante de ses réactifs S_i ($i \in [1 \dots n]$ tel que $m_{ij} \neq 0$).
- À un instant donné t et en fonction des nombres de molécules $x_i(t)$, chaque réaction possède un nombre de combinaisons distinctes de ses réactifs moléculaires $h_j(t) = h_j(\bar{x}(t))$ (par ex. pour $S_1 + S_2 \xrightarrow{c} S_3$, $h(t) = x_1(t)x_2(t)$).
- La probabilité par unité de temps pour chaque réaction R_j est alors le coefficient dit de propension (propensity) $a_j(t) = h_j(\bar{x}(t))c_j$.
- À partir de ces probabilités on tire au hasard la prochaine réaction R qui aura lieu.
- Le temps de la réaction est déterminée à partir d'une distribution de Poisson de coefficient la somme de toutes le propensions.
- Les quantités des réactifs sont mises à jour en fonction de la réaction r sélectionnée : $x_r \leftarrow x_r - m_{ir}^{\text{pré}} + m_{ir}^{\text{post}}$.



Encadré 4.2 – **L’algorithme de Gillespie.** À droite l’algorithme de Gillespie suivant la publication originale [GILLESPIE, 1977].

chastiques continues. Ces approches permettent alors des études numériques plus efficaces pour suivre l’évolution du système chimique avec ses fluctuations stochastiques. Comme le conclut Gillespie [GILLESPIE, 2002] l’intérêt de ces formalisations est en grande partie aussi conceptuel, car elles permettent de comprendre explicitement la transition vers les équations de taux de réaction

Équation maîtresse L'évolution probabiliste du système chimique introduit dans l'encadré 4.2 commence avec une configuration initiale donnée des abondances en termes de nombre de molécules de chaque type. Les probabilités $P(\bar{x}, t)$ au cours du temps de chaque configuration \bar{x} est décrite par une équation, dite l'**Équation maîtresse** qui exprime pour chaque état microscopique le bilan des probabilités des réactions possible qui amènent le système dans cet état moins le bilan des probabilités des réactions qui font sortir le système de cet état :

$$P(\bar{x}, t + dt) = \text{Probabilité d'être à l'état } \bar{x} \\ - \text{Probabilité de partir de l'état } \bar{x} \\ + \text{Probabilité d'être à tout autre état et d'arriver à l'état } \bar{x}$$

$$P(\bar{x}, t + dt) = P(\bar{x}, t) \left(1 - \sum_{j=0}^m a_j \cdot dt \right) + \sum_{j=0}^m a_j(\bar{x}(t)) P(\bar{x} + d_j, t)$$

L'équation maîtresse décrit exactement l'évolution de probabilités de chaque micro état *discret* du système stochastique.

Équation de Fokker-Planck chimique L'équation de Fokker-Planck est une limite mésoscopique de l'équation précédente qui décrit la variation dans le temps de la densité de probabilités des états. Elle peut être construite quand les variations entre chaque micro-état sont petites par rapport à la plage de variation des abondances, c'est-à-dire la granularité des états microscopiques négligeables. Les variables x_i de description du système sont alors des variables considérées continue, les *concentrations* de chaque espèce chimique. Avec des hypothèses sur les échelles de temps microscopiques du système étudiées attentivement dans [GILLESPIE, 2002] une équation différentielle partielle peut être construite (d'écriture trop complexe pour être utilement reproduite ici) qui décrit la fonction de distribution de la représentation continue des états du système.

Équation de Langevin chimique Quand la granularité des états microscopiques est petite on peut aussi approximer l'évolution stochastique du système des équations précédentes par les trajectoires stochastiques des quantités moyennes $x_i(t)$ de chaque espèce. L'équation, dite de Langevin, étend l'équation différentielle de l'évolution déterministe classique de la cinétique chimique à laquelle s'ajoutent des termes stochastiques supplémentaires représentant les fluctuations. L'établissement de ces termes de fluctuations est délicate et dépend des mêmes conditions sur les échelles de temps microscopiques que l'équation de Fokker-Planck et a été aussi étudiée en détail par Dan Gillespie [GILLESPIE, 2002].

Quand le système devient macroscopique et que la granularité des micro-états devient *négligeable* alors l'équation de Langevin se réduit à l'équation classique d'action de masse de la cinétique chimique. Les conditions justifiant un point de vue stochastique ou déterministes sont délicates à déterminer. Reprenant Dan Gillespie [GILLESPIE, 2002] :

« It has never been clear how the Chemical Master Equations gets supplanted, in the thermodynamic limit, by the conventional reaction rate equation. »

de la cinétique chimique dans la limite thermodynamique de système à grand nombre de molécules. En effet, dans cette limite les conditions pour les simplifications de l'équation maîtresse en équations de Langevin et de Fokker-Planck sont d'une part vérifiées, et d'autre part on peut montrer que la partie stochastique dans l'équation de Langevin devient négligeable par rapport à une partie déterministe qui correspond à l'équation de la cinétique chimique. De manière pratique il est cependant plus aisé, et sûr, de nos jours, grâce aux moyens de calcul numérique disponibles, de traiter les systèmes chimiques stochastiques à l'aide de la résolution Monte-Carlo exacte de l'équation maîtresse qu'avec les équations stochastiques approximés. Celles-ci exigent d'une part l'écriture parfois délicate des équations approximées et une vérification non immédiate de leurs conditions de validité [GILLESPIE, 2001].

4.4.3 Cinétique chimique pour les réseaux de régulation génétiques

De nombreuses applications de la littérature de modélisation de biologie des systèmes se fondent sur cette description des réseaux biochimiques cellulaires comme des systèmes chimiques mélangés qui permettent une mise en équation inspirée de la cinétique chimique avec des taux de réactions donnés par la loi d'action de masse, typiquement le produit des concentrations des réactifs. Le point de départ de l'écriture mathématique des modèles sera encore leur description comme ensemble de réactions chimiques comme discuté plus haut. Restons dans cette discussion sur la modélisation des réseaux de régulation génétique, plus spécifique à la biologie des systèmes, les mêmes principes se retrouvent dans les applications sur les réseaux de transmission de signaux et de voies métaboliques. En cohérence avec la compréhension des mécanismes moléculaires, la transcription d'un gène se fait tant que la région régulatrice de l'ADN en amont du gène, le promoteur du gène, est associé à une combinaison donnée de facteurs activateurs ou inhibiteurs qui y sont liés ou non. Dans le formalisme de la cinétique chimique le taux de transcription du gène dépendra de la quantité de promoteur positivement lié. Il s'exprimera proportionnellement à la *concentration* en promoteur lié, comme on peut le voir dans l'illustration de l'encadré ci-dessous.

Cette application de la cinétique chimique pour l'activité des réseaux de régulation génétique et biochimiques intracellulaires pose de nombreux problèmes d'interprétation car la limite à partir de laquelle l'approximation de modèles microscopiques stochastiques plus détaillés devient justifiée n'est pas toujours claire. On devrait la concevoir davantage comme un guide de construction de modèles phénoménologiques à valider à posteriori. Le fond de la question est dans le sens que l'on attribue au concept de concentration de promoteur lié. Pour un système chimique avec un grand nombre de réactifs les concentrations ont une signification claire. Dans le cadre de gènes et de facteurs de transcription dans le noyau cellulaire ou un cytoplasme bactérien la justification de cette limite ne semble pas a priori si valable. Néanmoins comme discutée notamment dans le cours de biologie des systèmes de Terence Hwa [HWA, 2017]— un des physiciens qui s'est le plus penché sur les fondements théoriques des modèles de régulation génétique— et dans une série d'articles méthodologiques [BINTU et collab., 2005a,b; BUCHLER et collab., 2003; GERLAND et collab., 2002], on peut y attribuer un sens qui expliquerait les succès des applications de cette nature. Dans une vision statistique on peut lire la concentration de promoteur occupé comme l'occupation moyenne d'un promoteur sur de nombreuses copies du système. En faisant appel au théorème ergodique dans nos systèmes considérés mélangés, ceci est égal à la fraction temporelle d'occupation d'un promoteur lorsqu'il est observé sur un temps long. Fraction du temps qui correspond à la période durant laquelle la transcription a lieu et contribuant positivement au taux de transcription. Ce raisonnement ergodique qui n'est pas toujours explicité est notamment nécessaire pour justifier les approches de modélisation de l'activité génétique basée sur la thermodynamique statistique introduites initialement par [ACKERS et collab., 1982] dès le début des années 80.

Ce cadre permet de calculer l'occupation moyenne d'un promoteur de structure complexe par des facteurs de transcription dont sera proportionnel le taux de transcription, suivant la lo-

gique discutée plus haut supposant un système bien mélangé et le théorème ergodique. Suivant [ACKERS et collab. \[1982\]](#) l'occupation des sites de liaison dans le promoteur, et leur conséquences successives sur la transcription, sont déterminés par les probabilités d'équilibre statistique thermodynamique dépendantes des énergies libres de liaison des facteurs de transcription pour former les complexes multi-protéiques et les associations avec leurs sites spécifiques d'appariement sur l'ADN spécifiques aux conditions de régulation particulières de chaque gène. Ce cadre repris et formalisé plus tard par le groupe de T. Hwa dans une série d'articles méthodologiques [\[BINTU et collab., 2005a,b; BUCHLER et collab., 2003; GERLAND et collab., 2002\]](#) permet de calculer l'occupation moyenne d'un promoteur de structure complexe par des combinaisons de facteurs de transcription dont sera proportionnel le taux de transcription. Il présuppose les hypothèses discutées plus haut de système dans une limite de grands nombres, bien mélangé et le raisonnement ergodique souvent supposés par défaut.

Approximation quasi-statique

L'idée de considérer l'occupation d'un promoteur comme un état d'équilibre est en réalité une autre manière de mettre en œuvre l'approximation discutée dans la section 3.2 du chapitre 3 d'un système détaillé avec des échelles de temps très différentes supposant les phénomènes les plus rapides quasi-statiques. Même un promoteur complexe comme celui du bactériophage λ peut être modélisé sous forme d'un ensemble de réactions chimiques comme discuté dans la section précédente 4.3 pour décrire toutes les associations et dissociations qui y interviennent. La cinétique chimique donnera un très grand nombre d'équations, avec de très nombreux paramètres cinétiques, inconnus la plupart.

1) Hypothèse quasi-statique

On suppose que les réactions d'association et de dissociation des facteurs de transcription sur leurs sites d'ADN conjugués dans le promoteur sont très rapides par rapport aux processus de transcription et qu'elles aient eu le temps d'atteindre un état stationnaire $\frac{d[F]}{dt} = \frac{d[DNA]}{dt} = \frac{d[F \cdot DNA]}{dt} = 0$. L'équation d'équilibre qui en résulte, $\frac{[F][DNA]}{[F \cdot DNA]} = K = \frac{k_-}{k_p}$ combinée à la relation de conservation $[DNA]_{tot}$

Conduit à l'expression simple pour la transcription où intervient la fonction hyperbolique.

2) *équilibre thermodynamique* Le facteur de transcription F possède une énergie libre de liaison sur le promoteur $\Delta G \dots$

Encadré 4.4 – Simplification des modèles : hypothèse quasi-statique et statistique thermodynamique pour l'occupation du promoteur. Illustration basée sur le modèle minimal de l'encadré 4.3.

Parmi les réactions présentes, celles de dissociation et association entre protéines et entre protéines et acide nucléique devraient avoir des taux individuels plus rapides que les processus sophistiqués et progressifs de transcription et traduction. Il est dès lors tentant et usuel de supposer légitime de séparer les dynamiques et de supposer qu'à l'échelle de variation des réactions les plus lentes (typiquement production de l'ARNm et des protéines), les réactions plus rapides (typiquement formation de facteurs de transcription multimériques, liaison du multimère sur le promoteur) ne varient pas. Formation et dissociation de multimères, association et dissociation de facteurs de transcription sur l'ADN peuvent suivant cette idée être considérés pratiquement stationnaires par rapport aux autres processus.

La fraction de promoteurs liés par un activateur se calculera donc à l'état stationnaire. Pour un promoteur simple avec une protéine régulatrice sur un site de liaison, ceci correspondra à la même approximation quasi-stationnaire pour le complexe Enzyme/Substrat du modèle Michaelis-Menten des réactions enzymatiques, le promoteur ayant le rôle de l'enzyme, et de la même manière le taux de transcription sera proportionnel à la fonction hyperbolique, dite aussi de Michaelis-Menten (ou à la fonction de Hill en cas de multimérisation du facteur), qui exprime la fraction d'association enzyme/substrat, ici la fraction d'association activateur/promoteur. Pour des promoteurs plus

complexes les expressions de régulation donnant l'occupation moyenne en fonction des abondances des différentes protéines facteurs de transcriptions seront les mêmes, qu'elles soient dérivées de simplifications quasi-statiques d'équations cinétiques ou bien de l'application de poids de Boltzmann pour calculer l'occupation moyenne.

L'usage du principe ergodique et d'approximations quasi-statiques conduit ainsi à des simplifications des modèles d'équations différentielles, et on justifie ainsi certaines formes de fonctions intervenant dans les modélisations directement empiriques ou bien on a disposition les outils pour construire des formes probables adaptées au système de régulation particulier d'un gène donné. Cette méthodologie est raisonnable avec un grand nombre de réactifs mais plus incertaine pour la régulation génétique étant donné le nombre limité d'éléments en interaction et notamment du nombre de copies de gènes.

Modèles à retard

Une ultérieure extension et possibilité de simplification de modèles est possible avec les méthodes tenant compte de retards. L'utilisation de systèmes d'équations différentielles ordinaires suppose comme simultanées les actions et réponses de tous les processus représentés dans les équations. Dans le cas de la modélisation de processus de régulation génétique plusieurs processus biologiques sont sophistiqués du point de vue moléculaire mais répétitifs comme la synthèse de l'ARN messenger par la polymérase durant sa progression le long de la séquence codante ou la polymérisation de la protéine par le ribosome le long de l'ADN. Dans un objectif de simplification on ne souhaite pas inclure ces étapes explicitement dans un modèle (sauf si elles étaient l'objet d'étude). On peut en faire abstraction, si les temps caractéristiques des processus individuels sont très inférieurs aux autres échelles temporelles auxquelles on observe le système. Une hypothèse qui n'est pas toujours explicitée. Le cas contraire, il est possible d'en tenir compte toute en simplifiant le modèle en condensant toutes les étapes intermédiaires dans des temps de retard explicites dans les modèles au bout desquels une action produira son effet. Des modèles à base d'équations différentielles à délai ont été utilisés dans de nombreuses situations biologiques et appliqués pour la première fois à des circuits génétiques également avant la vague de la biologie des systèmes (voir la discussion dans [?], chapitres 1 et 2).

L'inclusion de retards dans un système d'équations différentielles est une idée ancienne de simplification de la modélisations. Tout le déroulé de processus ou d'effets qui contribuent à construire une action sur les variables principales du système mais dont les détails ne sont pas d'intérêt principal sont capturés dans un terme d'action unique qui inclut un temps caractéristique représentatif de la durée entre leur initiation et la manifestation de l'action sur les grandeurs d'intérêt. Toutes les étapes intermédiaires sont ainsi condensées dans une interaction possédant un retard d'action sur les processus principaux modélisés. Des modèles à retard ont été utilisés avec succès dans une variété de contextes biologiques comme des désordres de la dynamique physiologique de la production de cellules sanguines (maladies hémato-poïétiques), des altérations du rythme respiratoire (respiration Cheyne-Stokes), ou encore des modèles écologiques les chapitre 1 et 2 du livre de J.D. Murray [MURRAY, 2002] traitent de ces exemples.

Le produit d'un gène impliquant une séquence de processus, la modulation de sa régulation aura un effet retardé sur le produit synthétisé et motive l'usage d'un délai dans les modèles entre l'initiation de la transcription et la production de la protéine. Les modèles à retard sont une manière efficace de rendre compte et capturer simplement les différents processus responsables des délais biologiques entre une action et son effet. Le principe d'un retard avait été introduit très tôt dès les premières publications de modèles mathématiques de circuits génétiques dans les années 1960 : dès les premières études de voies biochimiques [GRIFFITH, 1968a,b] avaient ainsi montré qu'une simple boucle rétroactive négative admettait des oscillations stables seulement en présence de retards, ou bien si la boucle négative contenait au moins trois variables fortement non linéaires. L'introduction de retards a pour effet d'augmenter la dimensionnalité des modèles et de

produire de nouveaux comportements dynamiques. Dans le contexte de la régulation génétique leur introduction semble remonter à AN DER HEIDEN 1979 qui a mis en évidence l'apparition d'oscillations dans le modèle précurseur de boucle de rétroaction génétique de GOODWIN 1965.

Bien que moins communs d'usage que des systèmes d'équations différentielles ordinaires, les systèmes à retard ont été repris dans plusieurs travaux de modélisation depuis le tournant du millénaire et la vague de la modélisation de circuits biochimiques. On les retrouve dans de modèles de génération de rythmes circadiens intracellulaires [SCHEPER et collab., 1999], d'expression oscillatoire de certaines protéines comme (p53, Hes1 et NF- κ B, Hes1) [BERNARD et collab., 2006; JENSEN et collab., 2003; MONK, 2003; TIANA et collab., 2002], du fonctionnement de opérons *lac* [YILDIRIM et MACKEY, 2003] et triptophane [MACKEY et collab., 2004], de l'interrupteur lyse/lysogénie [SANTILLÁN et MACKEY, 2004], la génération de patrons spatiaux périodique le long de l'axe dorsal du poisson zèbre [LEWIS, 2003]. Dans le chapitre suivant, à la section 5.3.2 on trouvera des contributions de ma part suivant cette approche.

4.5 Conclusion modélisation en biologie des systèmes

Ce chapitre a passé en revue plusieurs méthodologies de modélisation mathématiques pour la description de l'architecture de systèmes d'interaction biochimiques et des dynamiques qu'ils soutiennent. Ces approches ont déjà une longue histoire remontant au début même de la biologie moléculaire mais se sont essentiellement développées depuis une vingtaine d'années avec la vague de la biologie des systèmes.

Ces méthodes d'analyse de réseaux, et d'études des dynamiques d'activité génétique et moléculaires ont été à la base de mes travaux de recherche se rapportant à la biologie des systèmes que je vais maintenant passer en revue dans le chapitre suivant.

4.6 Références

- ACKERS, G. K., A. D. JOHNSON et M. A. SHEA. 1982, «Quantitative model for gene regulation by lambda phage repressor.», *Proc Natl Acad Sci U S A*, vol. 79, n° 4, p. 1129–1133. 124, 125
- ARKIN, A. «Biospice», <http://biospice.lbl.gov/>. 120
- BANÂTRE, J. P., P. FRADET et Y. RADENAC. 2005, «Principles of chemical programming», dans *Electronic Notes in Theoretical Computer Science*, ISSN 15710661, doi :10.1016/j.entcs.2004.07.019. 117
- BERGMANN, F. T., B. E. SHAPIRO et M. HUCKA. «Software working with SBML», URL http://sbml.org/Special/SBMLMatrix/static-summary.html#{#}cat_{_}3. 119
- BERNARD, S., B. CAJAVEC, P. L. MENJOUET, MACKEY et H. HERZEL. 2006, «Modelling transcriptional feedback loops : the role of Gro/TLE1 in Hes1 oscillations.», *Philos Transact A Math Phys Eng Sci*, vol. 364, n° 1842, doi :10.1098/rsta.2006.1761, p. 1155–1170. 127
- BERRY, H. 2002, «Monte Carlo Simulations of Enzyme Reactions in Two Dimensions : Fractal Kinetics and Spatial Segregation», *Biophysical Journal*, vol. 83, n° 4, doi :10.1016/S0006-3495(02)73953-2, p. 1891–1901, ISSN 00063495. 120
- BINTU, L., N. E. BUCHLER, H. G. GARCIA et collaborateurs. 2005a, «Transcriptional regulation by the numbers : applications.», *Curr Opin Genet Dev*, vol. 15, n° 2, doi :10.1016/j.gde.2005.02.006, p. 125–135. 124, 125

- BINTU, L., N. E. BUCHLER, H. G. GARCIA et collaborateurs. 2005b, «Transcriptional regulation by the numbers : models.», *Curr Opin Genet Dev*, vol. 15, n° 2, doi :10.1016/j.gde.2005.02.007, p. 116–124. [124](#), [125](#)
- BISARIA, A., P. HERSEN et M. N. MCCLEAN. 2014, «Microfluidic Platforms for Generating Dynamic Environmental Perturbations to Study the Responses of Single Yeast Cells», dans *Methods in molecular biology (Clifton, N.J.)*, vol. 1205, p. 111–129, doi :10.1007/978-1-4939-1363-3_8. [107](#)
- BOLOURI, H. et E. H. DAVIDSON. 2002, «Modeling transcriptional regulatory networks.», *Bioessays*, vol. 24, n° 12, doi :10.1002/bies.10189, p. 1118–1129. [115](#), [119](#)
- BOTTANI, S. et A. MAZURIE. 2007, «Genetic networks : between theory and experimentation», chap. Chapter 9, COMPLEX POPULATION DYNAMICS, World Scientific Lecture Notes in Complex Systems, p. 215–2326. [119](#)
- BUCHLER, N. E., U. GERLAND et T. HWA. 2003, «On schemes of combinatorial transcription logic.», *Proc Natl Acad Sci U S A*, vol. 100, n° 9, doi :10.1073/pnas.0930314100, p. 5136–5141. [124](#), [125](#)
- CALDER, M., S. GILMORE et J. HILLSTON. 2006, «Modelling the Influence of RKIP on the ERK Signalling Pathway Using the Stochastic Process Algebra PEPA», p. 1–23, doi :10.1007/11905455_1. [118](#)
- CARDELLI, L. 2005, «Brane Calculi», Springer, Berlin, Heidelberg, p. 257–278, doi :10.1007/978-3-540-25974-9_24. [118](#)
- CARSON-PARKER, S. et A.-L. RICHAIR. 2017, «La mécanobiologie, la nouvelle discipline qui bouleverse la recherche scientifique - The Monthly Digest», URL <http://www.monthly-digest-loreal.com>. [107](#)
- CHABRIER-RIVIER, N., F. FAGES et S. SOLIMAN. 2005, «The Biochemical Abstract Machine BIOCHAM», p. 172–191, doi :10.1007/978-3-540-25974-9_14. [118](#)
- CHAOUIYA, C., D. BÉRENGUIER, S. M. KEATING et collaborateurs. 2013, «SBML qualitative models : a model representation format and infrastructure to foster interactions between qualitative modelling formalisms and tools.», *BMC systems biology*, vol. 7, n° 1, doi :10.1186/1752-0509-7-135, p. 135, ISSN 1752-0509. [119](#)
- CHIAVERINI, M. et V. DANOS. 2003, «A Core Modeling Language for the Working Molecular Biologist (Abstract)», Springer, Berlin, Heidelberg, p. 166–166, doi :10.1007/3-540-36481-1_16. [118](#)
- CIOCCHETTA, F. et J. HILLSTON. 2008, «Bio-PEPA : An Extension of the Process Algebra PEPA for Biochemical Networks», *Electronic Notes in Theoretical Computer Science*, vol. 194, n° 3, doi : 10.1016/j.entcs.2007.12.008, p. 103–117, ISSN 15710661. [118](#)
- CLÉMENT, R., P. BLANC, B. MAUROY, V. SAPIN et S. DOUADY. 2012a, «Shape Self-Regulation in Early Lung Morphogenesis», *PLoS ONE*, vol. 7, n° 5, doi :10.1371/journal.pone.0036925, p. e36 925, ISSN 1932-6203. [104](#), [105](#)
- CLÉMENT, R., S. DOUADY et B. MAUROY. 2012b, «Branching geometry induced by lung self-regulated growth», *Physical Biology*, vol. 9, n° 6, doi :10.1088/1478-3975/9/6/066006, p. 066 006, ISSN 1478-3967. [104](#)
- COLLINS, R. 2010, «Electronic version of the Biochemical Pathways Chart : Collins Lab Blog Life in the Collins lab», URL <http://blogs.cornell.edu/collinslab/2010/04/08/electronic-version-of-the-biochemical-pathways-chart/>. [109](#)
- CORSON, F., O. HAMANT, S. BOHN et collaborateurs. 2009, «Turning a plant tissue into a living cell froth through isotropic growth», *Proceedings of the National Academy of Sciences*, vol. 106, n° 21, doi :10.1073/pnas.0812493106, p. 8453–8458, ISSN 0027-8424. [104](#)

- DANOS, V. et C. LANEVE. 2004, «Formal molecular biology», dans *Theoretical Computer Science*, vol. 325, Elsevier, ISBN 0304-3975, ISSN 03043975, p. 69–110, doi :10.1016/j.tcs.2004.03.065. [117](#), [118](#)
- DEMATTE, L., C. PRIAMI et A. ROMANEL. 2008, «Modelling and simulation of biological processes in BlenX», *ACM SIGMETRICS Performance Evaluation Review*, vol. 35, n° 4, doi :10.1145/1364644.1364653, p. 32, ISSN 01635999. [118](#)
- DOUADY, S. et Y. COUDER. 1992, «Phyllotaxis as a physical self-organized growth process», *Physical Review Letters*, vol. 68, n° 13, doi :10.1103/PhysRevLett.68.2098, p. 2098–2101, ISSN 0031-9007. [104](#)
- DOUADY, S. et Y. COUDER. 1996, «Phyllotaxis as a Dynamical Self Organizing Process Part I : The Spiral Modes Resulting from Time-Periodic Iterations», *Journal of Theoretical Biology*, vol. 178, n° 3, doi :10.1006/JTBI.1996.0024, p. 255–273, ISSN 0022-5193. [104](#)
- ENDY, D. et R. BRENT. 2001, «Modelling cellular behaviour.», *Nature*, vol. 409, n° 6818, doi :10.1038/35053181, p. 391–395. [115](#), [119](#)
- FAGES, F., S. SOLIMAN et C. N. RIVIER. 2004, «Modelling and Querying Interaction Networks in the Biochemical Abstract Machine {BIOCHAM}», *Journal of Biological Physics and Chemistry*, vol. 4, p. 64–73. [120](#)
- FERNANDEZ-SANCHEZ, M.-E., T. BRUNET, J.-C. RÖPER et E. FARGE. 2015, «Mechanotransduction's Impact on Animal Development, Evolution, and Tumorigenesis», *Annual Review of Cell and Developmental Biology*, vol. 31, n° 1, doi :10.1146/annurev-cellbio-102314-112441, p. 373–397, ISSN 1081-0706. [107](#)
- FLEURY, V., N. R. CHEVALIER, F. FURFARO et J. L. DUBAND. 2015, «Buckling along boundaries of elastic contrast as a mechanism for early vertebrate morphogenesis», *European Physical Journal E*, vol. 38, n° 2, doi :10.1140/epje/i2015-15006-7, ISSN 1292895X. [105](#), [106](#)
- FLEURY, V., A. V. MURUKUTLA, N. R. CHEVALIER et collaborateurs. 2016, «Physics of amniote formation», *Physical Review E*, vol. 94, n° 2, doi :10.1103/PhysRevE.94.022426, p. 1–19, ISSN 24700053. [102](#), [105](#), [106](#)
- GERLAND, U., D. J. MOROZ et T. HWA. 2002, «Physical constraints and functional characteristics of transcription factor-DNA interaction.», *Proc Natl Acad Sci U S A*, vol. 99, n° 19, doi :10.1073/pnas.192693599, p. 12 015–12 020. [124](#), [125](#)
- GIBSON, M. et E. MJOLSNESS. «Modeling the Activity of Single Genes», dans *From Computational Modeling of Genetic and Biochemical Networks*, édité par J. M. Bower et H. Boluri, chap. 1, MIT Press, Cambridge, p. 4–48. [114](#), [117](#)
- GIBSON, M. A. et J. BRUCK. 2000, «Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels», *The Journal of Physical Chemistry A*, vol. 104, n° 9, doi :10.1021/jp993732q, p. 1876–1889. [121](#)
- GILLESPIE, D. T. 1977, «Exact Stochastic simulation of coupled chemical reactions», *The Journal of Physical Chemistry*, vol. 81, p. 2340–2361. [121](#), [122](#)
- GILLESPIE, D. T. 2000, «Chemical Langevin equation», *Journal of Chemical Physics*, vol. 113, n° 1, doi :10.1063/1.481811, p. 297–306, ISSN 00219606. [120](#), [121](#)
- GILLESPIE, D. T. 2001, «Approximate accelerated stochastic simulation of chemically reacting systems», *The Journal of Chemical Physics*, vol. 115, n° 4, doi :10.1063/1.1378322, p. 1716–1733. [124](#)

- GILLESPIE, D. T. 2002, «The chemical Langevin and Fokker-Planck equations for the reversible isomerization reaction», *Journal of Physical Chemistry A*, vol. 106, n° 20, doi :10.1021/jp0128832, p. 5063–5071, ISSN 10895639. [120](#), [121](#), [122](#), [123](#)
- GOODWIN, B. C. 1965, «Oscillatory behavior in enzymatic control processes.», *Adv Enzyme Regul*, vol. 3, p. 425–438. [115](#), [120](#), [127](#)
- GRIFFITH, J. S. 1968a, «Mathematics of cellular control processes. {I}. {N}egative feedback to one gene.», *J Theor Biol*, vol. 20, n° 2, p. 202–208. [115](#), [120](#), [126](#)
- GRIFFITH, J. S. 1968b, «Mathematics of cellular control processes. {II}. {P}ositive feedback to one gene.», *J Theor Biol*, vol. 20, n° 2, p. 209–216. [120](#), [126](#)
- GUELZIM, N., S. BOTTANI, P. BOURGINE et K. FRANÇOIS. 2002, «Topological and causal structure of the yeast transcriptional regulatory network.», *Nat Genet*, vol. 31, n° 1, doi :10.1038/ng873, p. 60–63. [110](#), [113](#)
- HALL-SWAN, S., J. CRAWFORD, R. NEWMAN et L. J. COWEN. 2018, «Detangling PPI networks to uncover functionally meaningful clusters», *BMC Systems Biology*, vol. 12, n° Suppl 3, doi :10.1186/s12918-018-0550-5, p. 24, ISSN 17520509. [113](#)
- HARTWELL, L. H., J. J. HOPFIELD, S. LEIBLER et A. W. MURRAY. 1999, «From molecular to modular cell biology.», *Nature*, vol. 402, n° 6761 Suppl, doi :10.1038/35011540, p. C47—52. [114](#)
- AN DER HEIDEN, U. 1979, «Delays in physiological systems.», *J Math Biol*, vol. 8, n° 4, p. 345–364. [127](#)
- HEINER, M., D. GILBERT et R. DONALDSON. «Petri Nets for Systems and Synthetic Biology», dans *Formal Methods for Computational Systems Biology*, Springer Berlin Heidelberg, Berlin, Heidelberg, p. 215–264, doi :10.1007/978-3-540-68894-5_7. [118](#)
- HUCKA, M., A. FINNEY, H. M. SAURO et collaborateurs. 2003, «The systems biology markup language (SBML) : a medium for representation and exchange of biochemical network models», *BIOINFORMATICS I. I. Goryanin*, vol. 19, n° 9, doi :10.1093/bioinformatics/btg015, p. 524–531. [119](#)
- HWA, T. 2017, «Course on Quantitative Molecular Biology», URL <https://matisse.ucsd.edu/courses/winter-2015-physics-176276-quantitative-molecular-biology/>. [124](#)
- JENSEN, K. 1992, *Coloured Petri Nets*, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-662-06291-3, doi :10.1007/978-3-662-06289-0. [118](#)
- JENSEN, M. H., K. SNEPPEN et G. TIANA. 2003, «Sustained oscillations and time delays in gene expression of protein {H}es1.», *FEBS Lett*, vol. 541, n° 1-3, p. 176–177. [127](#)
- JEONG, H., S. P. MASON, A. L. BARABASI et Z. N. OLTVAI. 2001, «Lethality and centrality in protein networks.», *Nature*, vol. 411, n° 6833, doi :10.1038/35075138, p. 41–42. [110](#)
- JEONG, H., B. TOMBOR, R. ALBERT, Z. N. OLTVAI et A. L. SI. 2000, «The large-scale organization of metabolic networks.», *Nature*, vol. 407, n° 6804, doi :10.1038/35036627, p. 651–654. [110](#)
- DE JONG, H. 2002, «Modeling and simulation of genetic regulatory systems : a literature review.», *J Comput Biol*, vol. 9, n° 1, doi :10.1089/10665270252833208, p. 67–103. [115](#), [119](#)
- JUNIER, I., O. MARTIN et F. KÉPÈS. 2010, «Spatial and Topological Organization of DNA Chains Induced by Gene Co-localization», *PLoS Computational Biology*, vol. 6, n° 2, doi :10.1371/journal.pcbi.1000678, p. e1000678, ISSN 1553-7358. [113](#)

- KAMPEN, N. 2007, *Stochastic Processes in Physics and Chemistry*, North Holland Personal Library, ISBN 0444529659, 464 p.. [121](#)
- KANEHISA, M., S. GOTO, S. KAWASHIMA et A. NAKAYA. 2002, «The KEGG databases at Genome-Net.», *Nucleic acids research*, vol. 30, n° 1, p. 42–6, ISSN 1362-4962. [110](#)
- KAUFFMAN, S. A. 1992, *The origins of order : Self-organization and selection in evolution*, vol. 65, Oxford University Press, ISBN 0195058119, 153–181 p., doi :10.1126/science.260.5113.1531. [115](#)
- KAUFFMAN, S. A. 1993, *The Origins of Order : Self-Organization and Selection in Evolution*, Oxford University Press. [103](#)
- KWIATKOWSKI, M. et I. STARK. 2008, «The Continuous π -Calculus : A Process Algebra for Biochemical Modelling», p. 103–122, doi :10.1007/978-3-540-88562-7_11. [118](#)
- LADOUX, B. et R.-M. MÈGE. 2017, «Mechanobiology of collective cell behaviours», *Nature Reviews Molecular Cell Biology*, vol. 18, n° 12, doi :10.1038/nrm.2017.98, p. 743–757, ISSN 1471-0072. [107](#)
- LEWIS, J. 2003, «Autoinhibition with transcriptional delay : a simple mechanism for the zebrafish somitogenesis oscillator.», *Curr Biol*, vol. 13, n° 16, p. 1398–1408. [127](#)
- LOTEM, E. Y., S. SATTATH, N. KASHTAN et collaborateurs. 2004, «Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction.», *Proc Natl Acad Sci U S A*, vol. 101, n° 16, doi :10.1073/pnas.0306752101, p. 5934–5939. [110](#)
- MACKEY, M. C., M. SANTILLÁN et N. YILDIRIM. 2004, «Modeling operon dynamics : the tryptophan and lactose operons as paradigms.», *C R Biol*, vol. 327, n° 3, p. 211–224. [127](#)
- MALYGIN, A. G. 2006, «Morphodynamics of phyllotaxis», *The International Journal of Developmental Biology*, vol. 50, n° 2-3, doi :10.1387/ijdb.052052am, p. 277–287, ISSN 0214-6282. [104](#)
- MAZURIE, A., S. BOTTANI et M. VERGASSOLA. 2005a, «An evolutionary and functional assessment of regulatory network motifs.», *Genome biology*, vol. 6, n° 4, ISSN 14656914. [110](#)
- MAZURIE, A., S. BOTTANI et M. VERGASSOLA. 2005b, «An evolutionary and functional assessment of regulatory network motifs.», *Genome Biol*, vol. 6, n° 4, doi :10.1186/gb-2005-6-4-r35. [114](#)
- MENDES, P. et U. KUMMER. «Complex Pathway Simulator», <http://www.copasi.org/>. [120](#)
- MICHAL, G. 2018, «Roche - Biochemical Pathways», URL <https://www.roche.com/sustainability/philanthropy/science{ }education/pathways.htm>. [110](#)
- MONK, N. A. M. 2003, «Oscillatory expression of {H}es1, p53, and {NF}-kappa{B} driven by transcriptional time delays.», *Curr Biol*, vol. 13, n° 16, p. 1409–1413. [127](#)
- MOODIE, S., N. LE NOVÈRE, E. DEMIR, H. MI et A. VILLÉGER. 2015, «Systems Biology Graphical Notation : Process Description language Level 1 Version 1.3.», *Journal of integrative bioinformatics*, vol. 12, n° 2, doi :10.2390/biecoll-jib-2015-263, p. 263, ISSN 1613-4516. [118](#)
- MURRAY, J. D. 2002, *Mathematical Biology*, Springer. [126](#)
- NICODEMI, M. et A. PRISCO. 2009, «Thermodynamic pathways to genome spatial organization in the cell nucleus.», *Biophysical journal*, vol. 96, n° 6, doi :10.1016/j.bpj.2008.12.3919, p. 2168–77, ISSN 1542-0086. [113](#)
- NOVÈRE, N. L., M. HUCKA, H. MI et collaborateurs. 2009, «The Systems Biology Graphical Notation», *Nature Biotechnology*, vol. 27, n° 8, doi :10.1038/nbt.1558, p. 735–741, ISSN 1087-0156. [118](#)

- OULION, S., P. LAURENTI et D. CASANE. 2012, «Organisation des gènes HOX», *médecine/sciences*, vol. 28, n° 4, doi :10.1051/medsci/2012284005, p. 350–353. [112](#)
- PAYNE, J. L. et A. WAGNER. 2015, «Function does not follow form in gene regulatory circuits», *Scientific Reports*, vol. 5, n° 1, doi :10.1038/srep13015, p. 13 015, ISSN 2045-2322. [114](#)
- PEDERSEN, M., A. PHILLIPS et G. D. PLOTKIN. 2015, «A high-level language for rule-based modeling», *PloS one*, vol. 10, n° 6, doi :10.1371/journal.pone.0114296, p. e0114 296, ISSN 1932-6203. [118](#)
- PEDERSEN, M. et G. D. PLOTKIN. 2010, «A Language for Biochemical Systems : Design and Formal Specification», Springer, Berlin, Heidelberg, p. 77–145, doi :10.1007/978-3-642-11712-1_3. [118](#)
- PEREIRA-LEAL, J. B., A. J. ENRIGHT et C. A. OUZOUNIS. 2003, «Detection of functional modules from protein interaction networks», *Proteins : Structure, Function, and Bioinformatics*, vol. 54, n° 1, doi :10.1002/prot.10505, p. 49–57, ISSN 08873585. [113](#)
- PRIAMI, C. 1995, «Stochastic Calculus», *The Computer Journal*, vol. 38, n° 7, doi :10.1093/comjnl/38.7.578, p. 578–589, ISSN 0010-4620. [118](#)
- PRIGOGINE, I., R. LEFEVER, A. GOLDBETER et M. HERSCHKOWITZ-KAUFMAN. 1969, «Symmetry breaking instabilities in biological systems.», *Nature*, vol. 223, n° 5209, p. 913–6, ISSN 0028-0836. [115](#)
- RAJAPAKSE, I., M. D. PERLMAN, D. SCALZO et collaborateurs. 2009, «The emergence of lineage-specific chromosomal topologies from coordinate gene regulation.», *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, n° 16, doi :10.1073/pnas.0900986106, p. 6679–84, ISSN 1091-6490. [113](#)
- RAMASWAMY, R. et I. F. SBALZARINI. 2010, «A partial-propensity variant of the composition-rejection stochastic simulation algorithm for chemical reaction networks», *The Journal of Chemical Physics*, vol. 132, n° 4, doi :10.1063/1.3297948, p. 044 102, ISSN 0021-9606. [121](#)
- RAMASWAMY, R. et I. F. SBALZARINI. 2011, «A partial-propensity formulation of the stochastic simulation algorithm for chemical reaction networks with delays», *The Journal of Chemical Physics*, vol. 134, n° 1, doi :10.1063/1.3521496, p. 014 106, ISSN 0021-9606. [121](#)
- RAVASZ, E., A. L. SOMERA, D. A. MONGRU, Z. N. OLTVAI et A. L. BARABÁSI. 2002, «Hierarchical organization of modularity in metabolic networks.», *Science (New York, N.Y.)*, vol. 297, n° 5586, doi :10.1126/science.1073374, p. 1551–5, ISSN 1095-9203. [113](#)
- REGEV, A. et E. SHAPIRO. 2002, «Cells as computation.», *Nature*, vol. 419, n° 6905, doi :10.1038/419343a, p. 343, ISSN 0028-0836. [117](#)
- SANTILLÁN, M. et M. C. MACKEY. 2004, «Influence of catabolite repression and inducer exclusion on the bistable behavior of the lac operon.», *Biophys J*, vol. 86, n° 3, p. 1282–1292. [127](#)
- SAURO, H. M. 2000, «Jarnac : a system for interactive metabolic analysis.», dans *Proceedings of the 9th International Meeting on BioThermoKinetics.*, édité par J. H. Hofmeyr, J. M. Rohwer et J. L. Snoep, Stellenbosch University Press. [120](#)
- SCHEPER, T., D. KLINKENBERG, C. PENNARTZ et J. VAN PELT. 1999, «A mathematical model for the intracellular circadian rhythm generator.», *J Neurosci*, vol. 19, n° 1, p. 40–47. [127](#)
- SHAI, R. MILO, S. MANGAN et U. ALON. 2002, «Network motifs in the transcriptional regulation network of *Escherichia coli*.», *Nat Genet*, vol. 31, n° 1, doi :10.1038/ng881, p. 64–68. [110](#)

- SMOLEN, P., D. A. BAXTER et J. H. BYRNE. 2000, «Modeling transcriptional control in gene networks—methods, recent results, and future directions.», *Bull Math Biol*, vol. 62, n° 2, p. 247–292. [115](#), [119](#)
- SPIRIN, V. et L. A. MIRNY. 2003, «Protein complexes and functional modules in molecular networks», *Proceedings of the National Academy of Sciences*, vol. 100, n° 21, doi :10.1073/pnas.2032324100, p. 12 123–12 128, ISSN 0027-8424. [110](#)
- THANH, V. H. et C. PRIAMI. 2015, «Simulation of biochemical reactions with time-dependent rates by the rejection-based algorithm», *The Journal of Chemical Physics*, vol. 143, n° 5, doi :10.1063/1.4927916, p. 054 104, ISSN 0021-9606. [121](#)
- THIEFFRY, D. et L. SANCHEZ. 2003, «Dynamical modelling of pattern formation during embryonic development.», *Curr Opin Genet Dev*, vol. 13, n° 4, p. 326–330. [120](#)
- THIEFFRY, D. et R. THOMAS. 1995, «Dynamical behaviour of biological regulatory networks—II. Immunity control in bacteriophage lambda*», *Bulletin of Mathematical Biology*, vol. 57, n° 2, doi :10.1016/0092-8240(94)00037-D, p. 277–297, ISSN 00928240. [116](#)
- THIEFFRY, D. et R. THOMAS. 1998, «Qualitative analysis of gene networks.», *Pac Symp Biocomput*, p. 77–88. [120](#)
- THOMAS, R. 1973, «Boolean formalisation of genetic control circuits.», *J. Theor. Biol.*, vol. 42, n° 3, p. 565–583. [115](#), [120](#)
- THOMAS, R. 1991, «Regulatory networks seen as asynchronous automata : A logical description», *Journal of Theoretical Biology*, vol. 153, n° 1, doi :10.1016/S0022-5193(05)80350-9, p. 1–23, ISSN 10958541. [120](#)
- THOMAS, R. et M. KAUFMAN. 2001, «Multistationarity, the basis of cell differentiation and memory. {II}. {L}ogical analysis of regulatory networks in terms of feedback circuits.», *Chaos*, vol. 11, n° 1, doi :10.1063/1.1349893, p. 180–195. [115](#), [116](#)
- TIANA, G., M. H. JENSEN et K. SNEPPEN. 2002, «Time delay as a key to apoptosis induction in the p53 network», *Eur. Phys. J. B.*, vol. 29, p. 135–140. [127](#)
- TURING, A. M. 1952, «The Chemical Basis of Morphogenesis», *Philosophical Transactions of the Royal Society B : Biological Sciences*, vol. 237, n° 641, doi :10.1098/rstb.1952.0012, p. 37–72, ISSN 0962-8436. [105](#)
- TYSON, J. J., K. CHEN et B. NOVAK. 2001, «Network dynamics and cell physiology.», *Nat Rev Mol Cell Biol*, vol. 2, n° 12, doi :10.1038/35103078, p. 908–916. [115](#), [119](#)
- VAN, A. L., H. A. SOULA et H. BERRY. 2014, «Space-induced bifurcation in repression-based transcriptional circuits.», *BMC systems biology*, vol. 8, n° 1, doi :10.1186/s12918-014-0125-z, p. 125, ISSN 1752-0509. [120](#)
- VAN KAMPEN, N. G. 2007, *Stochastic Processes in Physics and Chemistry, Third Edition (North-Holland Personal Library)*, 3^e éd., North Holland, ISBN 0444529659. [121](#)
- VLACHAKIS, D., E. BENCUROVA, N. PAPANGELOPOULOS et S. KOSSIDA. 2014, «Current State-of-the-Art Molecular Dynamics Methods and Applications», dans *Advances in protein chemistry and structural biology*, vol. 94, p. 269–313, doi :10.1016/B978-0-12-800168-4.00007-X. [119](#)
- WAGNER, A. et D. A. FELL. 2001, «The small world inside large metabolic networks», *Proceedings of the Royal Society B : Biological Sciences*, vol. 268, n° 1478, doi :10.1098/rspb.2001.1711, p. 1803–1810, ISSN 0962-8452. [110](#)

WATTS, D. J. 1999, *Small worlds : the dynamics of networks between order and randomness*, Princeton University Press, ISBN 9780691117041, 262 p.. [110](#)

WIXON, J. et D. KELL. 2000, «The Kyoto Encyclopedia of Genes and Genomes—KEGG», *Yeast*, vol. 1, n° 1, doi :10.1002/(SICI)1097-0061(200004)17:1<48::AID-YEA2>3.0.CO;2-H, p. 48–55, ISSN 0749-503X. [110](#)

YILDIRIM, N. et M. C. MACKEY. 2003, «Feedback regulation in the lactose operon : a mathematical modeling study and comparison with experimental data.», *Biophys J*, vol. 84, n° 5, p. 2841–2851. [127](#)

Chapitre 5

Biologie des systèmes : contributions

« What makes the present stage of biological science so extraordinary is that molecular biology is driving us to the innermost reaches of the cell's ultimate mechanisms, complexity, and capacity to evolve. At the very same time, work in mathematics, physics, chemistry and biology is revealing how far-reaching the powers of self-organization can be. These advances hold implications for the origin of life itself and for the origins of order in the ontogeny of each organism. »

Stuart A. Kauffman, *The origin of Order*, pp. xiv, 1993.

Sommaire

5.1 Contributions : survol	136
5.2 Caractérisation des réseaux génétique	137
5.2.1 Reconstruction de circuits génétiques	137
5.2.2 Structure du réseau de régulation de la levure	141
5.2.3 Évaluation des motifs de réseau dans le génome de la levure	144
5.3 Circuits génétiques oscillants	149
5.3.1 Oscillations génétiques : motivation et contexte	150
5.3.2 Oscillations génétiques et modèles à retard	151
5.3.3 Modèle minimal des oscillations génétique du système p53 avec des équations différentielles retardées	152
5.3.4 Le système génétique p53	153
5.3.5 Comparaison de modélisation déterministes et stochastiques et de schémas de réduction	158
5.4 Caractérisation de la réponse non linéaire de la transcription.	165
5.4.1 Réponse transcriptionnelle avec compétition	169
5.5 Dosage génétique, aneuploïdie et déséquilibres génomiques	172
5.5.1 Dosage, haploinsuffisance et équilibres génétiques	174
5.5.2 Mécanismes de compensation de dosage	176
5.5.3 Un lien entre robustesse au dosage et AND « poubelle »?	182
5.5.4 Dosage et dominance génétique	187
5.5.5 Conclusion dosage génétique	194
5.6 Variabilité phénotypique et relation génotype/phénotype	194
5.6.1 Notre projet de caractérisation de la variabilité phénotypique et de la relation G→P	197
5.7 Autres travaux de biologie ds systèmes : Réponse au stress osmotique et contrôle temps réel de l'expression	204
5.8 Conclusion du chapitre	204
5.9 Références	205

5.1 Contributions : survol

Ce chapitre présente des synthèses de mes différentes contributions dans le cadre la biologie des systèmes suivant les approches discutées au chapitre précédent. Sur une période débutant vers l'année 2000, je me suis intéressé à différentes questions relevant de cette démarche dans des projets avec différents collaborateurs souvent interdisciplinaires.

L'étude des systèmes génétiques et des réseaux de leurs interactions génétiques comporte en règle générale différentes étapes. En premier il s'agit de caractériser le réseau lui-même : le reconstruire à partir des données disponibles de génétique et de génomique fonctionnelle, c'est-à-dire identifier toutes les relations. Une fois la carte des interactions établies, on peut analyser ses particularités de structure par l'application de différentes méthodes issues notamment de la théorie mathématique des graphes. On peut à ce stade, simplement par la structure, déjà peut-être identifier des éléments amenés à des rôles particuliers. Par exemple, reconnaître des gènes centraux impliqués dans un nombre particulièrement grand de processus. Avec une meilleure connaissance des éléments des systèmes et de leurs interactions on souhaiterait aussi pouvoir anticiper les comportements du système quand les conditions ou certains constituants changent. La connaissance de la carte des interactions et caractérisation de sa topologie ne suffit pas cependant à prédire la propagation des changements. Une analyse dynamique est nécessaire via la modélisation du fonctionnement de chaque lien et de la propagation de cette activité sur tout l'ensemble.

Dans le contexte de telles études je me suis intéressé à plusieurs questions en lien avec l'organisation fondamentale des réseaux et le lien avec leurs propriétés fonctionnelles qui ont porté sur plusieurs sujets :

Sur des questions de structure de réseaux :

- avec Florence d'Alché-Buc (Institut Mines Télécom) pour la reconstitution par apprentissage statistique à partir de données transcriptomique entre 2001 et 2003
- avec François Képès (Université d'Évry) et Paul Bourguin (École Polytechnique) pour la première analyse réalisée d'un réseau génétique global eukaryote entre 2000 et 2005.
- avec Massimo Vergassola (Institut Pasteur) pour l'analyse d'hypothèses sur la modularité des réseaux génétiques et protéiques de la levure en 2006.

Sur des questions de dynamiques portées par des circuits génétiques et la réponse à des signaux :

- avec Basile Grammaticos (CNRS laboratoire IMNC) de 2007 à 2008 des études de systèmes dynamiques de petites boucles de rétroaction génétiques caractéristiques de différentes situations biologiques, en particulier dans le cadre de la régulation de l'oncogène P53.
- avec Pascal Hersen (Laboratoire MSC) à propos de la réponse transductionnelle à des signaux de stress hyper-osmotique et le contrôle temps réel de l'expression génétique.
- avec Suzannah Rutherford (Hutchinson Center Seattle) dans l'étude de la variabilité de la relation génotype/phénotype dans l'exemple de la réponse au stress hyper-osmotique, dans le cas aussi de la levure.

Sur des questions d'équilibres globaux de niveaux d'expression génétiques :

- avec Reiner Veitia (Institut Jacques Monod, Université Paris Diderot) en lien avec les nombres de copie de gènes, l'aneuploidie et les effets de fond génomique non codant, une collaboration toujours active depuis 2007 en rapport à de nombreuses situations de génétique et d'évolution.

Dans les sections suivantes de ce chapitre je présente les principaux travaux en lien avec les publications correspondantes aux sujets ci-dessus.

5.2 Caractérisation des réseaux génétique

Les études théoriques de la topologie des réseaux caractérisent la manière dont les connexions entre les nœuds des réseaux sont organisées. Elles se fondent essentiellement sur le domaine mathématique de la théorie des graphes, dont l'histoire est ancienne et concerne des structures faites de nœuds connectés les uns et les autres par des liens pouvant-être de types différents. Comme introduit au chapitre 2, section 2.6, les recherches sur les graphes, le terme plus traditionnel des mathématiques pour parler des réseaux, ont conduit à la définition de nombreuses quantités mesurables permettant de caractériser des réseaux irréguliers et décrire différentes formes de complexités. Appliqués aux réseaux biochimiques ces descripteurs permettent des analyses que l'on voudrait utiliser pour comprendre les propriétés et les rôles possibles de différents nœuds suivant leur place dans le réseau.

5.2.1 Reconstruction de circuits génétiques

Entre 2003 et 2005 j'ai participé auprès de Florence d'Alché-Buc à l'époque au Laboratoire d'Informatique de l'Université Paris 6 Pierre et Marie Curie (LIP6) à des travaux en bioinformatique pour la recherche automatique de réseaux de régulation génétique à partir de données d'expression de gènes. Ces recherches informatiques d'algorithmes pour des situations aussi complexes que la régulation génétique demandent en effet pour leur application, la construction de modèles adéquats pour la représentation mathématique des processus dont on veut automatiquement déduire les propriétés. C'est dans la construction de cette modélisation, la recherche de situations d'applications pertinentes et l'incorporation des données biologique dans le traitement informatique qu'à consisté mon implication dans ces travaux méthodologiques ayant conduit aux publications suivantes :

- « Gene networks inference using dynamic Bayesian network », B.E. Perrin, L. Ralaivola, A. Mazurie, J. Mallet, S. Bottani, F. d'Alche Buc, *Bioinformatics*, **19**, II138-II138, (2003) ;
- « Inférence de réseaux d'interactions génique à l'aide de réseaux bayésiens dynamiques », Ralaivola L, Perrin BE, d'Alché-Buc F, Mazurie A, Bottani S, In *Journées Post-Génomiques de la Doua*, La Doua, France, (2003) ;
- « Réseaux bayésiens dynamiques pour l'inférence de structure de réseaux de régulation », Ralaivola L, Perrin BE, d'Alché-Buc F, Mazurie A, Bottani S, dans '*Informatique pour l'analyse du transcriptome*', Boulicaut JF and Gandrillon O editors, Lavoisier, pp.291-319 (2004) ;
- « A recurrent artificial neural network based on inertia principle for modeling gene regulatory network », F. d'Alche Buc, P-J Lahaye, T. Vujasinovic, S. Bottani, A. Mazurie , chapitre dans *Bioinformatics using Computational Intelligence Paradigms*, Ed. U. Seiffert , World Scientific Publishing, (2004).

L'établissement des réseaux biologiques à partir des données de génomiques, qui devenaient au courant des années 2000 de plus en plus abondantes grâce aux techniques à haut débit, est au cœur de l'exploitation de ces techniques et de la biologie des systèmes. C'est pour cette raison que j'ai commencé à m'y intéresser encore au démarrage de mon intérêt pour la biologie des systèmes. Je ne m'attarderais cependant pas dans ce mémoire sur ces contributions déjà anciennes, car il s'agit d'un axe de recherche auquel finalement je n'ai plus participé après ces travaux initiaux. Une fois la problématique et les bases de modélisation posées, les méthodologies en jeu sont de nature fortement algorithmique. Cependant, même si je n'ai pas poursuivi dans cette direction, les travaux auxquels j'ai participé dans ce champ ont été importants, en tous cas à titre personnel, car ils m'ont aidé à positionner mon travail ultérieur dans le cadre de la biologie des systèmes, davantage dans l'analyse des réseaux et la modélisation de mécanismes de régulation et de processus. En raison de leur rôle dans mon parcours et l'importance du thème de l'apprentissage automa-

tique de réseaux, je veux néanmoins mentionner même brièvement les fondements et teneurs de ces travaux.

Afin de mettre en œuvre les méthodes mathématiques d'ajustement et les algorithmes informatiques il est nécessaire de disposer au préalable de modèles décrivant les interactions entre les composants biologiques et leurs effets. Des modèles de la dynamique d'expression et des modèles des interactions possibles, dont l'ensemble des paramètres recherché correspond au réseau lui-même et à ses paramètres dynamiques.

En vue d'applications sur des données expérimentales issues d'instruments et d'expériences réelles, il est également nécessaire d'identifier la nature des informations pertinentes, de sélectionner les jeux de données valables, et les situations permettant l'application des méthodes informatiques. Une interaction étroite entre expérimentation, connaissances biologique et savoir faire théorique est indispensable. Un ensemble de tâches qui implique la représentation mathématique des phénomènes dans des modèles, le travail avec des données d'expérience, la mise au point et le paramétrage d'algorithmes, et enfin l'inférence automatique des réseaux. Une entreprise de collaboration interdisciplinaire qui demande de rassembler les compétences nécessaires de mathématiques, d'informatique, de modélisation et de compréhension biologique.

Apprentissage automatique de réseaux

Le développement des données biologiques et du développement concomitant de la modélisation que nous avons précédemment discuté à naturellement incité à l'application de méthodes d'apprentissage pour faciliter la recherche de corrélations et de liens dans les données en déduire automatiquement les réseaux d'interactions.

Les techniques d'apprentissage automatique comprennent tous les algorithmes permettant la détermination des paramètres d'un modèle afin qu'il corresponde « au mieux » avec les données. C'est une procédure standard de la modélisation, bien connue et maîtrisée pour des applications simples, d'ajustement d'une courbe dépendant d'une équation à un, deux ou trois paramètres avec un jeu de données complet. Régressions linéaires, algorithmes d'ajustement (« fit ») sont d'usage standard dans la pratique scientifique pour la recherche des paramètres d'une droite, d'une variation exponentielle, d'oscillations périodiques, etc. . . Mais des applications aussi communes que la recherche des pics et leurs largeurs dans des données temporelles bruitées sont déjà plus délicates.

Le problème devient un défi mathématique et informatique dès lors que les modèles comportent plus d'une poignée de paramètres, correspondent à des comportements mathématiques non réguliers et sont appliqués à des données bruitées, incomplètes et comportant des erreurs. Le sujet de l'apprentissage automatique, on dit également de l'« optimisation mathématique » ou encore de l'« operation research » est un domaine de recherche très actif depuis longtemps en raison de son importance stratégique pour tout type d'application. Un domaine qui sous l'intitulé « machine learning » a connu des ouvertures spectaculaires portées par l'explosion des données massives (« big data ») dans tous les secteurs d'activité ces quinze dernières années et le renouveau de l'intelligence artificielle.

De nombreuses techniques d'apprentissage ont ainsi été développées, mettant en œuvre différents principes. Parmi les catégories principales et utilisées dans les travaux avec Florence d'Alché-Buc on peut citer les techniques dites EM (expectation / maximisation) et les algorithmes génétiques. Dans tous les cas il s'agit d'explorer l'espace des paramètres afin de maximiser une fonction dite de *vraisemblance* qui décrit à quel point les données sont plausible en fonction d'un jeu de paramètre. Chaque combinaison de valeurs de paramètres est associée à une valeur de vraisemblance suivant des critères définis par un spécialiste du domaine d'application pour juger de la pertinence du modèle employant ces paramètres.

Les **algorithmes génétiques** explorent l'espace des paramètres à la recherche de la meilleure vraisemblance suivant un principe de sélection et évolution. Les combinaisons de paramètres correspondent à des organismes virtuels susceptibles de mutations aléatoires et capables de se reproduire plus ou moins efficacement en fonction de leur valeur adaptative qui correspond dans ce cas la vraisemblance du modèle. L'évolution de génération en génération d'une population de tels organismes virtuels permet l'exploration du paysage adaptatif et l'optimisation des paramètres.

Les algorithmes d'**expectation maximisation**, explorent l'espace des paramètres en itérant, à partir d'un jeu initial de valeurs, des cycles i) d'estimation des valeurs de changements de la vraisemblance dans différentes directions de l'espace multi-dimensionnel des paramètres et ii) du choix de la direction de maximisation du changement. Fondamentalement il s'agit d'estimer les gradients de variation du modèle et d'en déterminer le maximum. Des opérations cependant délicates dans le cas de modèle à nombreuses dimensions et nécessitant des algorithmes appropriés.

Applications de l'apprentissage de réseaux

Dans les publications avec Florence d'Alché-Buc nous avons ainsi testé des modèles et algorithmes de déduction de réseau testés sur des données temporelles d'expression de gènes par puces à ADN ou par RT-PCR pour *Escherichia coli* disponibles publiquement [COURCELLE et collab., 2001; GARDNER et collab., 2003; RONEN et collab., 2002]. Pour tester les méthodes on a choisi un système d'*Escherichia coli* déjà très bien connu par des méthodes classiques biomoléculaires et génétique et l'objectif de le retrouver automatiquement à partir des données. Le système de réparation SOS de *Escherichia coli* satisfait ces critères car constitué d'un nombre de gènes et interactions traitables, 8 gènes et 12 interactions bien établies. Les données à disposition permettent de connaître pour chaque gène des profils de 50 niveaux d'expression toutes les 6 minutes à partir desquelles on souhaite retrouver les liens en les supposant initialement inconnus.

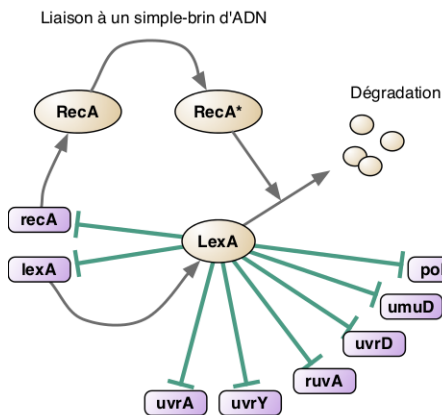


FIGURE 5.1 – Réseau de réparation SOS exploité par la bactérie *Escherichia coli* pour répondre aux dommages sur l'ADN. La protéine, RecA, a une affinité particulière pour les simples-brins d'ADN créés suite aux dommages. Une fois fixée à l'ADN, RecA subit une modification conformationnelle la rendant capable d'interagir avec une autre protéine, LexA qui réprime l'expression de plusieurs gènes dédiés à la réparation. La liaison de RecA à LexA induit la dégradation de cette dernière et entraîne une levée de la répression de ces gènes et la réparation de l'ADN [SUTTON et collab., 2000; WALKER, 1995].

Le modèle dynamique de réseau incorpore donc comme paramètres des poids pour chaque régulation possible, ainsi que des paramètres dynamiques pour l'expression temporelle de chaque gène. Dans l'optique de capturer de manière simple des effets de retard de l'expression du produit d'un gène suit à une modulation de sa régulation nous avons construit un modèle original effectif sous forme d'une équation différentielle du second ordre. On assimile ici l'effet du retard à un temps de relaxation dans un système inertiel amorti. Une analogie mécanique correspond à deux masses reliées par un ressort avec un fort amortissement : un déplacement d'une des deux masses induira celui de la seconde qui se stabilise au bout du temps caractéristique. Nous avons donc choisi de mettre en application l'algorithme d'apprentissage avec comme modèle de régulation génétique :

$$\frac{d^2 E_i(t)}{dt^2} + 2\lambda_i \omega_i \frac{dE_i(t)}{dt} + \omega_i^2 E_i(t) = \sum_j \omega_{ij} E_j(t), \lambda_i > 1. \quad (5.1)$$

où $E_i(t)$ est le niveau d'expression du gène i au cours du temps, ω_{ij} la force de couplage entre les gènes i et j , ω_i le coefficient caractéristique du temps de réponse du gène i à une modulation, λ_i un coefficient d'amortissement qui doit être supérieur à un pour éviter des oscillations sans signification biologique. Les interactions entre gènes sont représentées dans ce modèle de manière uniquement additive. Pour le cas du système SOS pris comme test ceci n'est pas un problème car les relations connues sont uniquement directes (un seul régulateur par gène). L'apprentissage du réseau signifie ici de déterminer à partir de données les valeurs ω_{ij} des paramètres d'interaction non nuls.

Pour tenir compte des variabilités et des bruits intrinsèques au système génétique et liés à la mesure nous avons fait une implémentation Bayésienne du modèle (5.1) dans laquelle les paramètres ne sont pas vus comme une valeur donnée avec un intervalle de confiance, mais sont plutôt caractérisés par des distributions des probabilités de valeurs de centre et largeurs devant être ajustées. Aussi le modèle est rendu stochastique en incorporant un terme de bruit gaussien, ainsi qu'une distinction entre l'état « caché » des gènes venant de leurs dynamiques propres suivant (5.1) et les quantités observables pour chaque gène sujettes à d'autres biais et fluctuations expérimentales.

La technique d'apprentissage d'expectation/maximisation a été appliquée sur cette version Bayésienne du modèle. Sa mise en œuvre n'est pas immédiate, elle demande une programmation dédiée du modèle et de l'algorithme et de nombreux paramètres techniques (seuils, initialisation, nombre de cycles, nombre de répétitions) doivent être réglées pour stabiliser les résultats. Une synthèse de ces explorations se trouve dans notre publication [D'ALCHÉ-BUC et collab., 2005]. Comme le montre la figure 5.2 la méthode est capable à bien capturer la dynamique du réseau : les variations des gènes les plus exprimés (recA, lexA, uvrA, et dans une moindre mesure umuD et uvrD) sont bien reproduites.

Après chaque apprentissage de la dynamique les paramètres peuvent être analysés pour extraire les régulations dans le réseau. Étant donné que les paramètres sont des valeurs continues dans un système bruité cette étape demande également des réglages afin d'établir les seuils caractérisant une probable interaction. Le réseau reconstitué est présenté dans le graphe de droite de la figure 5.2 en comparaison avec le réseau réel sur la partie gauche. On peut voir que 8 interactions probables sont identifiées, dont les inhibitions de lexA lui-même, recA et uvrA par lexA, ainsi que l'activation de lexA par recA. On trouve aussi un faux positif uvrA par recA (qui pourrait correspondre à une régulation indirecte $recA \rightarrow RecA \dashv LexA \dashv uvrA$.)

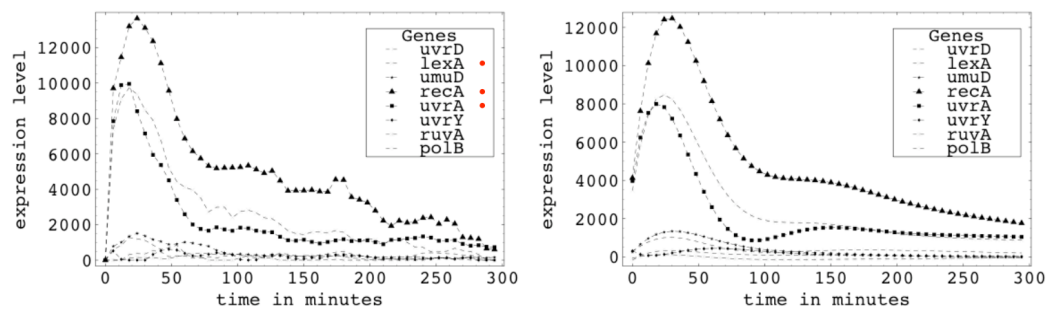


FIGURE 5.2 – **Profils d'expression des 8 gènes du réseau SOS.** Gauche : profils d'expression de 8 gènes du réseau SOS. Droite : profils moyens 'expression appris avec le modèle inertiel 5.1 bayésien et l'algorithme d'expectation/maximisation. Source : [D'ALCHÉ-BUC et collab., 2005]

Ces résultats sur le système de réparation SOS de *Escherichia coli* étaient très prometteurs de l'algorithme d'apprentissage et du modèle considéré même si l'intégralité du réseau cible n'a pas été parfaitement reconstruit. Il faut cependant tenir compte que le système testé restait de petite taille, avec une connectivité de faible degré et des régulations simples. Le passage à une échelle plus grande, ainsi que des régulations plus complexes restait, et l'est encore de nos jours un défi considérable. D'un point de vue de la modélisation, le modèle inertiel atténué considéré dans

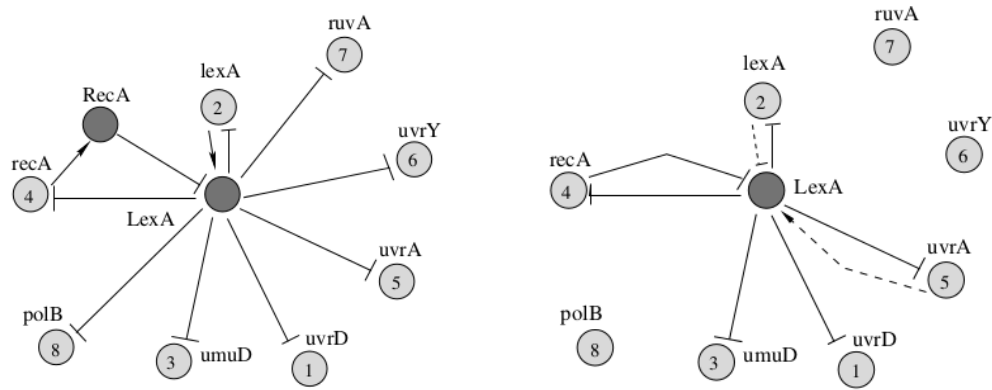


FIGURE 5.3 – **Comparaison du réseau SOS connu et inféré.** Réseau du système SOS connu à gauche et à droite le réseau établi par la méthode d'apprentissage. Source : [D'ALCHÉ-BUC et collab., 2005]

l'étude est intéressante et a permis un traitement de l'optimisation dans un cadre linéaire. Mais son application dans un contexte d'interactions non linéaires n'est pas évident.

Depuis 2005 et ces travaux le domaine général de l'apprentissage machine (« machine learning ») a connu un développement spectaculaire dans tous les secteurs et s'est imposé dans de larges pans de recherche en biologie. L'inférence de réseau de régulation n'est pas en reste [CARRÉ et collab., 2017]. Ces recherches sont fortement spécialisées en informatique, intelligence artificielle et mathématiques appliquées et depuis mon exploration de ces questions j'ai suivi d'autres orientations.

5.2.2 Structure du réseau de régulation de la levure

Les études théoriques de la topologie des réseaux caractérisent la manière dont les connexions entre les nœuds des réseaux sont organisées. On a déjà introduit au chapitre 2 section 2.6 le champ de la théorie des graphes qui se rapporte à ces structures et quelques quantités mesurables permettant de les caractériser. Appliquées aux réseaux biochimiques ces descripteurs permettent des analyses utiles pour comprendre leurs propriétés et les rôles possibles de différents nœuds suivant leur place dans le réseau.

Au moment de ces recherches entre 2000 et 2006, plusieurs études avaient déjà porté sur la topologie des réseaux macro-moléculaires suggérant différents types de propriétés caractéristiques telle qu'une organisation de type « petit monde » dans des réseaux d'interaction protéine-protéine [SOL et collab., 2005] et des réseaux métaboliques [WAGNER et FELL, 2001], et pour ces mêmes réseaux aussi une distribution en loi de puissance invariante d'échelle [JEONG et collab., 2001] (réseau protéine-protéine), [ALBERT et BARABASI, 2002] (réseau métabolique).

Je présenterai ci-dessous tout d'abord la première caractérisation globale jamais effectuée des propriétés du réseau de régulation transcriptionnel chez l'organisme modèle eucaryote le plus simple la levure. Ensuite je discuterai d'une autre étude sur l'organisation modulaire toujours chez la levure du réseau de régulation additionné des interactions entre protéines.

Topologie du réseau transcriptionnel de la levure

Notre article « Topological and causal structure of the yeast transcriptional regulatory network » avec Nabil Guelzin, François Képes et Philippe Bourguin [GUELZIM et collab., 2002] est la première publication d'un réseau transcriptomique eucaryote composé à partir de données génétiques annotées de facteurs de transcriptions et de leurs cibles. Le réseau inclut 491 gènes de la levure et 909 interactions validées, figure 5.4.

Nous avons ensuite fait une analyse de la topologie de ce réseau complexe mettant en évidence

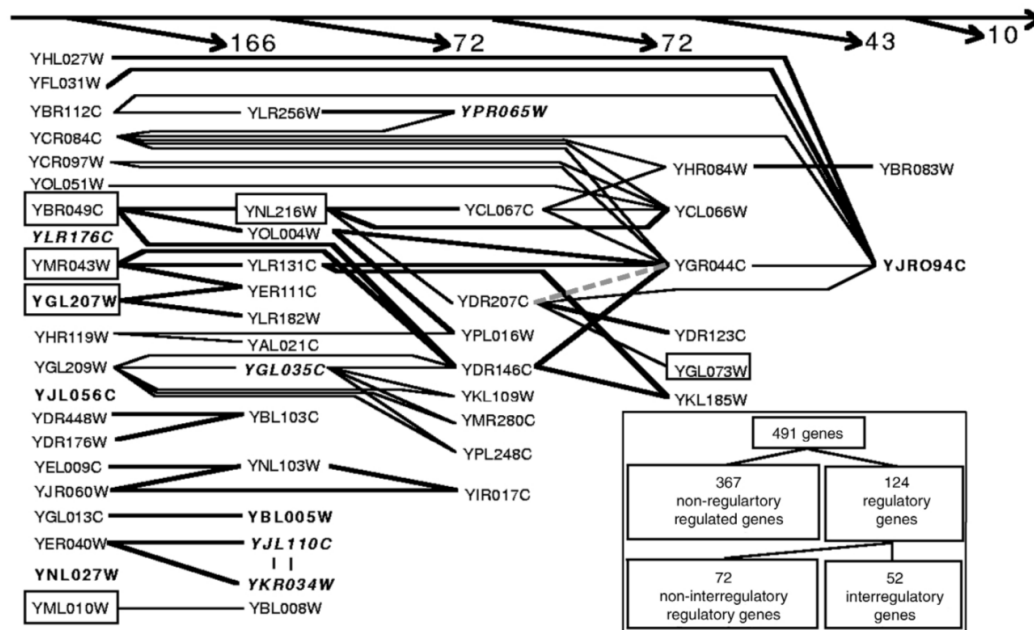


FIGURE 5.4 – Réseau d’*inter-régulation* transcriptionnelle chez la levure. Le réseau représente les interactions connues *entre* les 52 gènes identifiés codant pour des facteurs de transcription agissant sur d’autres facteurs de transcription (inter-régulations). Le réseau est ordonné de gauche à droite : les gènes ne possédant pas de régulateurs connu ou étant uniquement auto-régulés sont dans la colonne de gauche. Les gènes sont classés dans chaque colonne successive de telle manière que tous leurs régulateurs se trouvent sur la gauche. La ligne orientée et les flèches diagonales sur le haut indiquent pour les gènes de chaque colonne les nombres de gènes cibles non régulateurs contrôlés par les facteurs de chaque colonne (les noms de ces cibles ne sont pas indiqués). L’ordonnancement de gauche à droite indique donc une suite d’influence causale et les conséquences possibles sur d’autres gènes de l’altération de l’un en amont (plus sur la gauche). Les noms en gras indiquent les gènes avec s’auto-régulant, en italique l’auto-régulation négative et l’encadré entoure les gènes essentiels. Les liens en gras indiquent une activation, les fins une inhibition et en tirets une régulation duale (activation ou inhibition). L’encadré indique les nombres des différentes catégories de gènes de la base de données des régulations transcriptionnelles de la levure établie pour cette publication. Le réseau a été établi par une collecte manuelle de résultats sur les régulations dans *S. Cerevisae* publiées dans des articles scientifiques. Il ne s’agit pas ici d’une analyse de données à haut débit, mais de l’exploitation de résultats individuels sur les facteurs de transcription obtenus par des méthodes classiques et supposés fiables. Mais certainement pas exhaustifs. Source : [GUELZIM et collab., 2002]

une distribution exponentielle décroissante du nombre de régulateurs par gène régulé mais d’une distribution en loi de puissance du nombre de gènes cibles par protéine régulatrice, figure 5.5.

Pour caractériser plus finement la structure du réseau, différents descripteurs topologiques ont été calculés et comparés à trois modèles différents de réseaux aléatoires dirigés : un modèle basé sur les distributions de connectivités entrantes et sortantes empiriques, où les connections entre nœuds sont tirées aléatoirement mais selon ces distributions déterminées expérimentalement; un modèle basé sur les distributions ajustées de la figure 5.5 exponentielles pour les liens entrant et en lois de puissance pour les sortants; un modèle de distribution de Poisson pour les connections entrantes et sortante. Plusieurs descripteurs ont été calculés sur ces différents réseaux, indiqués dans le tableau 5.1 : les nombres moyens de liens directs, de liens secondaires, la distance moyenne sur le réseau, le nombre d’interactions triangulaires, des coefficients de clustering pour réseaux dirigés, les nombre de boucles d’auto-régulation et de n’importe quelle taille.

Ces comparaisons mettent en évidence comment le réseau biologique est beaucoup plus fragmenté que les différents modèles aléatoires qui sont caractérisés par des interconnexions locales et du clustering beaucoup plus importants. Ceci se reflète aussi dans le nombre de boucles de rétroactions directes ou impliquant plusieurs gènes qui sont 5 à 15 fois plus nombreuses que dans la modèles aléatoires. Dans l’ensemble le réseau de la levure combine un diamètre moyen faible, un

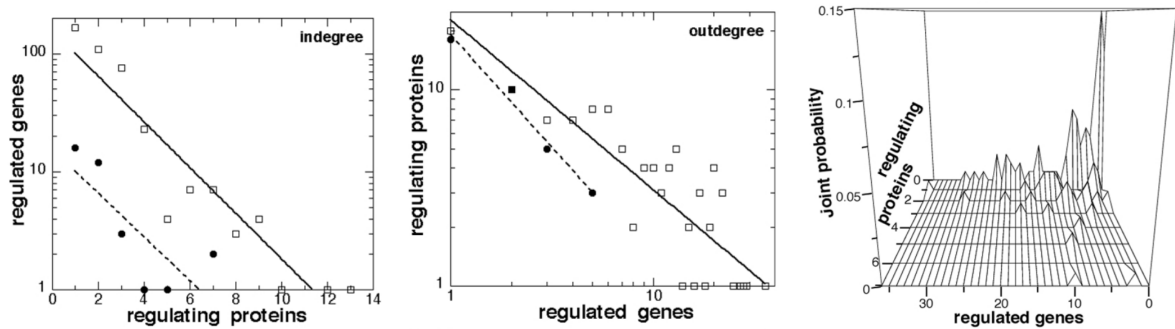


FIGURE 5.5 – **Connectivité du réseau de régulation transcriptionnel de la levure.** a) Distribution semi-log des connectivités entrantes. Les nombres de protéines régulatrices par gène régulé suivent une distribution exponentielle, aussi bien pour l'ensemble des 402 gènes régulés (367 gènes non régulateurs et 35 gènes inter-régulateurs, moins les 17 gènes inter-régulateurs sans régulateur connu, pour un ensemble de 909 liens, carrés et ajustement courbe continue : $p_k = 157e^{-0,45k}$, $R = 0,99$); que pour le sous ensemble des 35 gènes inter-régulateurs (72 liens, cercles pleins et ajustement en pointillés : $p_k = 15e^{-0,43}$, $R = 0,94$). b) Distribution des liens sortants en graphe log-log. La distribution des nombres des gènes régulés par facteur régulateur est ajusté au mieux par une loi de puissance, aussi bien pour les 124 gènes régulateurs (909 connections, carrés vides, ajustement ligne continue : $p_k = 23k^{-0,87}$, $R = 0,95$), que pour le sous-ensemble des 37 gènes inter-régulateurs (72 liens, cercles pleins, ajustement ligne en pointillés : $p_k = 19k^{-1,14}$, $R = 0,99$). Des ajustements inversés (loi de puissance pour les connectivités entrante et exponentielle pour les connectivités sortantes) donnent des coefficients de corrélation généralement plus faibles (respectivement $R = 0,83$; $0,98$ et $R = 0,88$; $8,91$) et des pentes significativement plus différentes entre les régulations totales et les inter-régulations. c) Distribution conjointe. La probabilité qu'un gène choisi aléatoirement possède un nombre de liaison entrantes donnée et un autre donné de liens sortants est représentées sur ce graphe linéaire en fonction des nombres de gènes en entrée (« regulating proteins ») et du nombre de gènes en sortie (« regulated genes »). Source : [GUELZIM et collab., 2002]

degré élevé de clustering local, et un grand nombre de boucles de rétroactions ainsi qu'une fragmentation importante. Ces caractéristiques qui le distinguent de réseaux aléatoires doivent refléter des contraintes évolutives.

Cette caractérisation a été effectuée dans les années 2000 à 2002, exploitant des données classiques de génétique. Elle englobait un nombre limité, mais a priori très fiable, de gènes et interactions. Déjà à l'époque de la publication des données issues de technologies à grande échelle ainsi que des analyses bioinformatiques de génomes devenaient disponibles. Ainsi à partir d'études d'homologie de séquences au moins 77 gènes de levure supplémentaires avaient été identifiées, et des nombres très variables d'autres protéines régulatrices supplémentaires mis en évidence par d'autres groupes avec des techniques de puces à ADN ou d'immunoprécipitation de la chromatine (ChIP-chip). Étant donné ces variabilités de résultats on avait choisi de ne pas les inclure dans notre étude qui se voulait focalisée sur des données à haute fiabilité. Dans une étude plus récente de 2009 centrée sur des questions d'organisation hiérarchique JOTHI et collab. [2009] ont constitué un réseau de 13 385 interactions entre 4503 gènes, incluant 158 facteurs se liant à l'ADN et gènes cibles à partir de données biochimiques et d'expériences ChIP-chip qu'il pourrait être intéressant de réanalyser selon l'approche de notre publication. L'étude de la régulation transcriptionnelle chez la levure et de son organisation est un sujet toujours actif, en particulier les questions de hiérarchies [ZAMAL et RUTHS, 2012]. Dans une revue de 2003 dans *Genetics* T.R. Hughes [HUGHES et DE BOER, 2013] conclut ainsi :

Despite the many advances—and the occasional perception among researchers outside the field (and, in our experience, funding agencies) that mapping yeast transcriptional regulatory networks is a task that has been completed—the functions of roughly half of all yeast TFs remain opaque, and it is difficult to explain what exactly is happening at most individual promoters. Where do we go now, to get a clearer and more comprehensive picture?

Table 1 • Structure of the yeast transcriptional regulatory network

	Connectivity distributions			
	(a) Actual data	(b) Empirical	(c) Expo/Power	(d) Poisson
Giant component % ^a	0	0	0	77
$\langle in_1 \rangle (= \langle out_1 \rangle)$ ^b	1.9	1.9	2.5	1.9
$\langle in_2 \rangle (= \langle out_2 \rangle)$ ^c	1.0	1.3	1.4	3.7
$\langle in_{\infty} \rangle (= \langle out_{\infty} \rangle)$ ^d	4.0	7.2	7.0	NA ^h
N_{Δ} ^e	135	29	26	7
SC_{in} ^f	0.050	0.010	0.010	0.004
SC_{out} ^f	0.010	0.002	0.002	0.004
FC_1 ^g	10	0.7	0.6	2
FC_{∞} ^g	11	2.2	1.4	NA ^h

^aPercentage of genes in the largest connected component (with at least one oriented path between any couple of vertices). ^bAverage number of regulating proteins (regulated genes) one step away. ^cIdem two steps away. ^dIdem at all distances (average component size). ^eNumber of oriented triangular interactions. ^f SC_{in} (SC_{out}), upstream (downstream) semi-clustering coefficient. ^g FC_1 (FC_{∞}), number of feedback circuits comprising one (any number of) gene. ^hNot directly applicable, owing to the presence of a giant component in (d). The features of the actual genetic network (a) were compared with those of a directed random graph with an equal number of vertices and edges (see Web Note B online). Connectivity distributions were as empirically observed (b), or followed the exponential (power, respectively) law for arriving (departing) regulations that had been determined from Fig. 2 (c), or followed a Poisson law (d).

TABLEAU 5.1 – **Caractéristique du réseau de régulation transcriptionnel de la levure *S. cerevisiae*** Source : [GUELZIM et collab., 2002]

Il faut en effet garder à l'esprit que les descripteurs topologiques généraux sont facilement sensibles à des biais et des erreurs systématiques. Ainsi plusieurs nouveaux examens des mêmes réseaux, avec une autre méthodologie ou des données différentes, ont remis en question certaines conclusions comme l'invariance d'échelle des réseaux de protéines [PRZULJ et collab., 2004] ou la propriété de petit monde des réseaux métaboliques [ARITA, 2004]. Les conclusions générales sur les topologies des réseaux biologiques sont délicates et peuvent être facilement hâtives.

5.2.3 Évaluation des motifs de réseau dans le génome de la levure

Dans l'article « An evolutionary and functional assessment of regulatory network motifs in yeast » avec A. Mazurie et M. Vergassola, nous avons examiné l'hypothèse devenue très en vogue des motifs des réseaux biologiques. Une idée introduite peu de temps avant par Uri Alon [MILO et collab., 2002] suggérant l'existence de formes de sous circuits privilégiés dans les réseaux biologiques constituant des briques fondamentale de construction de ceux-ci : « simple building blocks of complex networks. » [MILO et collab., 2002]. Notre étude, basée sur des données du réseau mixte d'interaction de régulation transcriptionnelles et d'interactions protéines-protéines de la levure, nous a conduit à récuser la notion de l'existence de tels motifs de connectivité qui seraient sur-représentés dans la structure des réseaux en raison du rôle fonctionnel particulier de leur circuit de câblage.

Notion de motifs de réseau Selon les auteurs originaux et de nombreux partisans depuis, certains types de circuits seraient en surnombre dans les réseaux biologiques car sélectionnés par l'évolution naturelle en raison de propriétés remarquables qui leur seraient associées. Cet effet de surnombre est défini en comparaison avec une version aléatoire des réseaux étudiés, donc par rapport à un réseau de référence dénouée d'un historique d'évolution naturelle, de même taille, avec la même distribution de connectivité, mais des liens ré-organisés au hasard. Uri Alon et ses collègues ont ainsi mis en évidence l'existence de tels schémas de connections « surabondants » dans de nombreux réseaux complexes très différents. Dans des réseaux biologiques de régulation génétique, dans les réseaux de neurones, des réseaux tropiques (chaînes alimentaires) [LOTEM et collab., 2004; SHAI et collab., 2002], mais aussi dans des réseaux technologiques comme les circuits électroniques et l'internet [ALON, 2003; MILO et collab., 2002]. Dans tous ces exemples certaines formes de sous-circuits typiques, qui dès lors ont été nommés motifs, se comptent en

nombre plus élevé dans les données de connectivités réelles mesurées, que dans des versions réorganisées aléatoirement de ces réseaux préservant seulement la distribution de connectivité, voir figure 5.6.

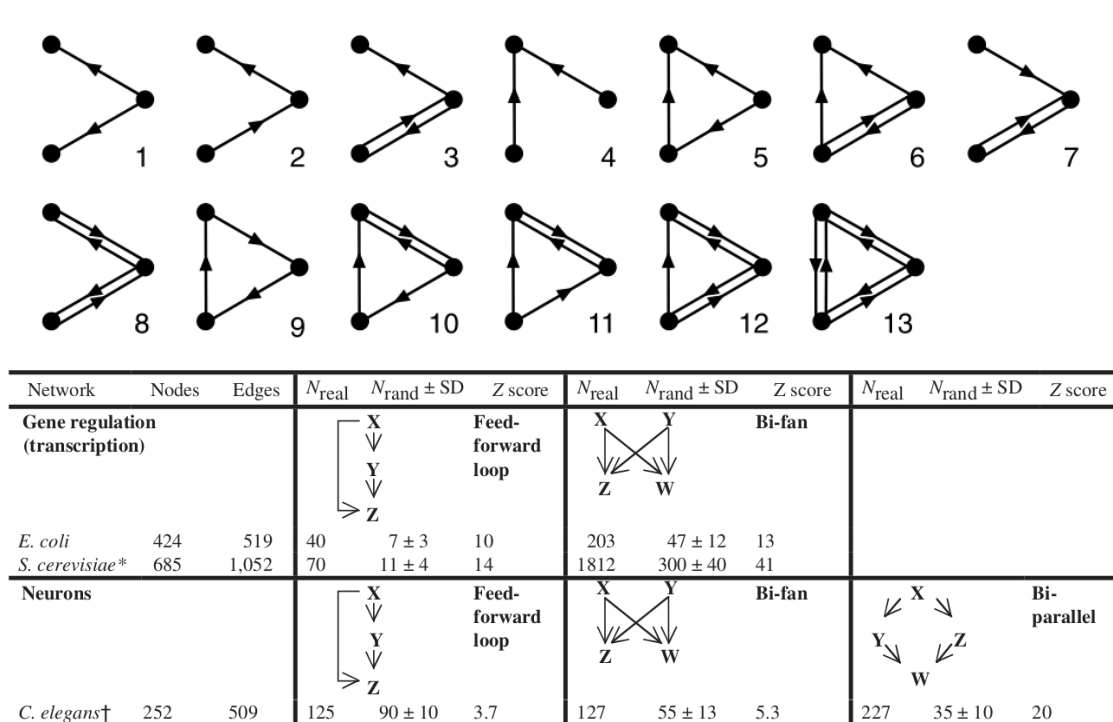


FIGURE 5.6 – Schémas de connexions à 3 gènes et motifs de réseau A) L'ensemble des 13 schémas de connexions possibles avec trois nœuds et un type possible d'interaction. La flèche représente tout type d'interaction suivant le contexte d'application : une régulation transcriptionnelle pour les réseaux génétiques, une synapse dirigée dans le cas de réseaux de neurones, une consommation alimentaire dans un réseau trophique, etc. ... La recherche de motifs est l'identification de schémas de cette nature sur-représentés dans un réseau réel par rapport à une version aléatoire. B) Motifs identifiés dans les réseaux de régulation transcriptionnelle de *E. coli* et *S. cerevisiae*, ainsi que dans le réseau des connexions synaptiques chez *C. elegans*. N_{real} et N_{rand} sont les nombres de motifs identifiés respectivement dans le réseau réel et sa version aléatoire, le score Z est la différence de ces valeurs sur la déviation standard des mesures sur les réseaux aléatoires. Source : [MILO et collab., 2002]

Cette méthode d'analyse de réseaux a conduit à proposer une théorie selon laquelle la surabondance de certains schémas de connectivité serait le résultat de propriétés particulières de ceux-ci. Ils seraient donc sélectionnés par l'évolution naturelle dans le contexte biologique, ou par les ingénieurs dans le cadre technologique. En biologie, chaque classe de motif représenterait alors un type d'activité biomoléculaire spécifique dépendant du schéma de câblage. La fonction correspondante sélectionnée par l'évolution naturelle conduirait à privilégier l'apparition de ces motifs dans les réseaux naturels [ALON, 2007; MANGAN et ALON, 2003; SHAI et collab., 2002]. Les motifs seraient similaires aux circuits et composants électroniques standards, qui se trouvent utilisés dans de nombreux contextes différents d'application. La figure 5.7 donne ainsi un exemple de dynamique fonctionnelle attribuée à un des motifs les plus représentatifs la « feed forward loop » [SHAI et collab., 2002]. la fonction de filtrage qui peut y être associé trouve un sens dans de nombreuses situations biologiques avec des gènes différents.

Malgré la simplicité de l'approche proposée par Uri Alon et son attrait biologique certain, notre étude conclut, sans ambiguïté pour ce qui nous concerne, à une interprétation contradictoire, selon laquelle l'abondance statistique des motifs n'a aucune connexion évidente avec des conséquences évolutives et fonctionnelles *in vivo*.

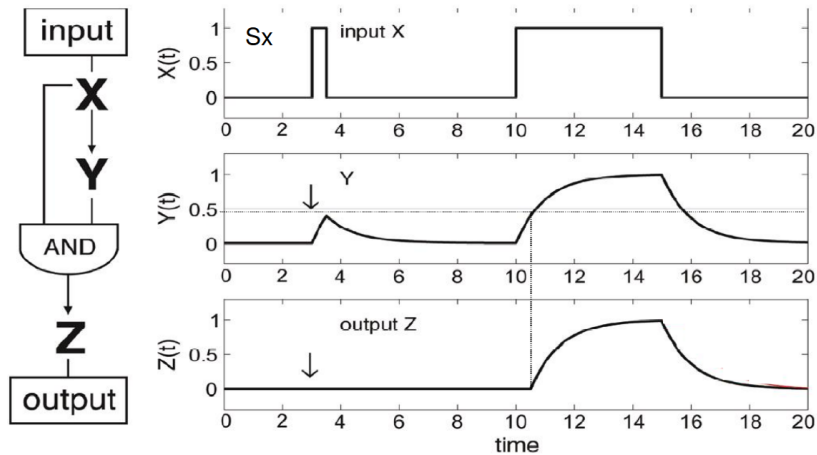


FIGURE 5.7 – Exemple du rôle fonctionnel attendu d'un motif de type « Feed Forward Loop » avec un opérateur « AND » sur le gène en sortie. Ce type de motif est associé au filtrage de signaux transitoires : Un pulse de signal S_x sur le gène en entrée X à l'instant $t = 3$ ne laissera pas suffisamment de temps pour une accumulation suffisante du produit de Y afin de dépasser le seuil d'activation de Z qui en raison de son opérateur de type « AND » nécessite la présence simultanée suffisante de ses deux régulateurs en entrée. Par contre un signal persistant S_x comme à partir du temps $t = 10$ permet l'expression de Z. Source : [SHAI et collab., 2002]

Le réseau de notre étude : régulation transcriptionnelle et interactions protéine-protéine chez *S. cerevisiae* Le réseau mixte sur lequel nous nous sommes basés reprend les données des interactions de transcription de ma publication [GUELZIM et collab., 2002] discuté plus haut (section « Topologie du réseau transcriptionnel de la levure ») et les données d'interaction de protéines de la « Database of Interacting Proteins » [DIP]. Nous avons appliqué le même algorithme de détections de motifs que l'article original à ces données pour des motifs à deux gènes (ordre $n = 2$) et à trois gènes ($n = 3$), voir figure 5.8. Les motifs représentés correspondent aux circuits de ces tailles, dont l'abondance dans le vrai réseau biologique de *Saccharomyces cerevisiae* est plus élevée que dans des réseaux aléatoires construits en connectant le même nombre de nœuds suivant la même distribution de connectivités que le réseau réel.

Les 12 motifs de la figure 5.8 sont présents en 504 instances dans le réseau, c'est-à-dire des combinaisons différentes des gènes spécifiques composant chaque instance d'un même motif. Cependant comme le montre la figure 5.9 qui représente le sous réseau de tous les motifs, seulement 4 de ces instances sont indépendantes les unes des autres. La grande majorité implique au contraire le partage d'une partie des mêmes gènes et des mêmes liens dans différentes instances du même motif ou d'autres motifs. Seulement 256 gènes différents en tout participent à des instances de motifs, alors que 1487 gènes différents seraient impliqués si toutes les instances étaient indépendantes. Le partage de liens et/ou de gènes entre plusieurs instances de motifs est la règle. Or le partage de mêmes liens et gènes avec circuits différents doit certainement affecter le type de fonction, le comportement physiologique, assuré par chaque instance. Les propriétés spécifiques que l'on voudrait attribuer à un schéma de connection doivent se retrouver sérieusement perturbées. Cette non indépendance des circuits complique en tous cas déjà l'interprétation originale des motifs et de leurs propriétés spécifiques. D'autres facteurs viennent encore compliquer le principe.

Pour clarifier notre impression venant de l'observation de tous les chevauchements d'instances, nous avons complété l'analyse par deux autres études. D'une part une analyse phylogénétique de conservation d'interactions de motifs entre espèces de levures différentes. D'autre part une étude fonctionnelle détaillée d'exemples d'instances.

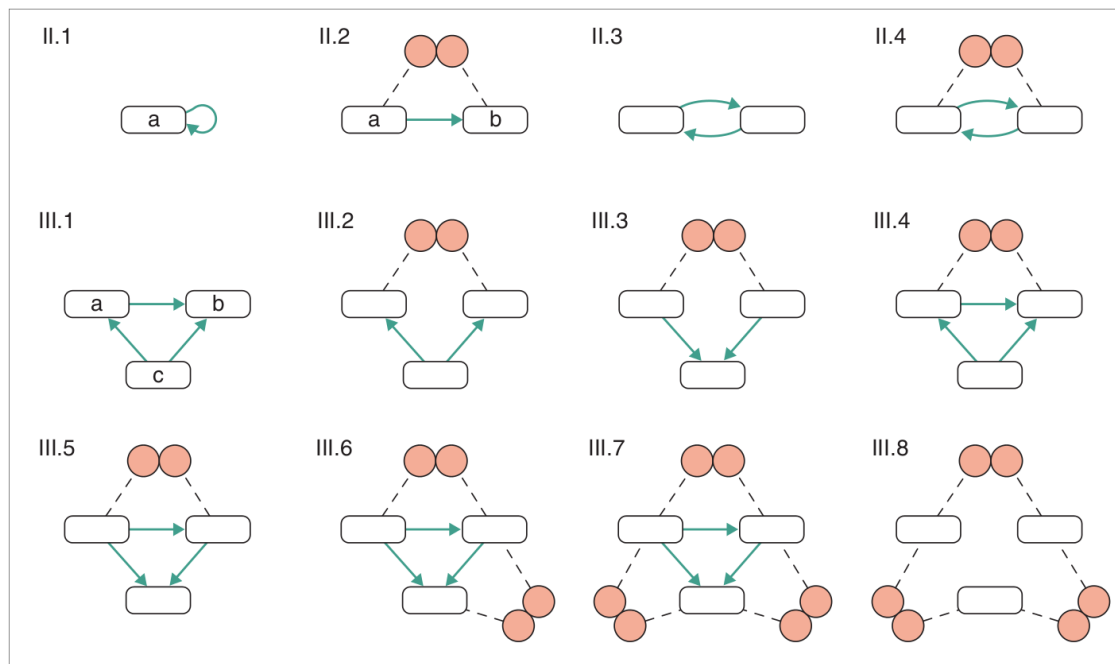


FIGURE 5.8 – **Types de motifs d'ordre $n = 2$ et $n = 3$ pour le réseau mixte transcriptionnel et protéines-protéines.** Les liens dirigés verts avec des flèches représentent les régulations transcriptionnelles, deux liens en tirets avec des cercles en contact représentent des interactions protéine-protéine non dirigées. Source : [MAZURIE et collab., 2005]

Études phylogénétiques des instances de motifs La comparaison de profils phylogénétiques d'instances de motifs entre espèces différentes de levure vise à identifier l'existence d'une pression de sélection agissant sur celles-ci. En effet, si les motifs avaient une signification évolutive particulière on s'attendrait à ce que leurs instances aient tendance à être conservées entre espèces. Afin de tester la présence d'une pression évolutive spéciale nous avons ainsi réalisé une étude comparative de conservation des gènes reliés dans les instances identifiées chez *Saccharomyces cerevisiae* avec des recherches d'homologie dans les séquences de 4 autres types d'hémiascomycètes, *Candida glabrata*, *Kluyveromyces lactis*, *Debaryomyces hansenii* et *Yarrowia lipolytica*.

L'analyse des profils phylogénétiques présenté dans la figure 5.10 mets très clairement en évidence qu'il n'existe aucune différence statistique entre les conservations d'interactions issues d'une instance de motifs et n'importe quelle interaction. Cette analyse comparative indique sans ambiguïté que les instances de motifs ne subissent aucune pression évolutive comparée à tout autre schéma d'interaction.

Études détaillée des fonctions d'instances de motifs La théorie des motifs de réseaux repose sur l'hypothèse d'une signification fonctionnelle particulière pour chaque type de schéma de câblage et étudiée dans les publications originales du groupe de Uri Alon.

Nous avons voulu vérifier si les comportements attendus se retrouvaient bien dans des situations biologiques très bien comprises. Nous avons ainsi sélectionné des exemples d'instances associées à des fonctions biologiques bien connues dans *S. cerevisiae*. Sur ces cas de figure nous avons recherché si les mécanismes fonctionnels envisagés par la théorie des motifs correspondaient effectivement à la compréhension biologique. Nous avons retenu cinq exemples de sous réseaux génétiques associés très bien documentées et incluant de nombreuses instances de motifs :

- le circuit de régulation de la synthèse de la méthionine
- le système de répression du catabolisme de sources azotées (« Nitrogen catabolite repression (NCR) system »)

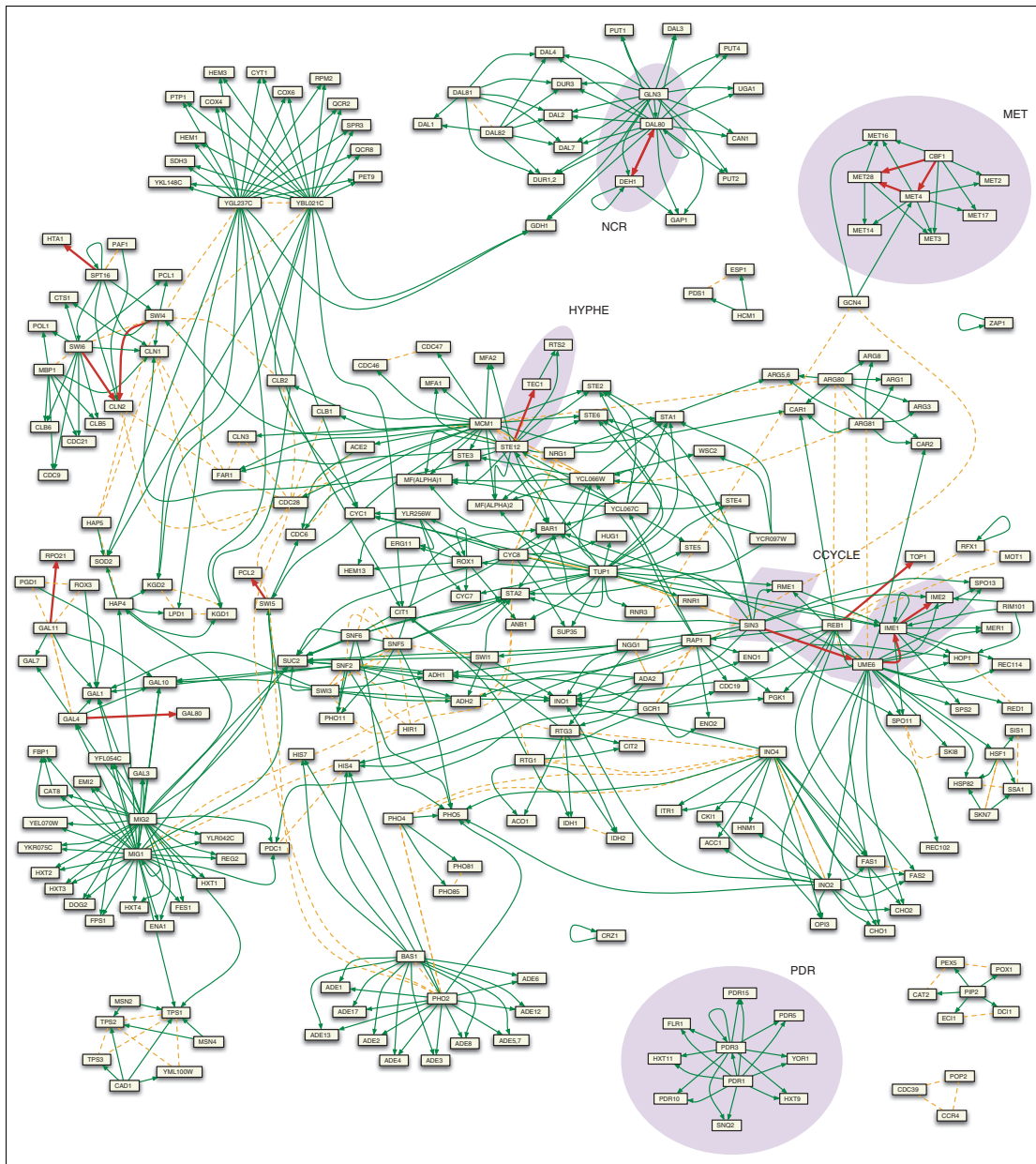


FIGURE 5.9 – **Instances des motifs dans la levure.** Le réseau représenté est le sous réseau mixte transcriptionnel et interactions protéine-protéine de *S. cerevisiae* constitué par toutes les instances de motifs d'ordre 2 et 3. Vert : régulation transcriptionnelle pure de la cible par le produit du gène régulateur; rouge, régulation transcriptionnelle et interaction protéine-protéine des deux gènes et de leurs produits; ligne en tirets, interaction protéine-protéine pure. Les circuits sous fond coloré ont été étudiés en détail dans la publication par rapport aux fonctions putatives de motifs.

- le système MAPK de choix entre croissance pseudohyphale ou accouplement (« Pseudohyphal growth/matingMAPK system »)
- la régulation des gènes d'initiation de la méiose (« Regulation of early meiotic genes »)
- le système de résistance pléiotropique aux drogues (« Pleiotropic drug resistance (PDR) system »)

Se fondant sur une littérature abondante de génétique et biologie moléculaire, nous avons étudié en détail le rôle des motifs dans les sous-réseaux de tous ces exemples très documentés. Dans tous ces cas, les schémas de connectivité des motifs qui s'y retrouvent ne semblent avoir aucun rôle central de régulation dans les processus biologiques concernés. Nous n'avons pas pu identifier dans les explications des comportements fonctionnels détaillés des systèmes biologiques

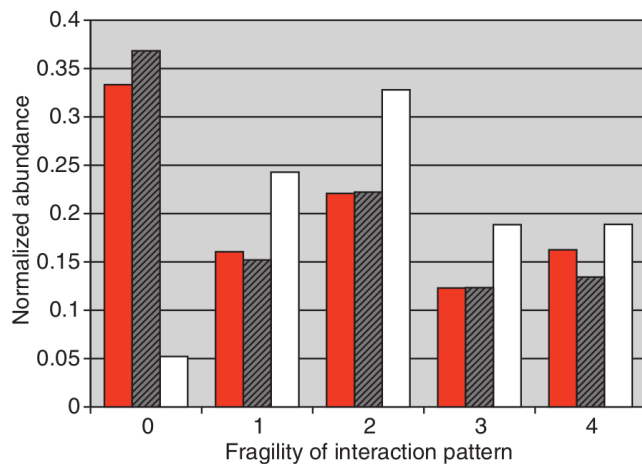


FIGURE 5.10 – **Profils phylogénétiques des interactions.** Rouge : 2 gènes liés dans une instance d'un motif; noir : 2 gènes liés dans une interaction de régulation chez *S. cerevisiae* mais dans un circuit quelconque, pas nécessairement suivant un motif; blanc : la fragilité de la conservation de 2 gènes pris au hasard indépendamment d'une interaction ou non. Chaque valeur de fragilité correspond au nombre de fois que les gènes de *S. cerevisiae* d'un couple des 3 types précédents sont co-conservés de manière homologue dans les 4 autres espèces. Une petite valeur indique que les 2 gènes d'un couples ont tendance à être conservés ensemble dans les autres génomes et indique une pression sélective pour préserver l'interaction. Source : [MAZURIE et collab., 2005]

ci-dessus aucun élément se référant aux propriétés attendues dans la théorie des motifs.

En effet, la réduction du fonctionnement des motifs aux régulations transcriptionnelles est trop simplificatrice pour rendre compte des comportements. La réception et le traitement de signaux cellulaires ne se réduit par à la régulation transcriptionnelle et des interrupteurs génétiques. D'autres mécanismes tels que phosphorylation, la dégradation contrôlée, la séquestration de protéines, le transport et la multimérisation dans des complexes multi-protéiques sont également au cœur de la logique de fonctionnement des sous-réseaux étudiés. Dans les exemples examinés ci-dessus il n'a pas été possible de dés-enchevêtrer la régulation génétique et les interactions protéines-protéines de l'ensemble de l'environnement cellulaire et biochimique pour interpréter leur fonctionnement. On ne peut réduire le comportement physiologique à un espace moléculaire donné.

Conclusion Nos différents résultats montrent clairement que la surabondance de motifs dans les réseaux de transcription et d'interactions protéines-protéines ne peut être reliée à un rôle fonctionnel particulier. Les instances de motifs se chevauchent fortement et interfèrent. Il n'existe pas de pression sélective pour conserver les instances de motifs. Enfin, les mécanismes de régulation des 5 processus biologiques impliquant des motifs s'expliquent sans aucune référence aux motifs et en contradiction avec l'interprétation fonctionnelle de ceux-ci.

Notre publication a ainsi remis en cause sérieusement la signification fonctionnelle des motifs dans les réseaux de régulation. D'autres travaux indépendants [GINOZA et MUGLER, 2010; INGRAM et collab., 2006; KNABE et collab., 2008; KONAGURTHU et LESK, 2008; MESHI et collab., 2007; PAYNE et WAGNER, 2015a; RANDRUP et collab., 2004; SOLÉ et VALVERDE, 2006] sont venus également confirmer nos conclusions, et on peut à ce jour rester très réservés sur cette hypothèse marquante durant les années 2000 de la biologie des systèmes.

5.3 Circuits génétiques oscillants

Les travaux regroupés dans cette section se sont étalés entre 2007 et 2011 en collaboration avec Basile Grammaticos du laboratoire IMNC. Nous nous sommes intéressés à l'étude de petits circuits génétiques formels oscillants impliquant un à quelques gènes en interactions et inspirés de situations biologiques. Notre objectif était de comprendre les facteurs théoriques mathématiques déterminants ces comportements et d'étudier des manières de les représenter mathématiquement.

quement. En particulier nous nous sommes intéressés à des modèles incluant des retards. Dans un premier temps de manière très classique sous forme d'équations différentiels, par la suite aussi dans un cadre stochastique Monte-Carlo. L'usage de retards dans des modèles, où intervient un délai entre l'instant d'une action et son effet, est classique et bien connue notamment en théorie du contrôle. Peut-être en raison de leur implémentation plus délicate et de difficultés analytiques, les modèles à retard avaient cependant été moins explorés dans le contexte génétique. Une raison pour laquelle il nous avait paru intéressant d'étudier les oscillations génétiques par ce biais. Mais avant de présenter nos applications particulières il convient d'introduire le contexte des oscillations génétique qui les a motivés.

5.3.1 Oscillations génétiques : motivation et contexte

Parmi les comportements temporels biologiques les plus caractéristiques, rythmes et processus répétitifs se retrouvent dans de nombreux processus du vivant dont ils déterminent les propriétés fondamentales. Des comportements périodiques ont aussi été identifiés au niveau moléculaire et génétique dans de nombreux organismes, une variété de contextes et des échelles temporelles allant de quelques heures à une journée. Les rythmes biologiques ont toujours suscité l'attention des mathématiciens et physiciens fascinés par l'origine des cycles et régularités dans les comportements naturels. En effet les périodicités du vivant sont les manifestations dynamiques régulières et robustes de multitudes d'éléments en interactions et soumises aux aléas des environnements naturels. Ce sont des comportements complexes issus d'actions réciproques et de régulations diverses dans les systèmes où ils ont lieu. L'étude du vivant, la biologie, identifie les comportements et les pièces en actions dans ces mécanismes mais nécessite les apports des mathématiques et de la physique pour décrire précisément de tels phénomènes dynamiques et en comprendre les ressorts [GOLDBETER, 2002]. C'est dans ce cadre que se situent mes travaux sur la modélisation de petits circuits génétiques oscillants.

De nombreux exemples d'oscillations moléculaires

Avec l'amélioration des techniques de marquage fluorescent et d'acquisition vidéo par microscopie, on a pu observer des variations périodiques d'activité génétique dans de nombreux contextes différents. Ainsi, le rythme physiologique bien établi du cycle circadien a pu être associé aux oscillations de la transcription de certains gènes manifestant l'existence d'une horloge génétique. Le cycle d'activité semble fondamentalement être le résultat de l'action négative exercée en retour par une protéine sur l'expression de son propre gène [DUNLAP, 1999; GONZE et collab., 2000]. Un autre cycle fondamental biologique bien connu, le cycle cellulaire est, comme son même le suggère déjà une répétition. Les phases répétitives du cycle de vie de la cellule, croissance et de reproduction sont contrôlées par un réseau transcriptionnel et post-transcriptionnel de signaux moléculaires dont les composants centraux sont les protéines kinases dépendantes de la cycline (cyclin-dependent protein kinases-Cdks) [LI et collab., 2004; TYSON et collab., 2001; TYSON et NOVAK, 2001] qui varient périodiquement. D'autres oscillations génétiques encore ont été mises en évidence dans le contexte très différent du développement embryonnaires des vertébrés. La structure antéro-postérieure de l'axe corporel est ainsi la répétition de structures générées par un processus rythmique qui fait intervenir un circuit génétique oscillant [POURQUIÉ, 2003]. Celui-ci implique les facteurs de transcription Her1 [LEWIS, 2003] et Hes1 [HIRATA et collab., 2002] formant des boucle d'auto-régulation négatives. En résulte une oscillation temporelle qui s'associe avec la segmentation et la formation du patron spatial périodique du développement morphologique.

D'autres exemples particulièrement intéressants dans différentes situations physiologiques et développementales impliquent des facteurs de transcription possédant un comportement oscillant. Deux cas significatifs sont les protéines NFκB [HOFFMANN et collab., 2002; NELSON et collab., 2004] et p53 [LAHAV et collab., 2004; LEV et collab., 2000; ZATORSKY et collab., 2006], impliquées

dans des voies de régulation de réponses cellulaires critiques à la survie en fonction de signaux environnementaux. Dans ces deux cas, les circuits génétiques centrés sur ces gènes sont sensibles à des signaux de stress inter- ou intra-cellulaire d'une variété de situations et peuvent déclencher l'expression de différents gènes en fonction des signaux en entrée.

Les rôles des oscillations de NF κ B et *p53* a été très débattu depuis la découverte du phénomène, et la signification de ce comportement périodique reste très hypothétique [LEV et collab., 2000; NELSON et collab., 2004]. En effet, ces facteurs semblent avoir des conséquences, en termes de choix et niveaux d'activation de leurs multiples cibles, qui dépendent du nombre, période et amplitude des cycles d'oscillations suivant des signaux, suggérant un système de codage ou de dosage via ce comportement temporel. Ainsi, les oscillations peuvent-être considérées comme un phénomène permettant une réponse progressive : après chaque pic d'expression d'une oscillation le système analyserait si la réponse a été suffisante. Dans le cas contraire, des répétitions du cycle se poursuivent tant que le problème n'est pas résolu. Enfin, en cas de signal excessif, une amplitude ou une période trop importante pourrait conduire à un autre type de réponse tel que le déclenchement irréversible de l'apoptose. Des oscillations génétiques permettraient ainsi d'encoder avec un même ensemble de gènes réduit des signaux moléculaires différents associés à une diversité de réponses [NELSON et collab., 2004].

Oscillations génétique et biologie de synthèse

Ces observations dans des systèmes naturels ont motivé dans le cadre de la biologie synthétique des tentatives de réalisation de circuits génétiques oscillants artificiels assemblés dans des bactéries [ELOWITZ et LEIBLER, 2000; FUNG et collab., 2005]. Ces constructions *in vivo* ont été conçues d'abord *in silico* à l'aide de modèles théoriques pour définir les interactions et les conditions donnant lieu à des oscillations. Le circuit oscillant « Repressilator » publié par M. Elowitz et S. Leibler en 2000 a ainsi été un étendard du programme de la biologie synthétique. Il a donné lieu depuis à une série de raffinements et de versions significativement améliorées avec la maîtrise accrue des principes de construction génétique [POTVIN-TROTTIER et collab., 2016; STRICKER et collab., 2008] et de contrôle des organismes synthétisés [MCMILLEN et collab., 2002]. De telles expériences ont permis de proposer des schémas de conception différents [O'BRIEN et collab., 2012; PURCELL et collab., 2010] et une exploration par la réalisation des principes de fonctionnement des oscillations génétiques et de leur modélisation.

Les situations naturelles et synthétiques ont ainsi motivé de nombreuses études théoriques basées sur des approches mathématiques et des hypothèses biologiques différentes pour capturer et représenter comment les différents objets moléculaires et génétiques interagissent et contribuent à la dynamique des oscillation et ses propriétés. Les facteurs fondamentaux à l'origine de ces rythmes sont résumés dans l'encadré 5.1. Ils sont subjacents à toutes les explications théoriques qui reprennent les éléments généraux de modélisation de l'expression génétique déjà présentés.

5.3.2 Oscillations génétiques et modèles à retard

J'ai discuté au chapitre 4, section 4.4.1 des approches de modélisation par systèmes d'équations différentielles. Jusqu'aux années quand j'ai commencé à m'intéresser, les modèles à retard avaient été relativement peu appliqués dans le contexte des réseaux génétiques mais connaissaient un regain d'intérêt. En dépit de l'échelle réduite des circuits dans ces applications, qui n'incluent qu'une poignée de gènes, les non-linéarités des interactions et le retard conduisent à une dynamique complexe. Mon approche avec Basile Grammaticos était ainsi motivée par la recherche de conditions permettant des oscillations stables dans des modèles génétique avec boucles de rétroaction de cette nature. Ceci a donné lieu à trois publications discutées ci-dessous sur :

- Un modèle minimal inspiré des oscillations du facteur de transcription P53 impliqué dans

Échelles de temps microscopiques des réactions et phénomènes biologiques :

- transcription et traduction (s ↔ min)
- liaison, dissociation de complexes moléculaires (ms ↔ h)
- modifications chimiques diverses (phosphorylation/déphosphorylation)
- transport moléculaire (s ↔ h)
- le jour et la nuit (24h)
- les saisons (365j)

Dynamique hors d'équilibre :

- instabilités dynamiques
- bifurcations (changements majeurs de la dynamique en fonction des paramètres)

Encadré 5.1 – Aux origines de la périodicité du vivant : échelles de temps et hors d'équilibre

de multiples fonctions cellulaires importantes dont l'apoptose et qui est un des oncogènes majeurs connus

- Une étude inspirée par le système précédent de systèmes d'oscillations génétiques contrôlés par la dégradation des produits.
- La réduction de modèles à retard stochastiques et déterministes.

5.3.3 Modèle minimal des oscillations génétique du système p53 avec des équations différentielles retardées

La protéine p53 est un facteur de transcription chez les mammifères qui régule l'expression de gènes intervenant dans la réparation de l'ADN, l'arrêt du cycle cellulaire ou l'apoptose en réponse à divers stress génotoxiques et cytotoxiques. Il a un rôle clé dans la prévention de la prolifération de cellules génétiquement instables, comme en témoigne la présence de mutations sur le gène dans le gène du plus de 55% des tumeurs humaines. En raison de son intérêt médical et son implication dans de nombreux processus cellulaires critiques, de très nombreuses recherches biologiques et génétiques ont mis en évidence une architecture de régulation autour du gène p53 particulièrement complexe impliquant de nombreux gènes et interactions. Le système p53 était voici dix ans, au moment de mes travaux, un des systèmes génétiques les plus étudiés [WAHL et collab., 2005]. Il l'est toujours. Plus de 60 000 publications lui font référence dans la base de données bibliographique Pubmed et des conférences internationales y sont régulièrement consacrées [GNANAPRA-DEEPAN et collab., 2018; LEROY et collab., 2013, 2017; LEVINE, 2011]. Ce réseau a également suscité de nombreux travaux théoriques, en particulier depuis la découverte de ses oscillations d'expression et celle des gènes associés.

Dans le travail avec Basile Grammaticos sur le système du gène p53, nous avons effectué une analyse détaillée des oscillations d'un modèle minimal construit à partir d'un nombre très réduit de composants biologiques, les mieux établis. Notre modèle est soigneusement établi à partir d'informations biologiques et est basé techniquement sur des équations différentielles retardées. Peu de temps avant notre travail, en 2004, des observations sur des cellules uniques par l'équipe de Uri Alon [LAHAV et collab., 2004] avaient mis en évidence un comportement inédit : les oscillations du système p53 ont un comportement « numérique » ! La réponse à un stress provoque la répétition de pics d'expression de la protéine d'amplitude constante, selon un nombre de fois proportionnel à l'intensité de l'agression, figure 5.11. L'analyse de bifurcations de notre modèle montre comment avec des ingrédients très simples et un temps de retard le cœur supposé du circuit génétique impliquant p53 est effectivement capable d'une transition par rapport à un signal vers des oscillations soutenues d'amplitude constante.

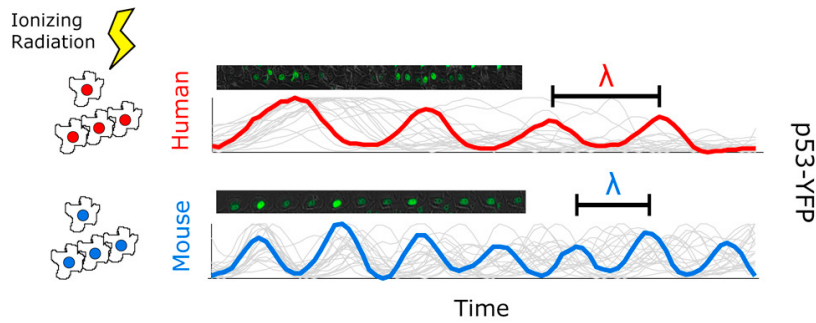


FIGURE 5.11 – **Oscillations pulsées de P53 en réponse à un stress de rayonnement, chez l'humain et chez la souris.** Les expériences récentes de cette figure reproduisent les observations originales [LAHAV et collab., 2004] de la dynamique de p53 dans des cellules individuelles après ce type de stress. La concentration de la protéine de p53 exprime des pics de concentration d'amplitude constante répétés un nombre de fois proportionnel à l'intensité de la radiation. Source : [STEWART-ORNSTEIN et collab., 2017]

5.3.4 Le système génétique p53

Lorsque p53 est activé et est suffisamment exprimée, il induit la transcription des gènes impliqués dans la réparation de l'ADN, l'arrêt du cycle cellulaire et l'apoptose. Dans des conditions normales, la concentration de la protéine p53 est faible en raison d'un boucle négative entre p53 et son principal antagoniste la protéine MDM2. Celle-ci est une ubiquitine ligase qui marque p53 pour faciliter sa dégradation par des protéases. MDM2 est elle-même activée par p53.

Cette boucle est le circuit de régulation central du système, qui reçoit et traite les signaux sur l'intégrité du génome. En réponse à un stress certains paramètres cinétiques des réactions chimiques qui la composent se trouvent altérés. Les travaux biologiques montrent en particulier que la protéine P53 est moins dégradée dans ce cas. Elle s'accumule donc dans la cellule et aboutit à l'activation des mécanismes de réparation moléculaire qui sont sous son contrôle, figure 5.12. Le retour à un niveau stable et faible d'expression de p53 n'a lieu qu'une fois les dommages réparés et le niveau de stress diminué. Si les dommages sont excessifs et de réparation complète insuffisamment rapide, la cellule entre dans une voie vers l'apoptose (suicide cellulaire). En raison de ce rôle critique pour la destinée cellulaire on comprend que le système p53 incorpore de multiples régulations et systèmes de contrôle afin de ne pas commettre de choix irréversible inapproprié. Comme système de sauvegarde majeur des cellules, il n'est pas surprenant non plus de le trouver dysfonctionnel dans de nombreux cancers où il ne peut jouer son rôle protecteur.

L'existence d'oscillations après un stress, plutôt qu'une simple bascule vers un niveau d'expression élevé et stable était déjà connue depuis longtemps [HAUPT et collab., 1997; LEV et collab., 2000]. Ces observations pouvaient s'interpréter classiquement comme les oscillations d'un système atténué vers un nouvel état. La découverte en 2004 [LAHAV et collab., 2004] d'une réponse sous forme d'un nombre discret de répétitions d'un cycle d'activité, proportionnellement au signal en entrée a causé beaucoup de surprise. Il s'agissait d'une forme de réponse dynamique encore inédite dans un système biologique. Ce comportement et la raison de cette forme particulière de réponse génétique étonne encore et son étude est toujours d'actualité [DEMIRKIRAN et collab., 2017; ELIAŠ et collab., 2014; HAFNER et collab., 2017; MÖNKE et collab., 2017; STEWART-ORNSTEIN et collab., 2017; STEWART-ORNSTEIN et LAHAV, 2017].

Une explication possible à cette propriété voit dans la répétition d'un nombre de pulses plus ou moins grand, un moyen de contrôle temporel des mécanismes déterminés par p53. En effet, étant donné les conséquences critiques et potentiellement létales, telle que l'apoptose, ce comportement permettrait une mesure pondérée et contrôlée dans le temps des processus mis en œuvre par p53. La génération d'un temps d'attente par la répétition de pulses donnerait le temps aux processus de réparation d'intervenir et d'ajuster leur continuation, leur arrêt ou d'enclencher un suicide cellulaire.

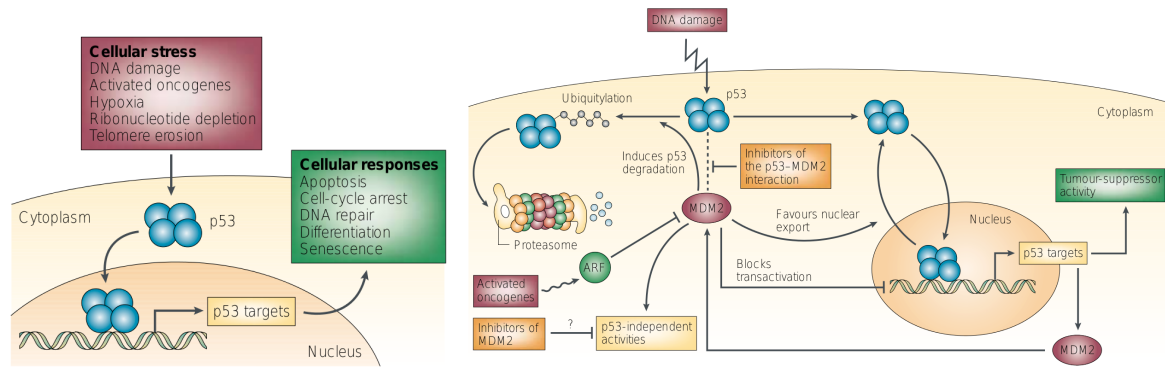


FIGURE 5.12 – **Réponse de p53 à des stress** Gauche : En présence de différentes formes de stress la protéine p53 s’accumule dans la cellule, forme un complexe tétramérique et induit la transcription de différents gènes impliqués dans le contrôle du cycle cellulaire, l’apoptose, la réparation de l’ADN, la différenciation et la sénescence. La perte ou le dysfonctionnement du gène p53 suite à des mutations permet la prolifération des cellules endommagées. Cette prolifération non contrôlée conduit au développement de tumeurs. Droite : le cœur du système p53 est formé par la boucle de régulation avec le gène MDM2. p53 stimule l’expression de MDM2; MDM2 inhibe p53 de différentes manières (blocage de sa transcription, export hors du noyau, dégradation stimulé). Différents signaux cellulaire comme des dommage de l’ADN ou des oncogènes activent la phosphorylation de p53 inhibant son association avec MDM2. Différents signaux peuvent intervenir à différentes étapes de l’interaction entre p53 et MDM2 affectant la dynamique du système. Source : [CHÈNE, 2003]

Plusieurs travaux théoriques ont naturellement suivi la découverte de ce phénomène utilisant différentes hypothèses biologiques et des approches mathématiques différentes [CILIBERTO et collab., 2005; LAHAV et collab., 2004; MA et collab., 2005] discutées en détail dans notre article [BOTTANI et GRAMMATICOS, 2007a].

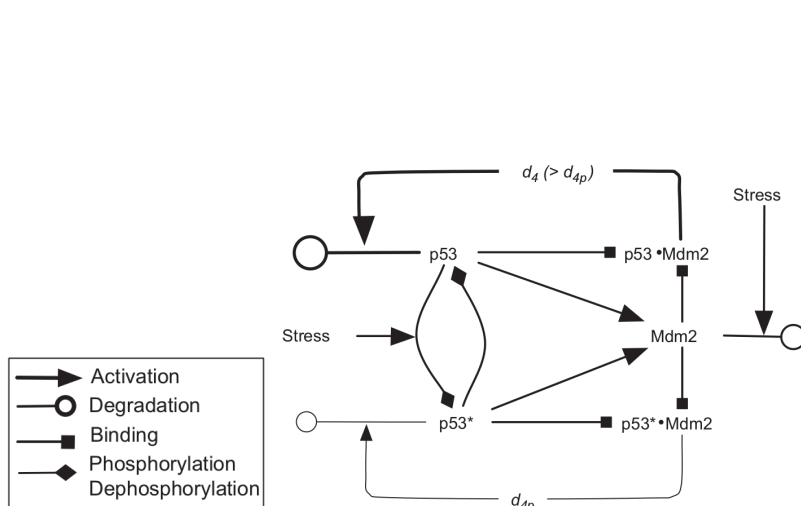


FIGURE 5.13 – **Diagramme du circuit minimal p53-Mdm2.** p53* désigne la protéine phosphorylée, p53•Mdm2 le complexe formé par les deux protéines. d_4 et d_{4p} les coefficients de dégradation. La protéine p53 est exprimée constamment à un niveau basal et peut être phosphorylée ($p53p$) via un processus binaire catalysé par le niveau de stress au-dessus d’un seuil. La phosphorylation de p53 sélectionne un mode avec plus grande dégradation (lignes épaisses) ($d_4 > d_{4p}$) que la dégradation non phosphorylée (lignes fines). Source : [BOTTANI et GRAMMATICOS, 2007b]

Un modèle minimal des oscillations de p53

Notre travail avec Basil Grammaticos s’est inscrit dans dans cette mouvance. En ce qui nous concerne nous nous sommes intéressés à un modèle minimal de la dynamique p53-MDM2 incluant seulement les interactions centrales, les plus établies, entre les deux gènes et leurs produits afin d’identifier les propriétés dynamiques de base sans autre élément. La figure 5.13 présente le

circuit que nous avons retenu, dont il serait trop détaillé de reprendre ici la justification détaillée dans notre article [BOTTANI et GRAMMATICOS, 2007b]. Chaque lien peut ensuite se formaliser en termes de réactions chimiques suivant la démarche discutée dans le chapitre 4, à partir desquelles on construit les équations différentielles du modèle. Ici nous avons aussi introduit un terme de retard pour rendre compte de la production de la protéine MDM2 induite par p53 qui la régule sous forme de tétramère. Après d'autres simplifications d'écriture le cœur du modèle devient le système réduit suivant :

$$\begin{aligned}\frac{dx}{dt} &= a - xs - bx - cxz \\ \frac{dy}{dt} &= xs - by - d yz \\ \frac{dz}{dt} &= h + fF(\bar{x} + \bar{y}) - g(s)z\end{aligned}\tag{5.2}$$

$$\tag{5.3}$$

La signification des symboles est la suivante. La protéine p53 existe en deux états, déphosphorylé x et phosphorylé y (la déphosphorylation est très lente on peut la négliger). La variable booléenne s correspond à la sortie d'un module de transduction de signal en présence de stress qui indique la présence de dommages sur l'ADN, $s = 1$ représente pour nous la présence de stress. Une simplification qui se justifie au niveau auquel on se place en considérant que les signaux en sortie de voies de transduction du signal sont typiquement connus pour être fortement non linéaires, donc s'apparenter à des interrupteurs. La production de la protéine Mdm2 z se produit avec un retard τ après l'initiation de la transcription induite par p53. La fonction F est définie par $F(u) = u^v / (K_t^v + u^v)$ avec $v = 4$. La barre dans l'argument de F indique qu'il est calculé à l'instant $(t - \tau)$, c'est-à-dire avec un retard dans le temps. Les paramètres c et d correspondent à la dégradation induite de 53 (x) et p53* (y) et expriment les différentes dégradations de ces deux formes suite à leur liaison avec MDM2, avec $d < c$. Le taux de dégradation de z est exprimé comme une fonction binaire $g(s) = (1 - s)g_1 + sg_2$ pour tenir compte de la dégradation accrue de MDM2 durant le stress.

Le système 5.2 est maintenant écrit sous une forme mathématique suffisamment simple pour permettre une analyse de stabilité conjointement à sa simulation numérique. La figure 5.14 montre le comportement temporel par intégration numérique des équations différentielles avec avec un algorithme Runge-Kutta du 4^e-ordre et une liste chaînée pour la gestion des retards. Elle montre pour des jeux de paramètres motivés biologiquement, la compatibilité de deux régimes dynamiques avec stress et sans stress du passage d'un état stable de faible expression à des oscillations persistantes toute la durée du stress. La figure 5.15 montre le diagramme de bifurcation en fonction de la durée de demie-vie de MDM2. En absence de production basale le passage entre un régime stable et instable de l'absence de stress au stress impose une contrainte forte sur les demies vie de MDM2 dans ces deux régimes (cas a) et b). En présence d'un niveau basal de production de la protéine, le diagramme de bifurcation fait apparaître un décalage plus important de la zone d'instabilité entre les deux états levant les contraintes sur la durée de vie pour le comportement de transition attendu.

Avec l'écriture simplifiée du système sous la forme 5.2 il est aussi possible de mener une analyse de stabilité semi-analytique. Cette analyse permet d'explorer plus rigoureusement l'espace des paramètres en déterminant l'existence de points fixe de stabilité et les comportements des variables en leur voisinage sans recours aux simulations numériques. Elle permet de caractériser les régimes de comportement dynamiques qualitativement possibles pour le système et les valeurs des paramètres quand ces changements se produisent (bifurcations). Nous avons effectué l'analyse de stabilité séparément pour $s = 0$, absence de stress et en présence de stress $s = 1$. La démarche de l'analyse de stabilité est classique, y compris pour un système à retard, et correspond à la recherche de valeurs propres du système linéarisé à proximité de chaque point fixe. Son

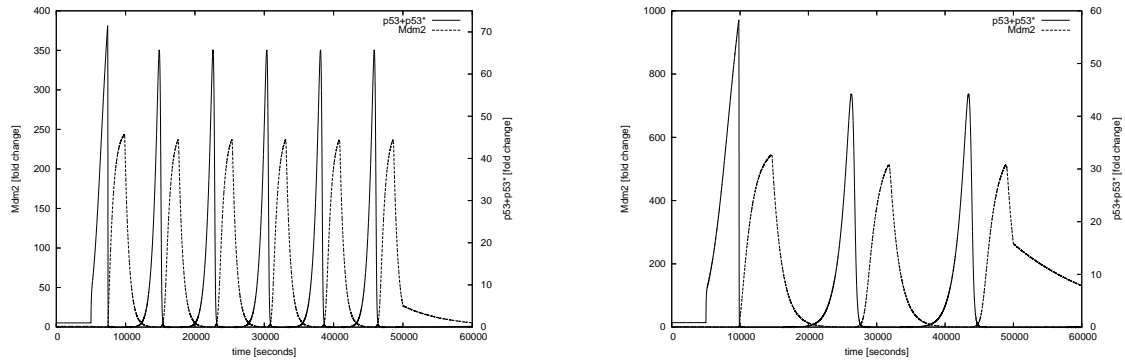


FIGURE 5.14 – **Simulations du modèle minimal de l'oscillateur central p53-Mdm2.** Simulation en absence d'expression basale de Mdm2 ($h = 0$), $g = 0.0016(0.1(1 - s) + s)$ (la demie vie de MDM2 est $\tau_m = \log(2)/g$), $c = 0.1$, pour des retards (a) $\tau = 2400$ s et (b) $\tau = 4800$ s. Les conditions initiales retardées sont initialisées à la valeur stationnaire pour $s = 0$. Le système démarra à l'état stationnaire sans stress ($s = 0$); à l'instant $t = 5000$ s on induit le stress $s = 0 \rightarrow s = 1$; à l'instant $t = 50000$ s le signal de stress s'éteint $s = 1 \rightarrow s = 0$. Suivant l'analyse détaillée des paramètres dans l'article les valeurs prises sont dans les domaines $a = 1$, $b = 1.10^{-4}$, $c \in [0.001, 1]$, $d \sim c/10$, $h \in [0, 0.1]$, $f = 1, K_t = 28$. On observe des oscillations durant le stress avec un retour au point fixe stable après le stress. Source : [BOTTANI et GRAMMATICOS, 2007b]

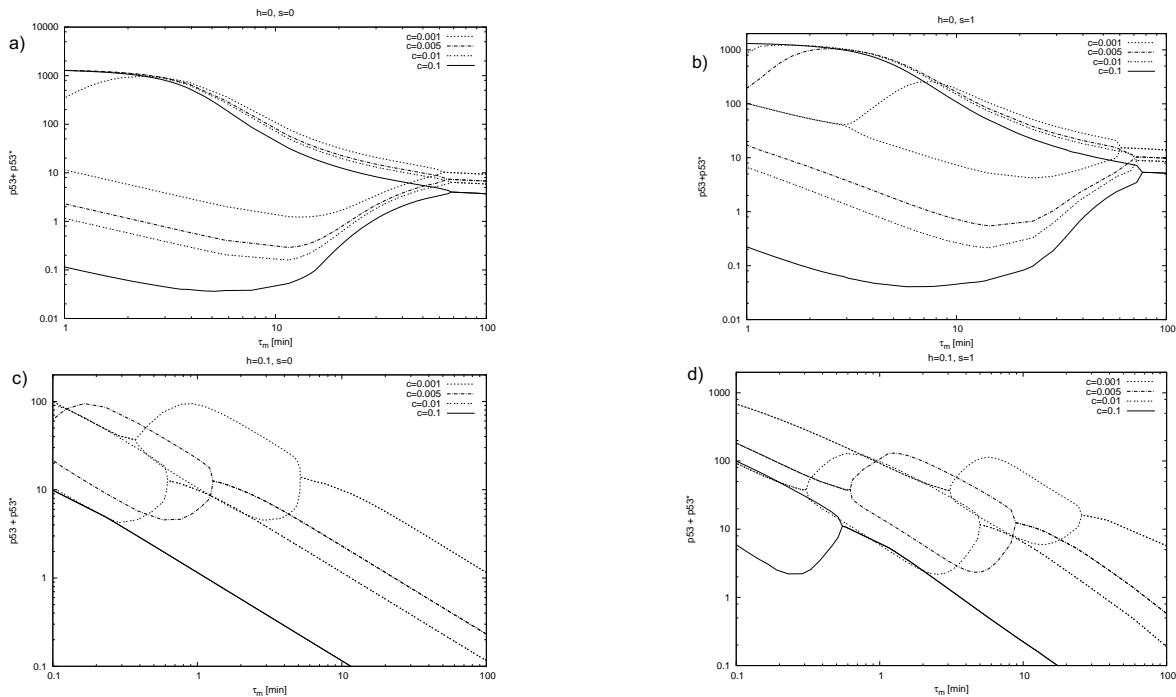


FIGURE 5.15 – **Diagramme de bifurcation en fonction de la demie-vie de MDM2.** Bifurcations en fonction de τ_m pour différentes valeurs c et $d = c/10$ de la dégradation induite de p53, en absence d'expression basale de MDM2 (a et b) avec une expression basale de MDM2 ($h = 0, 1$) (c et d). (a) et (c) en absence de stress $s = 0$. (b) et (d) avec stress $s = 1$. Le retard est $\tau = 2400$ s. Pour une valeur donnée τ_m une valeur unique indique la stabilité et deux valeurs indiquent une instabilité correspondant aux extréma des oscillations persistantes. Ici un point fixe stable existe en absence de stress que pour $\tau_m > \sim 60$ s. Le domaine d'oscillation en présence de stress impose une dégradation sous stress $\tau_m < \sim 8$ s. Source : [BOTTANI et GRAMMATICOS, 2007b]

application pour le système considéré ici demande en raison des non-linéarités importantes la résolution numérique d'équations transcendantales de valeurs propres.

Cette analyse montre clairement que dans les deux régimes distincts de stress et d'absence de celui-ci la dynamique change de comportement entre stabilité et oscillations comme attendu et que l'instabilité (oscillations) dépend particulièrement de la demie-vie de MDM2 entre les deux situations. L'analyse de stabilité permet de déterminer rapidement, sans intégration numérique des équations la sensibilité aux différents paramètres. Ainsi le taux de production basal de p53 apparaît comme un autre paramètre pouvant entraîner des changements sensibles de comportement.

En conclusion le modèle montre des régions d'instabilité avec des oscillations pour un large éventail de paramètres et de transitions continues vers la stabilité. Ces régions sont robustes et compatibles avec la survenue d'oscillations en fonction de l'état physiologique sous réserve d'hypothèses raisonnables sur le signal de déclenchement. Le circuit entre MDM2 et p53 au cœur du système est capable intrinsèquement à générer un comportement périodique dans des conditions assez larges permettant un basculement entre un état stable et des oscillations sous le contrôle du stress. Dans ce modèle la boucle de rétroaction entre p53 et MDM2 sert deux objectifs : elle assure la stabilité du point fixe à un niveau d'expression de p53 faible en absence de stress et elle permet la génération d'oscillations de l'expression de p53 (et MDM2) suite à des dommages.

Cette étude a fourni un aperçu utile de la dynamique d'un des modèles les plus simples possibles du système de p53. Même dans ce cas minimal les dynamiques sont suffisamment riches pour réussir à décrire le comportement observé. À partir de ce travail nous avons envisagé des développements du modèle et notamment une version stochastique.

Régulation par induction de protéolyse

Le type de régulation que nous venons de discuter dans le cas de p53 semble être très général. Dans notre article « A simple model of genetic oscillations through regulated degradation » [BOT-TANI et GRAMMATICOS, 2008] nous avons en effet relevé de nombreux exemples de circuits génétiques impliquant une boucle de rétroaction entre un facteur régulateur et une cible qui déclenche spécifiquement la dégradation de ce régulateur suivant le schéma de la figure 5.16 dans une variété de situations biologiques in a variety of biological situations [HUNG et SHAPIRO, 2002; JENAL et ARONIS, 2003; KURAS et collab., 2002; LAUB et collab., 2002; LIBEREK et GEORGOPOULOS, 1993; PRUTEANU et ARONIS, 2002; ROUILLON et collab., 2000; STRAUS et collab., 1990]. Dans ces exemples les protéines induisant la protéolyse sont induites tridimensionnellement par les facteurs de transcriptions qu'elles contribuent à dégrader.

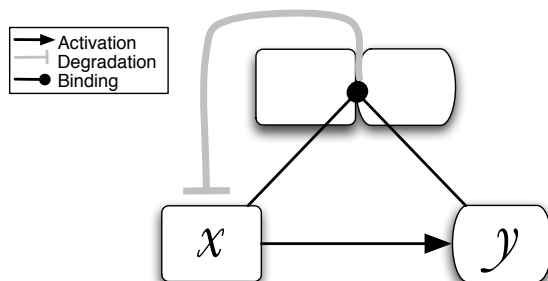


FIGURE 5.16 – Diagramme du circuit de régulation.

Les boîtes indiquent les produits des gènes. La protéine P53, produit du gène p53 est un facteur de transcription qui agit sur le gène Mdm2 (flèche). Les boîtes indiquent les produits des gènes. En se liant à MDM2 la protéine P53 subit une dégradation induite. Les lignes noires forment un motif de réseau génétique de type II.2 (voir section 5.2.3).

La dégradation de protéines comme dans le système p53 et ces autres exemples semble être une solution énergétiquement pas très favorable pour maintenir un niveau basal faible de protéines régulatrices. La régulation par la protéolyse pourrait cependant être un avantage par rapport à un contrôle transcriptionnel dans des situations nécessitant une réponse temporelle précise et une réactivité rapide sur des facteurs de transcriptions importants. En complément, par le biais de modifications enzymatiques (par exemple phosphorylation) la protéolyse peut être rapidement

inhibée amenant à une augmentation de la concentration cellulaire du régulateur beaucoup plus immédiate par rapport à une activation de sa production transcriptionnelle. Ce n'est donc pas très surprenant de rencontrer ce type de régulation dans des systèmes pour des situations d'urgence comme la réponse à un stress. Enfin, la dégradation contrôlée est aussi une nécessité si la cellule doit éliminer périodiquement certaines protéines comme à chaque période d'oscillations génétiques. Dans ce cas une interaction de ce type comme dans les modèles que nous avons étudié contribue à la génération d'oscillations fortes.

L'analyse de stabilité linéaire des modèles permet d'identifier des critères mathématiques des comportements et d'identifier les domaines de stabilité et, en cas d'instabilité les périodes des cycles limites. Ainsi il apparaît qu'un retard dans la boucle de rétroaction négative est insuffisant pour produire des oscillations. Celles-ci ne sont possibles qu'en présence de plus de deux équations de bilan de masse pour décrire la boucle, ou bien dans des modèles simplifiés avec deux équations mais suffisamment non-linéaires ($n > 2$) dans les fonctions de Hill qui résultent de la réduction du nombre d'équations par des approximations quasi-statiques. Pour les modèles qui admettent un régime instable, les oscillations ont lieu uniquement pour des délais au-dessus d'un seuil.

5.3.5 Comparaison de modélisation déterministes et stochastiques et de schémas de réduction

L'article « Stochastic and deterministic simulations of a delayed genetic oscillation model : Investigating the validity of reductions » [BOTTANI et GRAMMATICOS, 2012] présente des résultats méthodologiques sur la modélisation et la simulation de circuits oscillants génétiques basé sur le réseau formel minimal pour oscillations génétiques persistantes discuté plus haut. Nous avons étendu les approches précédentes dans un cadre stochastique numérique Monte-Carlo ce qui a nécessité le développement d'une version retardée dans le temps de l'algorithme classique de Gillespie pour les réactions chimiques stochastiques. À cette occasion nous avons étudié la mise en œuvre d'approximations quasi-statiques dans le cadre des simulations stochastiques.

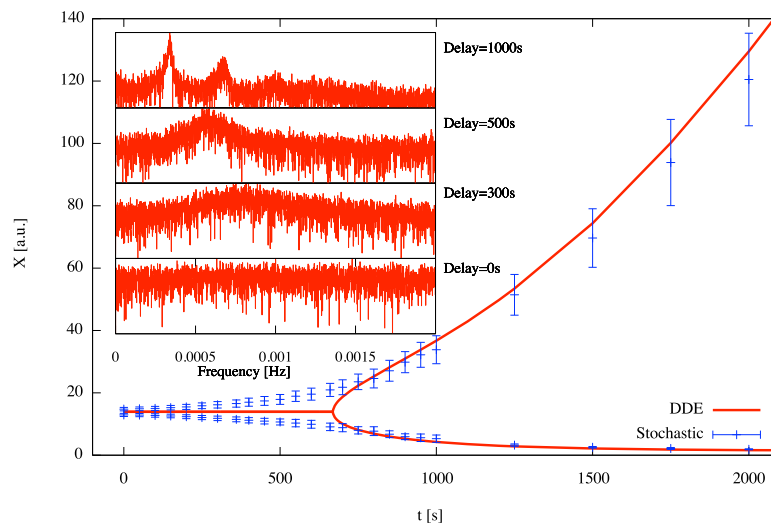


FIGURE 5.17 – **Diagramme de bifurcation stochastique et déterministe du circuit p53-Mmd2.** Diagramme de bifurcation du modèle complet en fonction du temps de retard (sans approximations quasi-statiques) pour un ensemble de paramètres caractéristique. Une valeur unique pour un retard donné indique une valeur stationnaire pour X; deux différentes valeurs indiquent les minimum et maximum d'oscillations stables. Pour le modèle complet stochastique les barres d'erreurs représentent une déviation standard autour de la moyenne calculées sur une longue exécution. Insert : spectre de puissance d'une simulation stochastique unique pour différents retards.

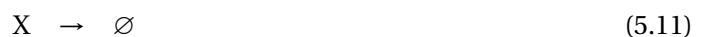
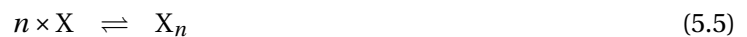
Réactions chimiques stochastiques et algorithme de Gillespie En complément des modélisations déterministes il est en effet important d'étudier les propriétés stochastiques de circuits de régulation en raison des fluctuations venant de petits nombres possibles des molécules participantes [MCADAMS et ARKIN, 1997]. Comme discuté dans la section 4.4.2 sur la modélisation stochastique du chapitre 4 ces systèmes se décrivent microscopiquement à l'aide de l'équation maîtresse généralement impossible à résoudre directement. En pratique les solutions d'un système stochastique sont généralement obtenues à l'aide de méthodes Monte-Carlo par l'algorithme de Gillespie [GILLESPIE, 1977] et ses variantes.

De telles simulations Monte-Carlo sont notoirement coûteuses numériquement. Leur application à des réseaux biochimiques devient très exigeante quand le nombre d'espèces en réaction augmente et comporte des réactions avec des échelles de temps très différentes. Des réactions de modification chimique, des interactions de liaison/dissociation de protéines avec l'ADN, sont généralement plus rapides que les différents processus génétiques comme la transcription, traduction et dégradation. Une simulation Monte-Carlo d'un circuit de régulation génétique même aussi réduit que ceux des figures 5.13 et 5.16 passera l'essentiel de son temps à traiter les événements rapides, les plus probables par unité de temps.

De nombreuses améliorations algorithmiques et méthodes d'accélération ont été proposées pour accélérer les simulations Monte-Carlo de ces systèmes de régulation [GIBSON et BRUCK, 2000; PAHLE, 2009; TURNER, 2004]. Une autre manière de procéder est de rechercher des simplifications dans la modélisation des systèmes en le réduisant aux interactions essentielles et en condensant de nombreuses réactions microscopiques dans une seule réaction effective préservant les propriétés macroscopiques. Cette démarche est commune déjà dans la simplification des systèmes déterministes avec les approximations quasi-statiques qui conduisent par exemple aux équations de Michaelis-Menten et aux fonctions de Hill pour représenter des réactions rapides.

Nous avons voulu dans notre étude explorer cette idée dans le contexte des simulations stochastiques avec l'algorithme de Gillespie et capturer l'effet de réactions rapides sur la probabilité d'initiation de « réactions » de transcription. Une autre simplification que nous avons introduite et qui reprend la même idée de l'approche équivalente déterministe, est l'inclusion de retards, pour capturer de manière effective l'effet reporté dans le temps d'un événement sans le détail de processus intermédiaires. Notre étude combine donc des approximations quasi-statiques dans l'algorithme de Gillespie avec des réactions retardées.

Système de réactions et équations déterministes retardées Le système de réactions qui représente le circuit général de la figure 5.16 que nous considérons comme cas d'étude s'écrit :



Ces réactions représentent : l'expression basale du gène x et la production de la protéine X (5.4); la formation d'un multimère d'ordre n agissant comme facteur de transcription(5.5); la liaison de ce complexe sur la région cis-régulatrice d'un gène y (5.6); la transcription et la traduction du gène y capturés dans une seule pseudo-réaction avec la libération du produit Y après un

retard τ (5.7); une transcription basale de y (5.8); la formation d'un complexe des protéines X et Y (5.10); la dégradation induite de X et Y (5.10); la dégradation intrinsèque des protéines (5.11-5.12). Ce modèle minimal néglige tous les détails des processus de transcription et traduction. Pour l'étude numérique de notre article nous considérons le cas non linéaire le plus petit avec $n = 2$, le facteur de transcription agissant sous forme de dimère.

La mise en équation déterministe du système des réactions (5.4)-(5.12) suit la procédure habituelle de la cinétique chimique discutée en 4.3 et amène directement pour $n = 2$ aux équations différentielles retardées :

$$\frac{dX}{dt} = k_1 + 2k_2X_2 + (k_3 + k_4)Z - k_4Z - k_5X - 2k_6X^2 - k_7XY \quad (5.13)$$

$$\frac{dX_2}{dt} = k_6X^2 + k_8y^* - k_2X_2 - k_9yX_2 \quad (5.14)$$

$$\frac{dy^*}{dt} = k_9yX_2 - k_8y^* \quad (5.15)$$

$$\frac{dY}{dt} = k_{10} + k_{11}\tilde{y}^* + (k_3 + k_4)Z - k_{12}Y - k_7XY \quad (5.16)$$

$$\frac{dZ}{dt} = k_7XY - (k_3 + k_4)Z \quad (5.17)$$

où X, Y et Z sont les concentrations de protéines, y et y^* respectivement la cible de la région cis-régulatrice du gène non lié et lié (par conservation $y + y^* = 1$). Le tilde dans l'équation (5.16) indique qu'il s'agit des concentrations à l'instant précédent $t - \tau$ pour prendre en compte le retard de transcription et traduction suivant le choix déjà discuté de simplification. Même pour un circuit aussi simple le nombre de paramètres est trop grand pour une analyse générale. Je renvoie à la publication [BOTTANI et GRAMMATICOS, 2012] pour la discussion étendue sur la paramétrisation et les choix biologiquement significatifs. On mentionnera ici seulement deux paramètres définis sur des intervalles et liés à la question des réductions qui nous intéressera : respectivement la constante de dissociation du dimère $K_{X_2} = k_2/k_6 = [9 - 900] \text{ s}^{-1}$ [ACKERS et collab., 1982; SAIZ et VILAR, 2006] et la constante de dissociation du dimère sur le promoteur $K_{y^*} = k_8/k_9 = [2 - 200] \text{ s}^{-1}$ [SAIZ et VILAR, 2006], qui sont rassemblés pratiquement dans un vecteur $\vec{K} = (K_{X_2}, K_{y^*})$. Il est important de noter que ces constantes de dissociations capturent de nombreuses situations de paramètres microscopiques différentes, tels que les rapports k_2/k_6 et respectivement k_8/k_9 sont les mêmes. À défaut d'informations, on prendra les taux (microscopiques) d'association identiques $k_6 = k_9 = k_+$ avec un paramètre k_+ pour le taux commun qu'on a exploré dans un intervalle large de variation biologique.

Comportements stochastique versus déterministe L'implémentation stochastique de ce système avec l'algorithme de Gillespie prend chacune de ces réactions associées à une probabilité par unité de temps et par molécule pour leur réalisation. Le diagramme de bifurcation de la figure 5.17 montre que pour un retard suffisant, le modèle déterministe ainsi que le modèle stochastique présentent une transition entre un état stationnaire stable et des oscillations entretenues. La transition est moins nette dans le cas stochastique : comme illustré sur la figure 5.19 pour des faibles retards l'expression des protéines montre des répétitions de pics quasiment périodiques, intermittentes, sur de courts intervalles mais avec spectre de puissance s'aplatissant sur une durée d'acquisition longue. Les oscillations stochastiques dans des régimes où le modèle déterministe n'oscille pas de manière entretenue sont surtout marquées dans des limites de taux d'associations petites avec des petits nombres ou des événements rares qui donne lieu à des événements pulsatiles.

Les oscillations stochastiques sont cependant sujettes à des variations de phase et amplitude de telle manière que la moyenne des simulations individuelle ne donne pas une image correcte

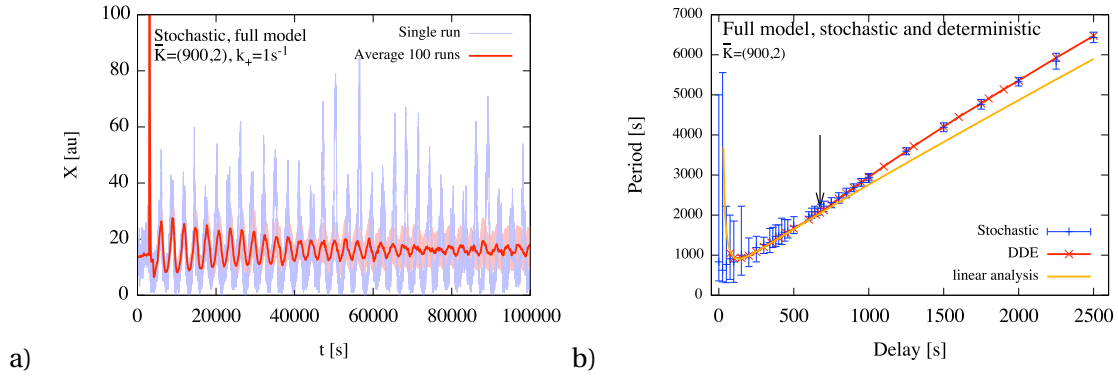


FIGURE 5.18 – **Évolution stochastique du modèle complet du circuit p53-mdm2.** Modèle avec $\bar{K} = (900, 2)$, $k_7 = 10 \text{ s}^{-1}$ et $k_+ = 1 \text{ s}^{-1}$ et retard $\tau = 1000 \text{ s}$. a) Évolution de la variable X pour une simulation après une perturbation du système initialisé à l'état stationnaire : ligne bleue, exemple d'une simulation unique; ligne rouge, moyenne de 100 simulations différentes; région grisée : une déviation standard autour de la moyenne pour 100 simulations. b) Périodes des modèles déterministes et stochastiques en fonction du retard, ainsi que la prédiction par l'analyse linéaire de stabilité

de la dynamique stochastique comme le montre l'exemple de la figure 5.18a. Par contre comme le montre la sous figure b) les périodes des oscillations stochastiques et déterministes (atténuées ou non) correspondent très bien, et sont également assez proches de l'estimation des périodes issues de l'analyse linéaire de stabilité du modèle.

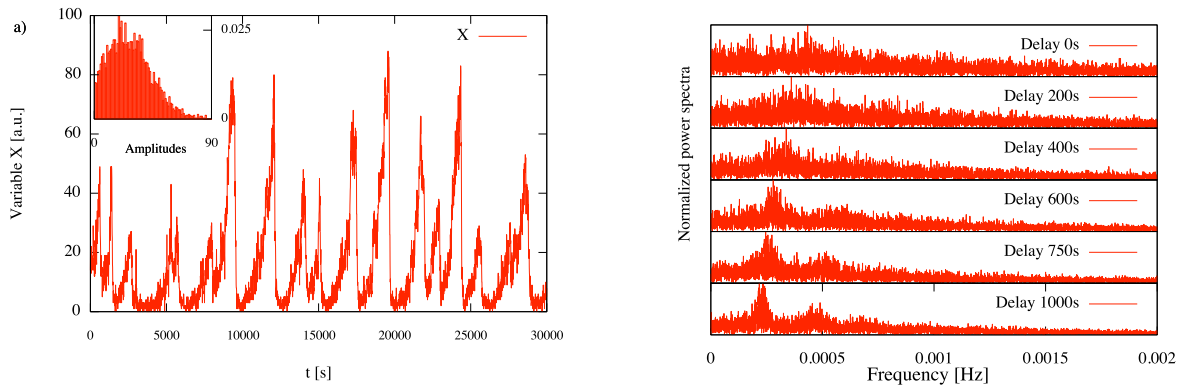


FIGURE 5.19 – **Modèle stochastique complet en absence de retard, avec des taux d'association faibles.** a) exemple sur une simulation unique Monte-Carlo, protéine X. Insert : distribution normalisée des amplitudes sur une durée d'acquisition de $4 \cdot 10^6 \text{ s}$. b) Spectre de puissance pour X pour des retards croissants 0, 200, 400, 600, 750, 1000, 1500 s et les mêmes paramètres que en a).

Réductions des modèles par approximations quasi-statiques En supposant les réactions de transcription et traduction comme lentes par rapport aux autres, différentes approximations quasi-statiques peuvent être envisagées sur ce système de réactions qui sont discutées en détail dans l'annexe de notre article [BOTTANI et GRAMMATICOS, 2012] :

1. R1 : approximation quasi-statique de la liaison du facteur de transcription sur le promoteur ($X_2 + y \rightleftharpoons y^*$)
2. R2 : approximation quasi-statique de la formation du dimère ($X + X \rightleftharpoons X_2$)
3. R3 : approximation quasi-statique de la formation du complexe des produits ($X + Y \rightleftharpoons Z$)
4. R4 : approximation quasi-statique de la formation du dimère ET de la liaison du dimère sur le promoteur (le deux premiers cas précédents ensemble)

5. R5 : Toutes les approximations précédentes ensemble.

Les implémentations déterministes et stochastiques de ces différentes réductions sont détaillées dans l'annexe de l'article [BOTTANI et GRAMMATICOS, 2012]. Notons que les réductions R4 et R5 qui combinent la dimérisation et la liaison sur le promoteur font intervenir une fonction de Hill avec une constante de dissociation $K_{MM}^2 = K_{X_2}K_{y^*}$ unique pour les différentes combinaisons de paramètres rassemblés dans \bar{K} et qui sont choisis comme cas d'étude dans la suite. Nous étudions les approximations ci-dessus à la fois dans le cadre déterministe que stochastique. La figure 5.20 montre l'effet des différentes réductions comparées avec le modèle complet pour la modélisation sous forme d'équations différentielles avec retard simulés numériquement. Les différentes réductions quasi-stationnaires ne sont pas équivalentes si on compare les trajectoires exactes même si qualitativement les comportements le sont. Toutes les réductions ne sont pas équivalentes, les réductions R1 (liaison de X_2 avec y quasi-statique), et R2 (dimerization quasi-statique), sont les plus proches du modèle complet. La réduction R4 (combinaison de R1 et R2), qui est la forme de réduction la plus commune dans les modélisations biochimiques de réseaux génétiques, est aussi proche du modèle complet, mais dans l'exemple simulé moins précise sur les amplitudes.

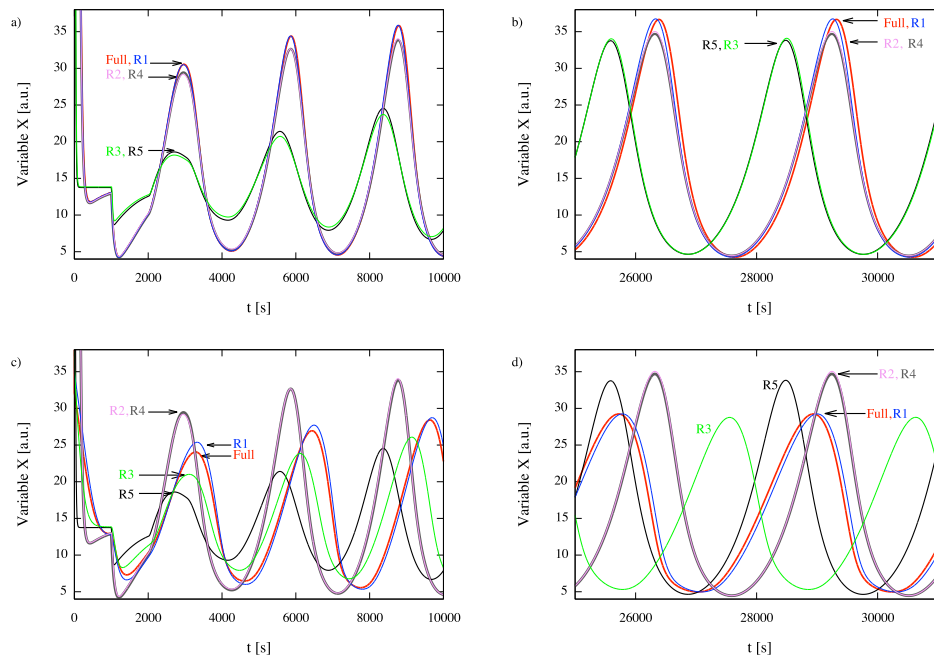


FIGURE 5.20 – **Modélisation déterministe par les équations différentielles avec retard.** Courbes pour la variable X pour des jeux de paramètres caractéristiques. Le système est initialisé à l'état stationnaire sur une durée avant l'instant initial égal au retard et avant une perturbation à l'instant $t = 0$ s. a) Comportement transitoire après la perturbation. Les courbes pour les différentes réductions sont proches les unes des autres ou se superposent quasiment. Ceci est le cas notamment pour le modèle complet et R1 ; R2 et R4, R3 et R5. b) La même simulation plus tard dans le temps après stabilisation. Les courbes superposées dans le transitoire le sont encore. c) et d) comme les deux sous figures précédentes pour un autre jeu de paramètres des constantes de dissociation. Certaines courbes se superposent comme dans le cas précédent (modèle complet, R1) et (R2,R4) mais plus (R3, R5).

Dans le contexte stochastique les réductions sont plus délicates à évaluer en raison de la variabilité et des fluctuations des amplitudes, fréquences et phases des oscillations. Les figures 5.21 et 5.22 présentent ainsi séparément respectivement les distributions des amplitudes des oscillations et les fonctions d'autocorrélation qui rendent compte des périodes et de phases.

Les réductions R1 et R3 se comportent comme le modèle stochastique complet en produisant des oscillations pulsatiles aussi pour des retards faibles alors que les versions déterministes

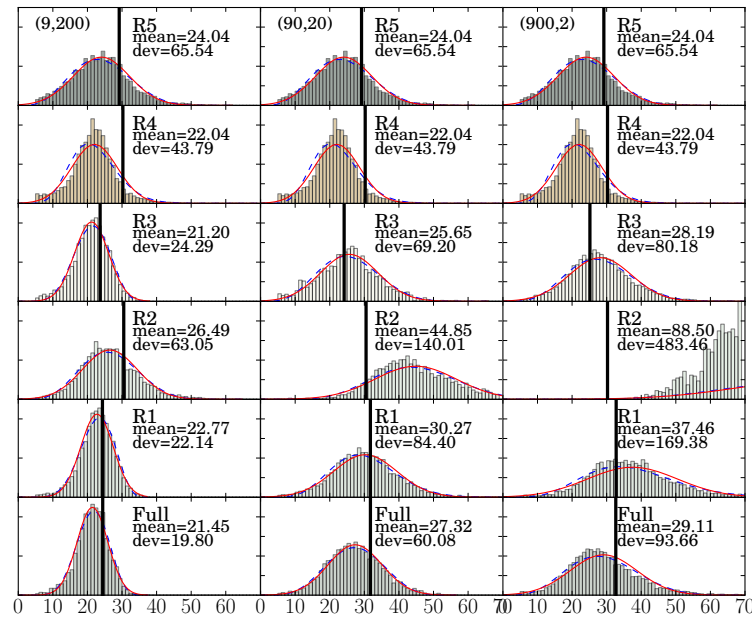


FIGURE 5.21 – **Distribution des amplitudes des oscillations stochastiques pour le modèle stochastique complet et toutes les réductions.** Les colonnes correspondent de gauche à droite aux copules de constantes de dissociation $\bar{K} = (9, 200), (90, 20), (900, 2)$. Chaque ligne est un type de réduction différent; du haut au bas R5, R4 (identiques pour les combinaisons \bar{K} considérées), R3, R2, R1 et le modèle complet. Les lignes verticales noires indiquent les valeurs des amplitudes pour les modèles déterministes retardés correspondants. Taux d'association : pour la formation du complexe Z $k_7 = 10 \text{ s}^{-1}$, pour la dimerisation et la liaison au promoteur (en fonction de la réduction) $k_6 = 2k_+, k_9 = k_+, k_+ = 1 \text{ s}^{-1}$. Les courbes rouges sont les meilleurs ajustement pour une distribution normale et les courbes bleus pour une distribution beta.

correspondantes sont stables. Ce n'est pas le cas des réductions R2, R4 et R5 qui n'ont pas d'oscillations stochastiques en dessous d'un seuil de retard. Étant donné que ces réductions capturent la dimérisation et la liaison sur le promoteur, on peut dès lors supposer que l'origine du comportement pulsatile est lié à la réaction de liaison sur le promoteur. Une réaction impliquant des petits nombres, une ou deux copies du promoteur, donc sujette à de fortes fluctuations. Les réductions R1 et R3 préservent explicitement cette réaction et ses fluctuations explicites, ce qui leur permet de mieux rendre compte du comportement stochastique dans ce régime.

Les distributions des amplitudes stochastiques, figure 5.21, sont généralement centrées à une déviation standard des valeurs du modèle déterministe correspondant. Les réductions R1 et R3 sont les plus proches du modèle complet pour les jeux de paramètres considérés, en accord aussi avec les réductions déterministes (figure 5.20). La réduction R4, usuelle dans les modélisations, donne des différences par rapport au modèle complet aussi bien dans sa version déterministe que stochastique. Les cas traités (y compris ceux non montrés ici) n'excluent cependant pas l'existence de domaines de paramètres avec un meilleur accord. Les mêmes conclusions valent pour la réduction la plus importante R5. Les fonctions d'autocorrélation des différentes réductions de la figure 5.22 indiquent la périodicité des oscillations déterministes et stochastiques et la décroissance dans le temps de celles-ci. Les réductions ne préservent pas la période exacte même si les différences restent faibles. On constate également une évolution des autocorrélations significativement différentes suivant les schémas de réduction.

Conclusions modèles stochastiques et déterministes et réductions Cette étude de nature méthodologique sur la modélisation de circuits génétiques et surtout de schémas de réduction a mis

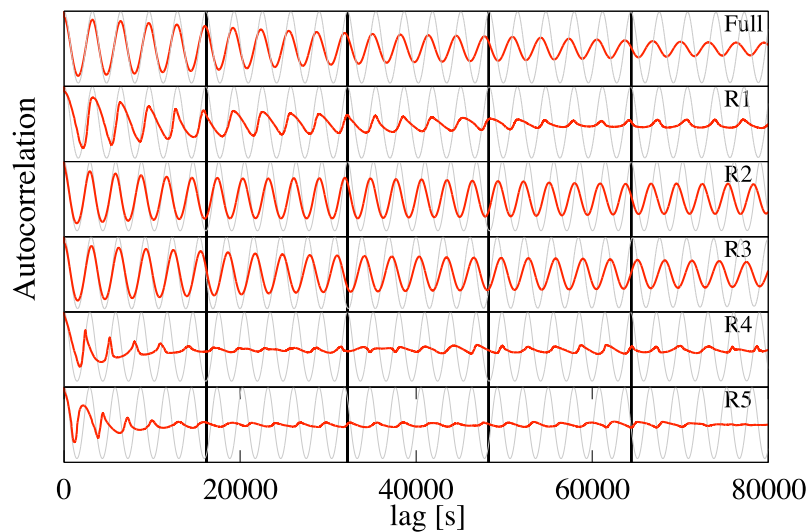


FIGURE 5.22 – **Fonctions d'autocorrélation.** Pour le modèle complet et les différentes réductions après un transitoire de 20000 s suivant une perturbation. Les valeurs des paramètres d'association sont $k_+ = 1\text{s}^{-1}$, $k_7 = 10\text{s}^{-1}$ et les constantes d'association pour la situation représentée ici $\bar{K} = (9, 200)$. Les autres combinaisons $\bar{K} = (90, 20)$, $(900, 2)$ considérées dans la figure 5.21 donnent des comportements similaires (courbes non montrées). Les courbes rouges représentent les 10^5 premiers décalages de la fonction d'autocorrélation de la variable X d'une simulation stochastique unique de longueur $> 10^6$. Les courbes grises sont les fonctions d'autocorrélation pour les modèles correspondants déterministes à retard de chaque réduction. Les lignes verticales noires sont des guides en phase avec les oscillations du modèle déterministe complet.

en évidence des variations significatives de comportement en fonction de choix spécifique de simplifications des modèles. De tels modèles et réductions ont également un intérêt dans d'autres champs d'application comme en écologie et épidémiologie.

Le modèle « complet » considéré défini par les pseudo-réactions 5.4 inclut clairement déjà de nombreuses simplification et hypothèses sur la situation qui a servi d'inspiration et inhérentes de la démarche de modélisation. Il a défini cependant dans cette étude la référence « microscopique » à partir de laquelle nous avons construits différentes simplifications ultérieures. Émettre des hypothèses de quasi-stationnarité est une démarche commune pour simplifier des modèles d'équations différentielles. Les systèmes réduits qui en résultent font intervenir moins de variables et d'équations que le système original et font généralement apparaître des termes non-linéaires déterminants pour la dynamique et qui lèvent la rigidité souvent trouvée dans les équations originales biochimiques d'action-de masse [CHOI et collab., 2008].

Ici nous avons étudié différentes combinaisons d'hypothèses de quasi-stationnarité pour des jeux de réactions pouvant être considérés rapides par rapport aux réactions de transcription et traduction dans le circuit dans les deux contextes du système déterministe retardé et du système stochastique retardé.

Le modèle stochastique produit des oscillations bruitées qui fluctuent en amplitude et période. Dans le régime instable, la moyenne d'un ensemble statistique de simulations stochastique ne converge pas vers la courbe oscillante déterministe et montre plutôt des oscillations atténuées, même si chaque simulation individuelle montre des oscillations entretenues.

Pour ces comportements non-linéaires stochastiques on ne peut attendre pour la moyenne sur la population d'être représentative du comportement caractéristique et des variations d'une courbe individuelle.

Quoique habituelles dans les modélisations déterministes l'application de réductions par des

approximations quasi-stationnaires sont moins communes et ont été moins étudiées dans le cadre stochastique, voir la discussion dans [GÓMEZ-URIBE et collab., 2008] et la conclusion de notre article [BOTTANI et GRAMMATICOS, 2012]. Cette étude visait à explorer sur l'exemple d'un circuit de petite taille. Malgré la manière assez empirique d'implémenter certaines réductions dans le cadre stochastique, on obtient des comportements proches du modèle complet au moins qualitativement. Les réductions combinant plusieurs réactions (comme nos réductions R4 et R5) qui correspondent aux hypothèses les plus communes dans le cadre déterministe, conduisent cependant à des approximations moins robustes et de domaine de validité par rapport au modèle microscopique plus restreint. Les fluctuations et les corrélations stochastiques ne sont pas préservées systématiquement dans les réductions. En conséquence l'usage des réductions s'il demeure possible, et fournit assez facilement des comportements qualitativement du même ordre qu'un modèle complet, microscopique, soit être bien examiné au cas par cas en cas de besoin de valeurs quantitatives plus précises. De la même manière, il est important de garder bien présent que les modèles déterministes ne représentent pas nécessairement la variation moyenne des systèmes stochastiques correspondants, et que leur pertinence doit être évaluée suivant les applications.

L'étude que nous allons maintenant voir concerne un autre aspect qui était inexploré sur des modèles classiques de régulation génétiques impliquant une simplification quasi-stationnaire des fonctions de Hill.

5.4 Caractérisation de la réponse non linéaire de la transcription.

Malgré le grand nombre d'études, tout n'est pas encore connu sur la fonction de réponse transcriptionnelle de promoteurs! Dans notre article « Hill function-based models of transcriptional switches : Impact of specific, nonspecific, functional and nonfunctional binding » [BOTTANI et VEITIA, 2016] nous nous sommes intéressés aux effets du nombre de cibles d'un facteur régulation sur la forme de fonction de réponse d'un promoteur qu'il régule. Une question en relation par le sujet du dosage génétique qui m'occupe avec Reiner Veitia sur laquelle je reviendrai plus loin 5.5.

La fonction de régulation, notons là $P([A])$, décrit le taux de production d'ARN messenger produit à partir d'un promoteur P en fonction de la concentration d'un facteur de transcription activateur A ¹. Celle-ci est communément décrite par une fonction de Hill avec deux paramètres, la constante de dissociation K et l'exposant non linéaire n , :

$$P([A]) \sim \frac{[A]^n}{K^n + [A]^n}. \quad (5.18)$$

Cette relation figure dans les manuels et peut se justifier facilement, voir l'encadré 1.2. La réponse d'un gène en fonction de l'abondance de ses facteurs de régulation est une des relations les plus fondamentales du comportement génétique. Le système de contrôle d'un gène est souvent décrit comme un interrupteur moléculaire dans lequel la présence ou non d'une molécule « de régulation » active ou non l'expression du gène cible. Des concentrations de facteur de transcription en dessous d'un seuil n'induisent pas de transcription, au-dessus la réponse est significative. L'existence d'un tel seuil permet notamment de discriminer un signal du bruit et de permettre des choix.

On sait que le niveau d'expression de la protéine produit du gène peut faire l'objet d'une succession de multiples niveaux de contrôle depuis l'initiation de la transcription de la séquence codante de l'ADN, la maturation éventuelle de l'ARN messenger avec de l'épissage pour les organismes supérieurs, la transcription des ces ARN messagers en protéines, la maturation de celles-ci, enfin leur activation éventuelle par liaison avec des métabolites ou d'autres protéines. Le point de départ et première étape principale reste l'activation de la transcription par la liaison, ou non, des

1. Pour simplifier on parlera dans le texte uniquement de la situation de facteurs de transcriptions activateurs. Les conclusions sont identiques et le traitement équivalent pour des inhibiteurs, les relations étant complémentaires.

protéines facteur de transcription sur la région dite régulatrice, ou encore promotrice, de l'ADN en amont de la séquence codante du gène, voir exemple sur la figure 5.23. Les protéines facteurs de transcription ont la propriété de pouvoir se lier à l'ADN plus particulièrement sur des sections d'une séquence spécifique de reconnaissance. Suivant la présence d'un ou plusieurs facteurs de transcriptions liés sur leurs séquences spécifiques dans la zone de régulation la polymérase, va pouvoir s'y assembler, ou être recrutée déjà composée, se lier à promoteur marquant le début de la séquence codante du gène en tant que machinerie moléculaire procédant à la transcription [GREENBLATT, 1997; LEHMAN et collab., 1998,?; REINBERG et ROEDER, 1987; ROEDER, 1996], voir figure 5.23.

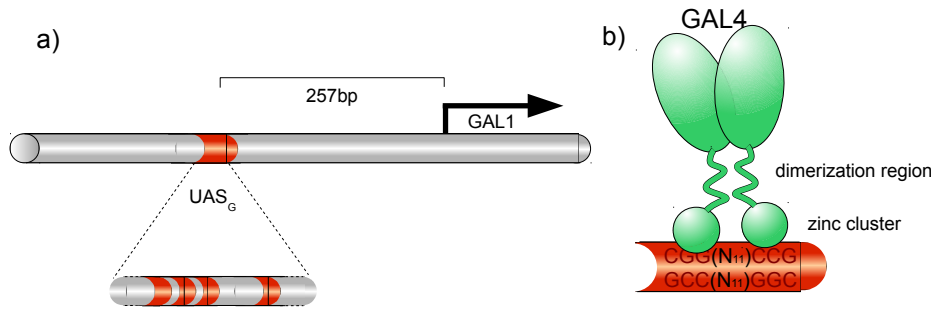


FIGURE 5.23 – **Régulation du gène GAL4.** a) Le gène GAL1 et la région de régulation GAL4 en amont. Les 4 sections colorées dans la région de régulation en amont correspondent à 4 séquences de liaison spécifiques pour le facteur de transcription GAL4. La zone de régulation s'étend sur 118 bases d'ADN et chaque séquence spécifique GAL4 est de 17 bases. b) Le facteur de transcription est un dimère de la protéine GAL4 est capable de se lier à une séquence palindromique de 17 bases d'ADN. (Le schéma ne montre que la section de GAL4 se liant à l'ADN). Figure d'après [PTASHNE et GANN, 2002].

Cette description implique une notion de condition combinatoire pour la transcription quand la bonne combinaison de facteurs est présente sur l'ADN. Un mécanisme qui s'interprète tout à fait comme un système d'interrupteur moléculaire. Cependant cette vision doit être complétée car elle ne suffit pas pour décrire l'état moyen d'occupation favorable de la zone de régulation par rapport aux abondances des facteurs de transcription dans la cellule ou le noyau.

En règle générale de multiples facteurs interagissant simultanément pour recruter la polymérase. Cette association a des conséquences très importantes car il s'agit du processus biologique qui génère la non-linéarité mathématique de l'activation d'un gène en fonction de l'abondance de ses régulateurs. Cette propriété est généralement insuffisamment reconnue dans les manuels et articles de référence et associée confusément avec la notion de coopérativité des facteurs de transcription. Comme discuté dans ma publication avec Reiner Veitia [BOTTANI et VEITIA, 2016] la non-linéarité de la réponse transcriptionnelle peut venir de deux mécanismes d'interaction reliés mais distincts : la *coopérativité* proprement dite d'une part et la *synergie* d'autre part.

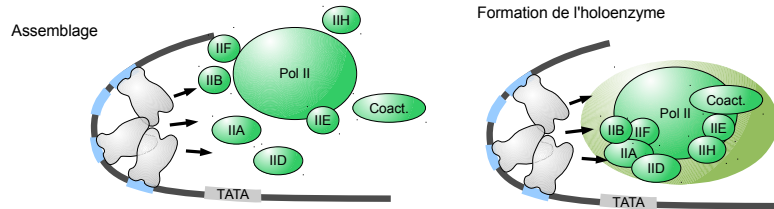
La coopérativité signifie un renforcement des interactions lorsque plusieurs unités agissent ensemble : l'interaction globale est d'intensité supérieure à la somme des interactions individuelles. Un processus qui intervient par exemple lorsqu'un premier événement facilite les suivants. C'est le cas quand la liaison d'un premier facteur de transcription TF accroît la probabilité de liaison d'un facteur suivant et ainsi de suite. Une propriété qui semble générale, mise en évidence par exemple sur une étude génomique sur *Drosophila megalogaster* [KAZEMIAN et collab., 2013]. La coopérativité implique une interaction protéine-protéine des facteurs agissant simultanément sur une partie tierce.

La synergie est un processus dans lequel plusieurs facteurs agissent ensemble, mais indépendamment, sans interaction directe entre-eux. L'action conjointe est simplement et seulement la somme des actions individuelles. Les protéines agissant ensemble sur une partie tierce n'interagissent pas entre elles. Les différentes situations de liaison sont représentées sur la figure 5.25, où sont distingués des cas supplémentaires lorsque les facteurs multiples en action sont identiques

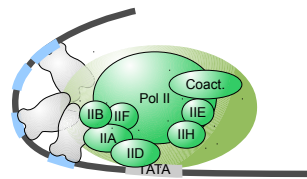
Étape 1: liaison du facteur de transcription



Étape 2: Recrutement de la machinerie de transcription



Étape 3: Liaison sur le promoteur



Étape 4: Initiation de la transcription

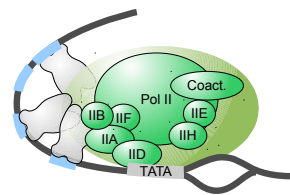


FIGURE 5.24 – **Schéma de la régulation de la transcription.** 1) En amont de la séquence codante du gène se trouvent un ou plusieurs sites de liaison de protéines, les « facteurs de transcription » avec la faculté de se lier spécifiquement aux séquences particulières qui s’y trouvent. 2) La présence d’une combinaison appropriée de facteurs de transcription permettra l’assemblage du complexe de transcription à polymérase, ou bien son recrutement comme complexe déjà formé l’« holoenzyme ». 3) La polymérase pourra se lier au démarrage de la séquence codante du gène, sur le « promoteur ». 4) La polymérase liée au promoteur pourra initier la transcription.

(situations « homo- ») et lorsque les facteurs sont différents (situations « hétéro- »).

La différence entre synergie et coopérativité peut bien se visualiser au niveau énergétique. Notons ΔG_1^0 et ΔG_2^0 les énergies libres de liaison indépendantes de chacun de facteurs 1 et 2 avec l’holoenzyme de transcription (polymérase). Alors l’énergie libre du complexe formé par l’holoenzyme et les deux facteurs liées simultanément sera la somme des énergies interactions individuelles $\Delta G_1^0 + \Delta G_2^0$ plus une contribution éventuelle (positive ou négative) d’interaction entre les facteurs² En cas de synergie les interactions des facteurs sont indépendante et ce terme est nul. En cas de coopérativité ce terme favorise la liaison des deux facteurs ensemble avec la polymérase et cette contribution est non nulle et négative (plus l’énergie libre de liaison est négative, plus la liaison est « forte »).

On peut lier les énergies de liaison à la constante d’équilibre (inverse de la constante de dissociation) \mathcal{K} par la relation classique $\Delta G^0 = -RT \log \mathcal{K}$, T la température en Kelvin, $R = 8.314 \text{ J.mol}^{-1}.\text{K}^{-1}$ la constante de Boltzman R. Ainsi la constante d’équilibre est simplement le produit des constantes individuelles en cas de synergie $\mathcal{K} = \mathcal{K}_1 \mathcal{K}_2$, alors que la coopérativité amène une constante augmentée du terme d’interactions $\mathcal{K}_{coop} = \mathcal{K}_1 \mathcal{K}_2 \mathcal{K}_{interaction}$. Dans les deux cas la liaison double conduit à une constante d’association bien plus importante que celle des liaisons individuelles, ce qui est la raison conduisant à la non linéarité de la fonction de réponse.

La modélisation par les approches classiques (voir encadré 1.2) amène à la fonction de Hill pour la courbe de réponse moyenne avec un exposant n de la non linéarité correspondant au nombre de sites de liaison dans la région cis-régulatrice. Plus grande est la synergie (ainsi que la coopérativité) plus la réponse est non-linéaire avec une sigmoïde prononcée (figure 5.26). La co-

2. Pour être complet il faudrait inclure dans les termes énergétiques d’éventuelles contributions dues à la déformation nécessaire des protéines ou de l’ADN pour permettre l’interaction. Ces termes contribuent de fait dans le terme de coopérativité, positivement ou négativement. La synergie qui implique des interactions indépendantes est donc probablement un cas limite.

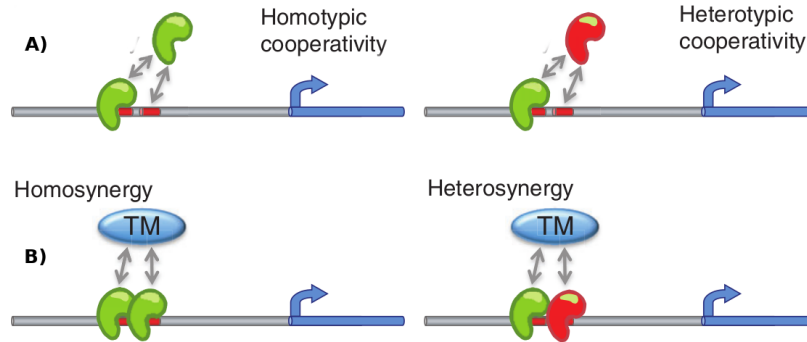


FIGURE 5.25 – **Interactions entre un promoteur et des protéines facteurs de transcription coopératives ou synergétiques.** Les interactions homotype ou hétérotype impliquent respectivement des interactions de molécules identiques ou différentes. A : exemple d'interaction coopérative. La liaison sur l'ADN d'un second facteur est le résultat de l'addition de son interaction individuelle sur son site *plus* son interaction avec le facteur déjà lié. B : exemple d'interaction synergétique. La liaison de la machinerie de transcription est la somme des interactions indépendantes avec chaque facteur. Source : d'après [BOTTANI et VEITIA, 2016]

opération entre facteurs de transcriptions conduit ainsi à une fonction de réponse présentant un seuil défini de telle manière qu'un petit changement de la concentration des protéines de régulation évoque un changement dramatique de la transcription.

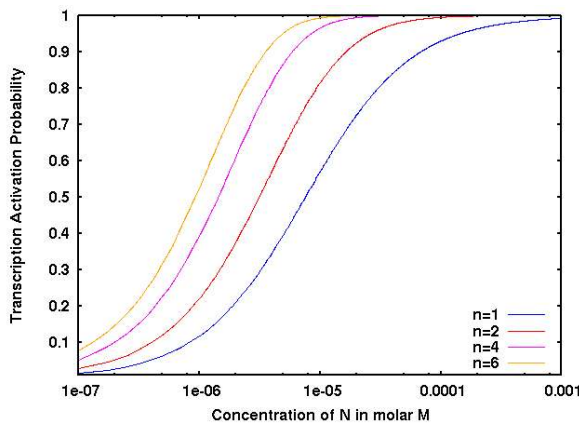


FIGURE 5.26 – **Courbe de réponse transcriptionnelle sigmoïde.** Courbe pour un modèle de promoteur avec n site de liaison pour une protéine activatrice, calculée avec un modèle thermodynamique comme dans [BINTU et collab., 2005a,b]. Le nombre de sites de liaison accroît la raideur de la réponse et décale le seuil en concentration.

Étant donné la simplicité de la fonction de Hill et ses paramètres clairement interprétables, elle est très souvent utilisée en faisant une confusion sur la nature de sa variable $[A]$ méprise par erreur, ou par simplicité, comme la concentration *totale* $[A_{tot}]$. Cette confusion est probablement aussi induite par la facilité à déterminer $[A_{tot}]$ par exemple par la fluorescence d'un rapporteur comme la GFP (Green Fluorescent Protein) fusionné au facteur de transcription [GREGOR et collab., 2007] alors que faire la part entre facteurs liés et libres est beaucoup plus difficile. Que ce soit pour l'expérimentateur contrôlant un régulateur, ou la cellule dans sa physiologie c'est la quantité *totale* de la protéine amenée dans le système ou produite qui est contrôlée. Il serait souhaitable alors d'exprimer la réponse transcriptionnelle comme une fonction $g([A_{tot}])$, mais celle-ci ne peut s'obtenir que comme une solution d'une équation non-linéaire solvable numériquement.

Si le nombre de sites de liaisons pour la protéine est petit et son expression élevée, alors ces nombres sont à peu près identiques et on ne s'attend pas à des différences entre les courbes de réponses exprimées en fonction d'abondances libres ou totales. Si par contre le nombre de cibles possibles pour A est élevé, les abondances libres et totales ne sont pas comparables et comme on l'a mis en évidence ceci affecte la courbe de réponse. Les conséquences de cette différence vont apparaître lorsqu'on considère les multiples sites de liaison d'un facteur de transcription dans le génome.

On considère une situation homo-synergétique dans laquelle une zone de cis-régulation notée P (promoteur) contient n sites identiques d'un activateur A. Si chaque site lié contribue au recrutement de la machinerie de transcription comme dans le cas synergétique de la figure 5.25B) alors, d'après les conséquences multiplicatives des facteurs d'association de la synergie, seul le promoteur pleinement lié PA_n recrutera effectivement la machinerie de transcription et contrôlera celle-ci.

La réponse transcriptionnelle sera proportionnelle à la fraction d'occupation du promoteur $\frac{[PA_n]}{[P]_{\text{tot}}}$ en vertu du principe ergodique discuté à la section 4.4.3 du chapitre 4. On supposera ici toutes les liaisons de i facteurs ($i=1 \dots n$) au promoteur à l'équilibre quasi-statique par rapport à la transcription (voir la même section). De telle manière que tout le processus de régulation se ramène à la réaction $nA + P \leftrightarrow PA_n$.

À l'équilibre thermodynamique la fraction de la concentration du produit PA_n sera [ACKERS et collab., 1982; BINTU et collab., 2005a,b; BUCHLER et collab., 2003; SHEA et ACKERS, 1985] :

$$\frac{[PA_n]}{[P]_{\text{tot}}} = \frac{[A]^n e^{-n\Delta G^0}}{RT} \frac{1}{1 + [A]^n e^{-n\Delta G^0} RT} = \frac{[A]^n}{\mathcal{K}_P^n + [A]^n}$$

où $\mathcal{K}_P = e^{n\Delta G^0}$ est la constante de dissociation pour chaque A individuel. On retrouve la fonction de Hill bien connue.

Notons que l'on peut aboutir très facilement aussi à la même relation avec le formalisme de la cinétique chimique, partant de la condition d'équilibre $\frac{[P][A]^n}{[PA_n]}$ et de la conservation $[P]_{\text{tot}} = [P] + [PA_n]$ et les mêmes hypothèses de synergie et quasi-stationarité.

Encadré 5.2 – Dérivation de la fonction de Hill pour la fonction de réponse transcriptionnelle

5.4.1 Réponse transcriptionnelle avec compétition

Les protéines facteurs de transcription interagissent de multiples manières avec l'ADN génomique avec de nombreux sites de liaison de nature différentes. En premier lieu, de manière générale un facteur possède plusieurs gènes cibles (voir plus haut section 5.2.2 pour *S. cerevisiae*), pour lesquels les zones de régulation peuvent posséder plusieurs sites de liaison homologues spécifiques. Outre ces liaisons de régulation directe, des sites de liaisons homologues spécifiques supplémentaires existent aussi dans les régions non fonctionnelles du génome sans être associées à des promoteurs et des séquences codantes. Enfin les protéines de régulation s'associent aussi de manière non spécifique, plus faible que la liaison homologue, avec l'ADN quel-qu'il soit, donc tout le long du génome.

Pour étudier les effets possibles de telles interactions multiples nous avons construit un modèle avec compétition. Pour simplifier ce modèle inclut seulement deux catégories de liaisons spécifiques, celles sur le promoteur P du gène d'intérêt dont on veut représenter la fonction de transcription, et celles d'autres sites spécifiques R qu'on associera à des liaisons non fonctionnelles. On ne considère pas les multiples liaisons non spécifiques, de liaison plus faible, on peut extrapoler une généralisation des conclusions. En supposant n facteurs se liant au promoteur R et m aux sites non fonctionnels le système correspondra aux réactions :



Réactions auxquelles il faut en outre rajouter la conservation du nombre de sites P (= 2 typiquement) et du nombre de sites R. Ce système peut facilement être mis en équation selon la cinétique chimique (de manière équivalente avec le formalisme thermodynamique aussi) et donner numériquement les relations des occupations fractionnées de P et R dont deux exemples sont dans la figure 5.27. Si le nombre de sites de liaison est petit par rapport à l'abondance des protéines

régulatrices, alors $[A]$ et $[A_{\text{tot}}]$ sont presque égaux et les courbes de réponse TR en fonction de $[A]$ et $[A_{\text{tot}}]$ se superposent presque. Ceci n'est plus le cas quand la quantité de A devient limitant par rapport au nombre de sites de liaison. L'effet de la compétition est d'accroître la non linéarité de la réponse dont la forme fonctionnelle correspond encore à une fonction de Hill, ainsi qu'à décaler la constante de demie occupation vers de plus grandes valeurs.

Ces effets se retrouvent en étendant la compétition en considérant également un compétition entre plus de promoteurs P ($n > 2$) correspondant à des situations biologiques discutées dans notre article [BOTTANI et VEITIA, 2016] et qu'il serait trop détaillé de reprendre ici.

Possibilité de compensation par un contrôle externe, spatial ou temporel Dans des systèmes réels, la concentration $[A_{\text{tot}}]$ d'un régulateur peut varier, soit dans le temps, soit dans l'espace. (C'est le cas bien connu par exemple de la variation exponentielle du facteur de transcription Bicoid induisant la réponse sigmoïde dans l'espace du gène Hunchback dans *Drosophila melanogaster*). La fonction de transcription est alors fonction de la position dans un gradient de régulateur $\text{TR}([A_{\text{tot}}](x))$ (ou bien du temps dans le cas d'une variation temporelle de A). En tenant compte de cette dépendance en présence de compétition on peut montrer qu'il est possible de compenser les effets de compétition par un changement du gradient externe. La présence de non-linéarité conduit en fait à la possibilité de compenser la compétition par un changement très modeste du gradient (ou, selon le cas, de la variation temporelle de la concentration de facteur de transcription). Ceci signifie que deux organismes différents avec des nombres différents de sites de liaison en compétition peuvent montrer des courbes de réponses transcriptionnelles similaires.

Effet stationnaire Les traitements précédents considéraient une situation à l'équilibre avec des quantités données constantes des protéines. De nombreux facteurs sont en fait produits et dégradés en permanence ce qui nécessite des modèles à l'état stationnaires. Le seul changement au modèle précédent est d'inclure des réactions de production et de dégradation à des taux s ($\text{M}\cdot\text{s}^{-1}$) et d (s^{-1}) :



qui associé aux réactions de liaison aux sites en compétition P et R



conduit au système :

$$\frac{d[A]}{dt} = \alpha - \gamma[A] - nk_+^P[A]^n[P] - mk_+^R[A]^m[R] + nk_-^P - [\text{PAN}] + mk_-^R[\text{RAM}] \quad (5.24)$$

$$\frac{d[\text{PAN}]}{dt} = nk_+^P[A]^n[P] - nk_-^P[\text{PAN}] \quad (5.25)$$

$$\frac{d[\text{RAM}]}{dt} = mk_+^R[R]^n[P] - mk_-^R[\text{RAM}] \quad (5.26)$$

$$[P] + [\text{PAN}] = [P]_0 \quad (5.27)$$

$$[R] + [\text{RAM}] = [R]_0 \quad (5.28)$$

A l'état stationnaire ce système conduit aux solutions :

$$[PA_n]_{\text{stat}} = [P]_0 \frac{\left(\frac{s}{d}\right)^n}{K^P + \left(\frac{s}{d}\right)^n} \quad (5.29)$$

$$[RA_m]_{\text{stat}} = [R]_0 \frac{\left(\frac{s}{d}\right)^m}{K^R + \left(\frac{s}{d}\right)^m} \quad (5.30)$$

qui correspondent à des fonctions de Hill avec des constantes de dissociation microscopiques $K^P = k_-^P/k_+^P$ et $K^R = k_-^R/k_+^R$. La solution stationnaire est une fonction de type Hill, où le rapport s/d replace les concentrations de facteur de transcription. Contrairement à la situation à l'équilibre ce résultat est donc indépendant de la présence de compétition.

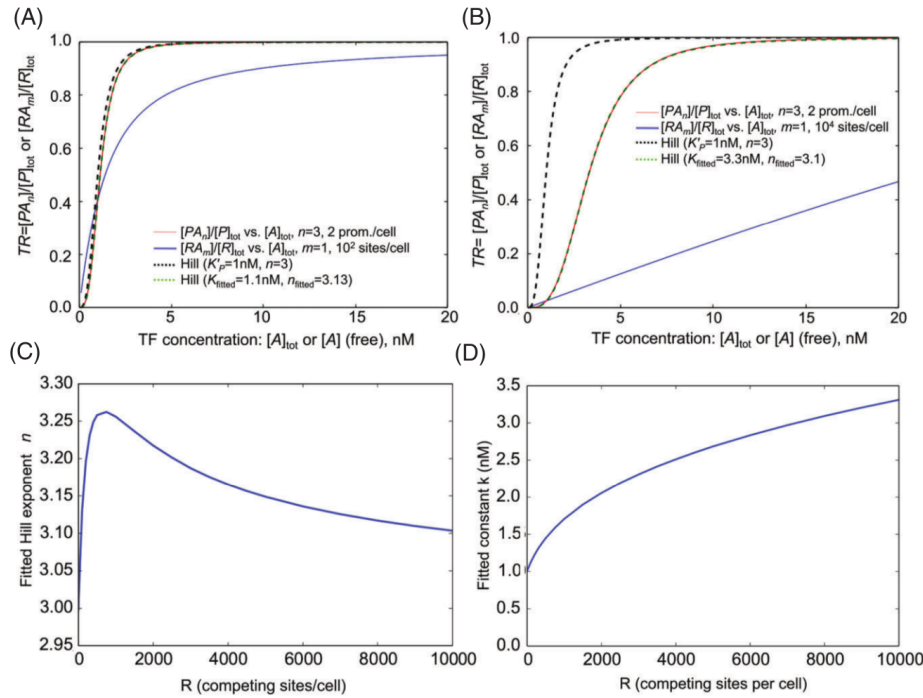


FIGURE 5.27 – **Fraction d'occupation d'un promoteur.** Fraction d'occupation en fonction de la concentration totale d'activateur $[A_{\text{tot}}]$ pour le promoteur P (en rouge, correspondant à la réponse transcriptionnelle) et les sites en compétition R (en bleu). Le promoteur principal d'intérêt P est présent en deux copies, possède $n = 3$ sites de liaison et une constante de dissociation microscopique pour chaque site $\mathcal{K}'_{\mathcal{P}} = 1$ nM. Les sites de liaison en compétition R peuvent lier chacun une molécule A ($m = 1$) avec la même valeur de la constante de dissociation $\mathcal{K}'_{\mathcal{R}} = 1$ nM et sont présents en 100 copies (sous figure A) et 10000 copies (sous figure B); le volume du noyau est supposé 10^{-1} l. La ligne pointillée verte est le meilleur ajustement de la fraction d'occupation du promoteur en fonction de A_{tot} (la courbe rouge) avec une fonction de Hill. Pour comparaison la fonction de Hill avec $\mathcal{K}'_{\mathcal{P}} = 1$ nM et $n = 3$ est représentée en pointillées noires. La fonction de transcription $\text{TR}(A_{\text{tot}}) = \frac{[P]}{[P_{\text{tot}}]}(A_{\text{tot}})$ s'approxime très bien par une fonction de Hill, *mais* avec des paramètres K et n qui ne s'interprètent pas classiquement. Les valeurs ajustées de ces paramètres sont représentées en C) et D) en fonction du nombre des sites de liaisons R en compétition. Les effets de ces sites de liaisons ont été obtenus en ajustant la courbe du promoteur lié P avec des fonctions de Hill en présence de nombre croissant de sites R. Source : [BOTTANI et VEITIA, 2016].

Conclusion effets de compétition L'analyse de cette étude de la fonction transcriptionnelle en présence de sites de liaison en compétition montre que l'utilisation commune de la concentration de facteurs libres $[A]$ comme argument de la fonction de Hill est justifiée quand la concentration effective des régulateurs est beaucoup plus grande que le nombre de promoteurs. Quand ce n'est pas le cas, la fonction de réponse reste une fonction sigmoïde de la concentration totale $[A_{\text{tot}}]$ mais avec des paramètres qui n'ont pas la même signification que dans l'interprétation classique en fonction des facteurs libres $[A]$.

L'influence des sites en compétition de notre modèle prédit qu'avec une augmentation de leur nombre, la réponse transcriptionnelle $TR = g([A]_{tot})$ des promoteurs cibles diminue en non-linéarité et en raideur. L'évolution naturelle, où bien artificiellement la bio-ingénierie, peuvent ainsi régler la raideur d'une fonction de transcription en ajustant le nombre de sites de liaison en compétition.

Nous avons aussi appliqué ce modèle à l'analyse de la sensibilité de la réponse transcriptionnelle à la perte d'un allèle codant pour un facteur de transcription pouvant conduire à de l'haploinsuffisance. Dans ce cas la non-linéarité venant de la synergie dans le modèle de compétition amplifie l'effet possible d'une telle perte si elle n'est pas compensée par d'autres mécanismes. L'haploinsuffisance et d'autres effets de nombre de copies de gènes est le thème central de ma collaboration avec Reiner Veitia que je développerai maintenant dans les parties suivantes.

5.5 Dosage génétique, aneuploidie et déséquilibres génomiques

Cette section regroupe plusieurs contributions durant les dix dernières années avec mon collègue généticien Reiner Veitia de l'Institut Jacques Monod et se rapportant à l'étude d'effets du nombre de copies de gènes sur la physiologie. La question des effets du dosage de gènes est une question de génétique de nature évidemment quantitative et systémique car dépendant de propriétés globales d'interactions dans le génome. Ces articles sont donc de nature principalement génétique, récapitulant de nombreux résultats biologiques d'effets moléculaires et phénomènes génomiques pour lesquels Reiner Veitia est un expert reconnu. L'originalité de ces travaux est l'inclusion de modèles mathématiques illustrant les comportements attendus de mécanismes observés ou attendus liés à l'expression génétique et les liaisons biochimiques des protéines dans les cellules. Les modélisations avec lesquelles j'ai contribué dans ces études sont généralement techniquement simples, et reposant sur les principes discutés précédemment (voir section 4.3). L'originalité de ces travaux de modélisation consiste ici davantage dans la conceptualisation des processus biologiques et leur formalisation. Sauf pour quelques exemples rapportés dans la suite plutôt à titre indicatif (par exemple dans la section 5.5.2) je ne reproduirais pas ici les détails de la modélisation mais discuterai davantage la problématique et des conclusions des modèles.

Les gènes sont en plusieurs copies. On entend par dosage, le nombre de copies d'un même gène, c'est-à-dire du nombre de copies d'une séquence codant pour une protéine, dans le génome d'organismes. Les génomes des animaux, insectes et plantes ont généralement deux ou davantage de copies de leurs chromosomes. En conséquence deux fois, ou davantage, de copies des mêmes gènes. Ces copies peuvent-être identiques, correspondre à exactement la même séquence d'ADN, ou porteuses de différences, certaines altérations de la séquence codante donnant des variantes de protéines produites. Les différentes variantes sont les allèles du gène.

Suivant les organismes les nombres de copies d'un même gène et ses allèles peuvent varier. Il peut en être ainsi entre leurs variantes au sein d'une espèce, entre les différentes cellules d'une population monocellulaire, ainsi que dans une proportion de différentes cellules d'êtres multicellulaires. Selon les circonstances ce « dosage génétique » peut avoir des conséquences sur le phénotype, ou ne produire aucun effet observable. Des effets sont généralement attendus car la quantité de la protéine issue d'un gène, dépend a priori du nombre de ses copies de séquence d'ADN dans le génome. Les phénomènes et les mécanismes concernés par le dosage génétique, conduisant ou non à des effets, ainsi que leurs conséquences évolutives sont encore très débattues et mal connues. Leur compréhension implique en effet des notions difficiles à la fois de génétique et de biologie des systèmes en raison des nombreuses interactions en jeu et de propriétés quantitative qui interviennent.

Répétitions de l'information génétique La plupart des espèces vivantes possèdent plusieurs copies de leur lot de chromosomes. Ce nombre d'exemplaires du jeu complet de chromosomes est nommé la ploïdie de l'organisme. Plusieurs situations de ploïdie existent définies dans l'encadré 5.3 et correspondent à des circonstances biologiques génétiques différentes. Les espèces sexuées portent un nombre pair de chromosomes permettant un brassage chromosomique entre les parents dans la production de gamètes pour la reproduction. La variabilité génétique permises par l'existence des multiples copies de l'information génétique, la ploïdie, donne un sens évolutif à ce phénomène. Ainsi chez les plantes, qui ont souvent plus de deux copies des chromosomes, la polyploïdie est un facteur-clé d'apparition de variétés nouvelles. Les humains et la plupart des animaux sont diploïdes, les cellules qui contiennent chacune les chromosomes en double.

Plusieurs définitions décrivent des situations différentes de nombres de copies de jeux de chromosomes qui peuvent avoir des conséquences biologiques différentes. On parle ainsi de cellule :

haploïde si elle possède 1 jeu complet, soit n chromosomes.

diploïde si elle possède 2 jeux, donc $2n$ chromosomes, organisés en n paires.

polyploïde si elle possède au moins 3 jeux : triploïde ($3n$ chromosomes), tétraploïde ($4n$ chromosomes).

aneuploïde si le nombre de jeux de chromosomes est considéré comme « anormal », bien que généralement viable chez les plantes.

autopolyploïdie si le nombre de jeux de chromosomes résulte d'une multiplication d'un même génome.

allopolyploïdie si les jeux de chromosomes proviennent de deux ou plusieurs génomes différents.

Encadré 5.3 – Définitions autour de la ploïdie, d'après Wikipédia : ploïde et aneuploïde.

La redondance de l'information génétique est surprenante car elle ne semble pas optimale et représente un coût métabolique pour les organismes. Aussi des raisons d'être doivent l'expliquer. Un des avantages de la polyploïdie pourrait simplement être une meilleure robustesse permise une continuité fonctionnelle assurée par une des copies de gène intacte en cas de mutations délétères de l'une des allèles. Cette explication simple n'est cependant pas suffisante comme le démontrent de nombreuses observations discutées dans [VEITIA, 2005] : des allèles, mêmes identiques, ne se substituent souvent pas. Dans ces situations pour une physiologie normale les deux copies doivent généralement être simultanément actives. D'autres exemples suggèrent au contraire la pertinence de l'idée initiale : pour certains gènes la présence de deux copies permet bien une résilience en cas de défaillance de l'une d'elles.

La ploïdie intervient également dans la définition des traits héréditaires des phénotypes : les variations des caractéristiques individuelles par rapport à celle des parents. La ploïdie correspond au nombre de copies de l'ensemble des chromosomes, donc du nombre de loci (positions) pour chaque gène, chaque position pouvant différer du gène, dite une forme allélique, ou simplement allèle. La notion de ploïdie, qui va au-delà des deux copies des chromosomes humains et concernent suivant les organismes des nombres de duplications très divers, est donc aussi reliée à la question des interactions entre allèles d'un même gène et la génération de la diversité des traits héréditaires.

Dans les deux sections suivantes je présente des contributions de modélisation auxquelles j'ai participé dans le cadre d'études prospectives sur des phénomènes et mécanismes déterminés et impliquant le dosage génétiques.

5.5.1 Dosage, haploinsuffisance et équilibres génétiques

Le niveau d'expression d'un gène est déterminé par les mécanismes de régulations multiples au niveau de l'ADN codant sur le promoteur, ainsi qu'au niveaux de la transcription, de la traduction des transcrits et de la maturation protéine. Tous ces processus peuvent a priori se dérouler en parallèle sur chaque copie de séquences d'ADN du gène présente dans le génome. On s'attendrait alors à une simple relation linéaire de l'abondance du produit d'un gène avec son dosage. Or le niveau d'expression d'un gène, la quantité de sa protéine produite peut-être déterminante sur le bon fonctionnement des processus dans lequel il est impliqué.

Haploinsuffisance Dans certaines situations, suite à une reproduction anormale, ou suite à une extinction génétique, seul un allèle se retrouve actif parmi les différentes copies existantes. On parle alors de haploinsuffisance, quand l'allèle en copie unique ne suffit pas à assurer la fonction biologique dans laquelle son produit est impliqué.

La première conséquence possible d'un manque d'un allèle actif, c'est, pour une cellule diploïde, la diminution de moitié de la quantité du produit du gène. Dans ce cas, pour les gènes haploinsuffisance la protéine correspondante existe bien dans la cellule, mais du fait qu'elle est synthétisée par une seule copie du gène, son abondance est réduite à un niveau qui ne permet pas d'assurer son rôle fonctionnel normal. La diminution, a priori d'un facteur un demi, de l'abondance des protéines affectées empêche le phénotype normal. D'où vient cette sensibilité critique à la réduction quantitative du niveau d'expression d'un gène dans l'haploinsuffisance ?

Haploinsuffisance et non-linéarité Les conséquences de désordres haploinsuffisants semble faire intervenir une amplification non-linéaire de la condition génétique. Comprendre l'haploinsuffisance implique donc d'identifier les mécanismes de cette amplification du signal génétique en entrée vers sa réalisation fonctionnelle. Ceci permettra également de mieux comprendre les mécanismes de robustesse existant dans les génomes qui conduisent aux exceptions à l'haploinsuffisance.

Ceci peut se voir déjà au niveau des fonctions de réponses de la transcription. Comme discuté aussi dans la section 5.4.1, les fonctions sigmoïdes associées à l'expression génétique se comportent comme des interrupteurs moléculaires. Quand la fonction de réponse est nettement non-linéaire en raison de synergie ou coopérativité l'existence d'un seuil marqué rend le système très sensible à des changements de dosage de l'activateur. Une diminution de moitié de la concentration de l'activateur peut l'amener en dessous du seuil d'activation et arrêter la réponse transcriptionnelle conduisant à des phénotypes anormaux haploinsuffisants. La non-linéarité des fonctions de réponse transcriptionnelles est une propriété nécessaire au contrôle précis de l'activation génétique, mais elle a pour conséquence la sensibilité à l'haploinsuffisance.

Haploinsuffisance et équilibres stœchiométriques fonctionnels Les gènes sont liés dans des réseaux d'interactions imposant des rapports précis entre les abondances des protéines pour accomplir leurs rôles fonctionnels. Ces équilibres viennent essentiellement des nombres relatifs des différents types de protéines au sein de complexes macromoléculaire. Il peut aussi s'agir de l'implication de protéines à des étapes successives d'une chaîne de réactions dans des rapports précis. Dans tous les cas la fonction biologique (macromolécule fonctionnelle, ou chaîne de réactions opérationnelle) exige un rapport correct entre les nombres produits des gènes disponibles. Dans son réseau d'interactions multiples le génome contiennent donc un grand nombre de contraintes sur l'équilibre des doses de tous les produits génétiques.

En conséquence tout déséquilibre dans les rapports stœchiométriques des doses de composants dans des complexes macromoléculaires ou bien dans des réseaux de réactions aura des conséquences sur le phénotype. Les changements de ces rapports peuvent résulter d'une mul-

tiplicité de facteurs soit au niveau des copies individuelles avec par exemples des mutations non fonctionnelles sur l'une d'elle, de changements de la régulation d'une copie, de variations lors reproduction ou d'événements exceptionnels de duplication de chromosomes, voir de génomes. Dans tous les cas l'altération des équilibres peut donc affecter le fonctionnement normal [BIRCHLER et collab., 2001]. Ce thème est développé dans notre publication [VEITIA et collab., 2008] présentée plus bas.

Dosage génétique et dominance allélique En raison des conséquences possibles sur les phénotypes ou des pathologies, la question des équilibres de dose entre produits de gènes nécessaires dans une fonction et sont aussi reliée des sujets classiques de génétique d'hérédité, de développement et de maladies. Ce lien fait l'objet depuis près d'une vingtaine d'années de questionnement par des généticiens comme Reiner Veitia et James Birchler avec lesquels j'ai eu la chance de collaborer. Dans plusieurs articles conceptuels ces chercheurs ont en effet relevé quantités d'observations et d'argumentations démontrant comment les génomes prennent en compte le dosage ou en sont affectés. C'est une question importante aussi d'un point de vue médical en raison des pathologies liées à des variations du nombre de copie de gènes dues en particulier à des nombres anormaux de chromosomes. C'est le cas des trisomies présentant une copie supplémentaire d'un chromosome. Mais également des monosomies où subsiste une seule copie.

Ces études ont conduit à la formulation de la **théorie de l'équilibre de dosage** qui explique l'importance de bilans entre niveaux d'expression de différents gènes et en conséquence l'existence de contraintes sur les variations de nombre de copies de gènes pour l'adaptation génétique dans l'évolution naturelle. Cette théorie se fonde sur la constatation simple de l'inter-dépendance de la plupart des gènes car les fonctions biologiques impliquent des actions combinées. Au niveau des protéines ceci se manifeste par exemple dans les complexes macromoléculaires liant ensemble plusieurs protéines, produits de gènes différents, pour former l'unité fonctionnelle. Un autre cas se retrouve dans les réseaux et voies de réactions biochimiques cellulaires qui impliquent toujours de nombreuses enzymes. L'activité des réseaux dépend des abondances respectives des différentes protéines impliquées. Ainsi les voies de signalisation font généralement intervenir des enzymes avec des activités opposées, comme une kinase ou une phosphatase, agissant sur un substrat commun. La bonne exécution du processus de signalisation dépend de la bonne abondance relative de l'une et l'autre [VEITIA, 2005], donc de l'équilibre de dosage des gènes correspondants.

L'adaptation naturelle doit donc avoir tendance à minimiser les déséquilibres que causeraient des mutations et tout accident touchant le génome. En conséquence on peut s'attendre à ce que des mécanismes on dû évoluer pour compenser tout déséquilibre de dosage d'un gène qui pourrait se produire. De telle manière, en cas de changement du nombre de copies d'un gène, diminuant de moitié ou doublant le niveau de son produit, un ajustement de l'expression de ses partenaires pourrait se produire pour suffisamment compenser la variation accidentelle initiale.

En cas d'absence de mécanisme de compensation, la nécessité d'un équilibre de dosage devient une contrainte pour l'évolution. La duplication de gènes et de génomes est la source principale de création de nouveaux gènes. Cependant les divergences apparaissant entre les unités dupliquées par l'effet de mutations vont générer en s'accumulant des déséquilibres de dosage croissants dont la résolution façonnera le génome. L'importance des effets de dosage lors de duplications de gènes avait déjà été relevée au début des années 1970 par OHNO [1970] évoquant le risque de changements de la concentration de produits de gènes par rapport au reste du génome non dupliqué. En conséquence des duplications du génome entiers seraient plutôt favorisés par rapport à celles de chromosomes individuels ou de gènes isolés, car elles évitent justement les déséquilibres de dosage [CONANT et collab., 2014].

Afin d'avoir à disposition un complexe protéique donné en quantité suffisante, l'expression des gènes correspondants est contrainte à les produire dans les bons rapports nécessaires à la physiologie. Ceci impose la préservation au niveau de l'expression génétique des rapports stœchiométriques entre leurs protéines dans les interactions où elles interviennent. En réalité, comme

on le verra plus loin la contrainte est même plus importante : pour la production suffisante du complexe, non seulement les composants doivent être présents en nombre suffisant, dans les bons rapports, pour le bon assemblage, mais leur cinétique de production doit également être ajustée en fonction des réactions d'assemblage possibles.

La formation de complexes macromoléculaire peut-être schématiquement décrite par des réactions d'association et dissociation faciles à décrire par la cinétique chimique. Les rapports d'équilibre correspondent aux fractions de composés formés en fonction des diverses abondances des monomères de base. J'ai eu la chance d'avoir été introduit à cette problématique par Reiner Veitia qui est un des premiers généticiens à avoir proposé d'analyser les propriétés des équilibres de dosage génétique par la modélisation mathématique [VEITIA, 2002a, 2003].

Les deux sections suivantes 5.5.2 et 5.5.3 présentent quelques éléments de deux publications [VEITIA et collab., 2008] et [VEITIA et collab., 2013] où nous avons développé ces idées et discuté des phénomènes associés aux effets de dosage et de mécanismes de compensation dont nous avons proposé des modélisations permettant des interprétations et des anticipations.

5.5.2 Mécanismes de compensation de dosage

Dans notre publication « Cellular reactions to gene dosage imbalance : genomic, transcriptomic and proteomic effects. » avec Reiner Veitia et John Birchler [VEITIA et collab., 2008] nous avons analysé les mécanismes conduisant à des effets de dosage et des processus de compensation au niveau transcriptionnel et traductionnel, montrant notamment comment la participation de protéines au sein de complexes peut imposer des contraintes d'équilibre, mais également intervenir dans des mécanismes de protection.

Les protéines produites par l'expression de gènes individuels sont souvent des monomères, c'est-à-dire des unités d'ensembles moléculaires plus grands, associant dans un complexe protéique plusieurs copies du même produit ou issus d'autres gènes. Les produits des gènes trouvent alors leur rôle fonctionnel en interaction avec d'autres. L'unité moléculaire active est ce complexe multi-protéique et non pas chaque protéine produit de gène individuel qui le compose. L'abondance en quantité suffisante pour la physiologie normale du complexe dépendra alors de l'abondance suffisante de tous ses composants. Et en conséquence elle dépendra des bons rapports d'abondance entre chaque type de monomère associé.

Un complexe composé des protéines A et B différentes peut exiger des nombres « stœchiométriques » différents de chaque type à commencer par les hétéro-trimères ABA et BAB³. Le résultat de la sous-expression, ou de la sur-expression d'une sous-unité dans une structure macromoléculaire dépend de ses connections avec d'autres dans le complexe, ainsi que du chemin pour l'assemblage de la structure complète. La figure 5.28 montre ainsi un exemple de l'effet de la sur-expression

Compensation au niveau transcriptionnel. De nombreux gènes impliqués sur des chromosomes aneuploïdes possèdent un niveau d'expression total identique à la situation normale [BIRCHLER, 1979; BIRCHLER et NEWTON, 1981]. Ceci signifie que des mécanismes de compensation du changement structurel du nombre de copies de gènes doivent exister dont le résultat est un niveau des produits proche de celui diploïde.

La figure 5.29 montre des scénarios de mécanismes transcriptionnels permettant une compensation à ce niveau de la synthèse des produits. Les situations représentées impliquent des gènes contrôlés négativement par des gènes régulateurs portés sur le même chromosome ou segment chromosomique. La variation du nombre de copies du gène cible sera donc associée à une

3. on s'intéresse ici uniquement aux rapports d'abondance entre monomères sans tenir compte des variantes d'arrangement.

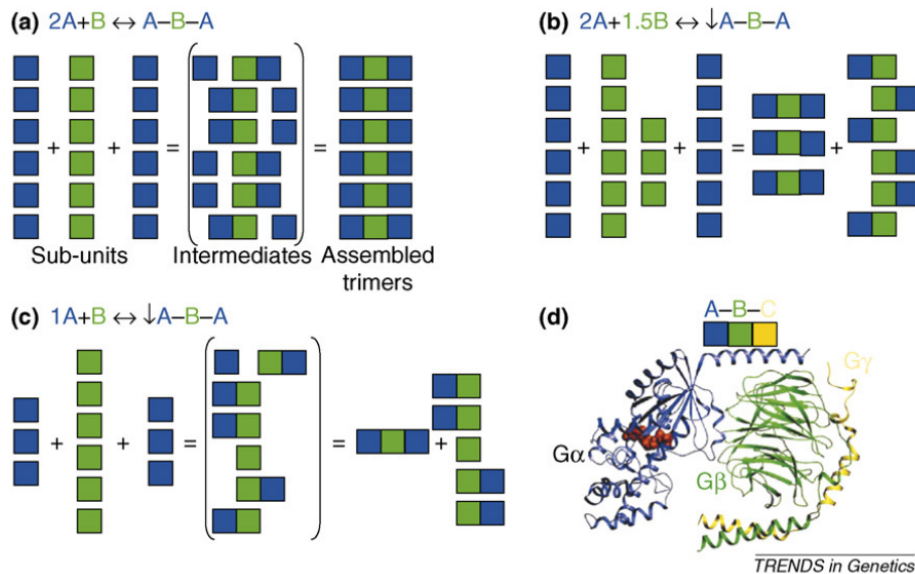


FIGURE 5.28 – Effet de la sur-expression (en raison de la duplication d'un gène) d'une sous-unité sensible au dosage dans un complexe macromoléculaire. a) assemblage du trimer ABA via la formation intermédiaire de dimères AB et BA sous des conditions irréversibles (cad ABA est très stable et ne se dissocie pas.) b) un excès de la molécule centrale B par rapport au cas a) amène à une diminution dramatique de la production de ABA en raison de la production d'intermédiaires qui ne peuvent être complétés à cause d'un nombre insuffisant de monomères A. c) Une diminution de A par rapport à la situation en a) amène à une diminution de ABA encore plus importante en raison de l'excès du monomère B. c) Structure d'un exemple d'hétérotrimère ABC, la protéine G ($G\alpha$ - $G\beta$ - $G\gamma$), impliqué dans la transduction de signal. La surexpression du monomère central β dans la levure conduit à un excès anormal de $\beta\gamma$. Source : [VEITIA et collab., 2008].

variation équivalente de son régulateur. En résulte une compensation du dosage au niveau transcriptionnel. Il est cependant improbable que chaque gène robuste au dosage soit lié avec un régulateur inverse. D'autres mécanismes plus globaux doivent exister.

Compensation au niveau des protéines De nombreux exemples montrent des situations dans lesquelles la participation d'une protéine dans un complexe peut affecter sa stabilité [ASHER et collab., 2006]. Une raison probable à cette observation est la protection à des attaques de protéases dont jouiraient les protéines liées les unes aux autres. Un effet qui a été désigné du nom de *stabilité coopérative* [TOMPA et collab., 2008]. Pour agir dans leur mission les enzymes protéases ont besoin de se fixer sur les molécules de protéine qu'elles vont digérer. Ceci se produit préférentiellement à certaines positions des protéines ciblées, correspondant à des sites de signaux de dégradation. La formation d'un complexe multi-protéique peut alors masquer ces sites d'attachement préférentiels et les rendre inaccessibles aux protéases.

Nous avons exploré cette idée dans le cas d'un trimère ABA en supposant que le monomère central B porte deux signaux de dégradation, chacun masqué indépendamment lors de la formation des deux dimères intermédiaires AB et BA. Une modélisation de cinétique chimique très simple permet de calculer les abondances attendues des différentes unités en fonction des scénarios pour les effets de dosage. Cette modélisation ne représente en soi aucune difficulté technique et suit les principes de la cinétique chimique. L'essentiel du travail est ici aussi dans la conceptualisation des processus, leur formalisation comme réactions chimiques et le déterminisme des paramètres. Pour illustration de cette démarche qui se retrouve dans la plupart des modélisations de cette section, je rapporte ci-dessous l'ensemble de réactions modélisant la formation du trimère ABA et les réactions de dégradations :

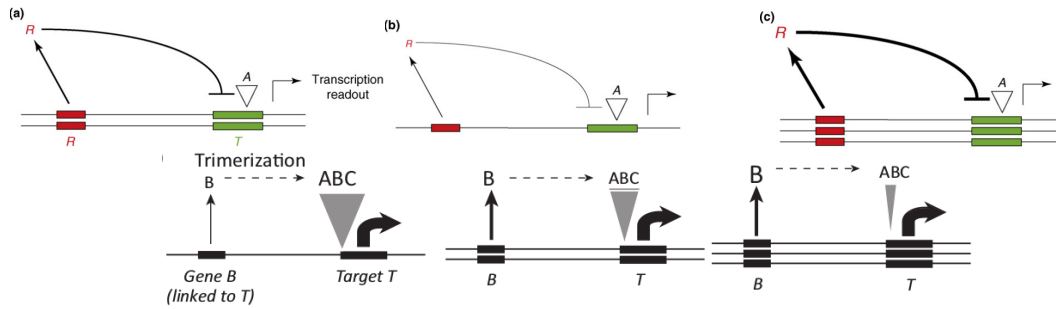
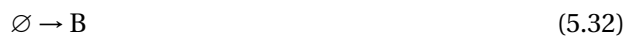


FIGURE 5.29 – **Compensation locale de dosage.** (a) La compensation peut-être attribuée à un effet de dosage inverse. Dans la situation diploïde, la régulation du gène cible (T) est obtenue par une balance entre les effets d'un activateur A et d'un régulateur négatif R encodé par le gène lié à T. (b) Lorsqu'il y a une copie unique de la région contenant la cible T, le gène encodant le régulateur inverse R sera aussi en copie unique. Une plus faible quantité de R et une quantité normale de A donnerait une régulation positive de T (c'est-à-dire de production de T par allèle). (c) Dans des situations de trisomie, il y aura trois copies à la fois du gène structurel et de son répresseur qui va inhiber le précédent. Dans les deux cas des niveaux à peu près diploïdes de l'expression des gènes cibles sont attendus. Figure du bas : les gènes B et T sont liés physiquement. Alors ils peuvent être co-supprimés ou co-dupliqués. La quantité normale de ABC n'est pas maximale mais suffisante pour permettre une activation transcriptionnelle de T. La suppression hétérozygote à la fois de B et T conduirait à une diminution de la quantité de B. La copie unique de T restante sera donc activé par une concentration double de ABC (par rapport à la situation diploïde normale). Si l'activation transcriptionnelle de ABC est directement proportionnelle à sa concentration, alors on a une compensation de dosage. (C) Quand B et C sont co-dupliqués, trois copies de T sont activés par seulement deux-tiers de la concentration normale de ABC, conduisant aussi à de la compensation. Source : [VEITIA et collab., 2008] et [VEITIA et collab., 2013]



Les réactions (5.31) et (5.32) correspondent à la production des composés A et B avec de taux s_A et s_B ; la formation du dimère par les réactions (5.33) et (5.34) avec des taux d'association p_{AB} et p_{BA} et de dissociation m_{AB} et m_{BA} , on considère ici des dimères asymétriques ($AB \neq BA$), la formation du trimère suit alors deux routes (5.35) et (5.36) avec des taux d'association p_{ABA_1} et p_{ABA_2} , le trimère peut être dissocié dans les deux composés intermédiaires avec des taux m_{ABA_1} et m_{ABA_2} ; les autres réactions décrivent la dégradation de chaque unité. Avec l'approche classique de cinétique chimique ces réactions conduisent au système d'équations différentielles ordinaires ci-dessous, qu'il serait possible de simplifier si une étude dynamique plus fine était nécessaire :

$$\begin{aligned}
 \frac{d[A]}{dt} &= sA + mAB * [AB] + mBA * [BA] + mABA_1 * [ABA] + mABA_2 * [ABA] \\
 &\quad - (pABA_1 * [AB] + pABA_2 * [BA]) * [A] \\
 &\quad - (pAB + pBA) * [A] * [B] - dA * [A] \\
 \frac{d[B]}{dt} &= sB + mAB * [AB] + mBA * [BA] \\
 &\quad - (pAB + pBA) * [A] * [B] - dB * [B] \\
 \frac{d[AB]}{dt} &= pAB * [A] * [B] + mABA_1 * [ABA] \\
 &\quad - (mAB + dAB) * [AB] - pABA_1 * [AB] * [A] \\
 \frac{d[BA]}{dt} &= pBA * [B] * [A] + mABA_2 * [ABA] \\
 &\quad - (mBA + dBA) * [BA] - pABA_2 * [BA] * [A] \\
 \frac{d[ABA]}{dt} &= (pABA_1 * [AB] + pABA_2 * [BA]) * [A] \\
 &\quad - (mABA_1 + mABA_2 + dABA) * [ABA]
 \end{aligned}$$

Comme déjà évoqué, la seule particularité de cette modélisation est la conceptualisation biologique des phénomènes, dans le système formel des équations biochimiques, et dans une moindre mesure la paramétrisation des équations avec des valeurs biologiquement motivées des différents taux. L'enjeu du modèle ici ne nécessite pas d'aller au-delà de l'intégration numérique du système et l'interprétation des courbes présentée dans la figure 5.30. Ces courbes montrent un comportement non intuitif : l'augmentation de la dégradation du monomère libre s'accompagne d'une plus grande robustesse face à des changements des taux de synthèse des protéines en cas d'accidents de dosage (duplication ou suppression d'une des copies alléliques). Ceci signifie que certaines propriétés constitutives de protéines, peuvent implémenter automatiquement une forme de compensation à des déséquilibres de dosage. Dans cet exemple des sous unités de complexes multi-protéiques portant plusieurs signaux de dégradation pouvant être masqués au sein d'un assemblage subissent une dégradation préférentielle lorsque les monomères en excès. Selon le modèle ci-dessus ceci suffit à assurer une certaine compensation de dosage sur les niveaux d'abondance du complexe protéique si l'une de ses unités subit une variation de dosage.

Compensation par effet de tampon génomique Le phénomène de tampon discuté dans notre autre publication « Gene dosage effects : nonlinearities, genetic interactions, and dosage compensation. » [VEITIA et collab., 2013] est un autre mécanisme favorisant une robustesse de la transcription en cas de variation du dosage génétique.

Ici nous avons étudié un effet global de tampon sur des variations de dosage venant de l'interaction non spécifique entre un facteur de transcription et l'ADN total du génome en dehors de ses cibles spécifiques dans des régions régulatrices. Un facteur de transcription reconnaît en effet avec une forte affinité des cibles dans les régions cis-régulatrices des gènes qu'il régule. Il se lie aussi de manière non spécifique, à plus faible affinité, avec tout le reste de l'ADN, donc sur un nombre considérablement plus élevé de sites. On notera ADN_s les éléments de liaison dans des régions cis-régulatrice et ADN_{ns} les sites non spécifiques répartis sur tout le génome. Des sites de liaisons spécifiques peuvent également exister sans rôle fonctionnel en dehors de promoteurs, ils contribuent aux bilans d'équilibre par rapport au paramètre nombre de promoteurs et on ne les distinguera pas explicitement.

Notre modèle considère une situation à l'équilibre dans laquelle un nombre donné de facteurs peut se lier en compétition entre les quelques sites spécifiques ADN_s et les nombreux sites non spécifiques ADN_{ns} . Ce modèle est mathématiquement la simple expression d'équilibres de cinétiques chimiques et je ne le reproduis pas ici. Le seul point d'attention est sa résolution numérique en raison des non linéarités. Il s'agit de calculer la fraction d'occupation des promoteurs

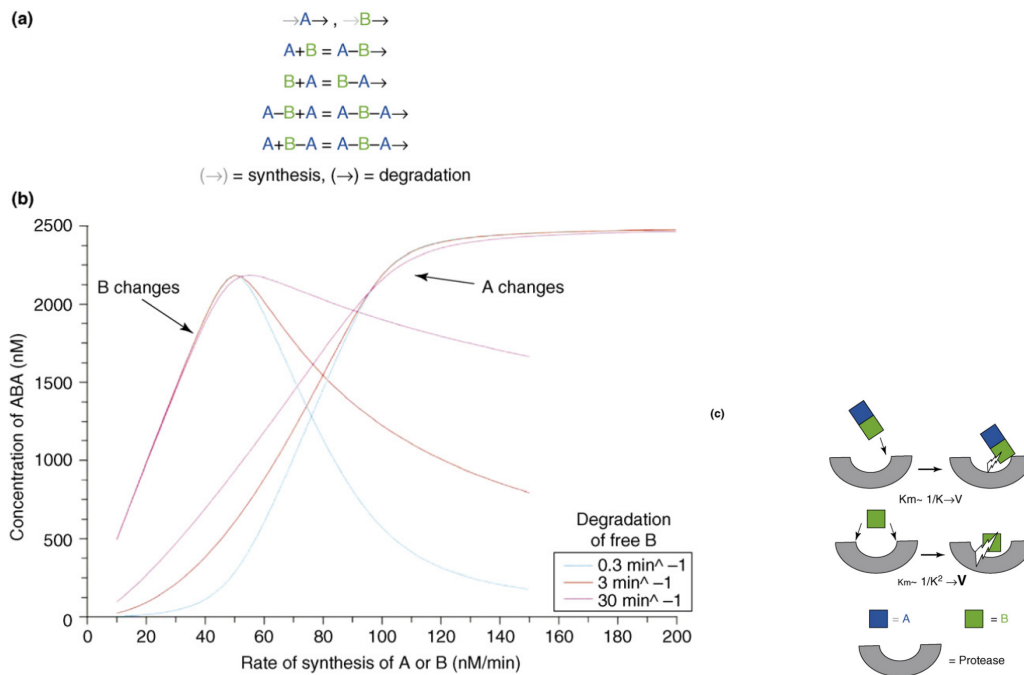


FIGURE 5.30 – **Compensation de déséquilibre de dosage au niveau des protéines par protection contre la protéolyse, exemple sur un trimère ABA.** a) Schéma de réactions pour la production de ABA, les flèches représentent les réactions de synthèse et de dégradation. b) courbes de la concentration de ABA (nM) en fonction des taux de synthèse de A et B (nM/min) à l'état stationnaire. Les courbes de formes sigmoïdes (marqués par la flèche « A changes ») correspondent à l'effet du changement du taux de synthèse de A en gardant celui de B constant (à 50 nM/min). Les trois courbes correspondent à des taux de dégradation constants de B différents suivant la légende colorées). On note que la concentration maximale de ABA est atteinte (plateau) quand la synthèse de A est à peu près le double de celle de B, correspondant au rapport des stœchiométries de A dans le trimère par rapport à B. En cas d'un défaut sur A conduisant à une plus faible synthèse la production de ABA décroît fortement. La courbe en cloche (marqué par la flèche « B changes ») représente l'effet d'un changement de la synthèse de B avec le taux de A constant à 100 nM/min. Le taux de production maximal de ABA est obtenu quand la synthèse de B est à peu près la moitié de celle de A, qui correspond à l'équilibre stœchiométrique entre les unités dans le trimère. On remarque que lorsque le taux de dégradation de B est élevé (courbe violette) la sensibilité à des variations de dosage est moins importante. (c) Scénario du modèle : quand la protéine B est associée à A (comme dans AB et BA) elle expose un seul signal de dégradation et la protéase le reconnaît avec une constante d'association K. Quand B est libre (lorsque la protéine est en excès par exemple), elle expose deux signaux de dégradation et la protéase possède une association beaucoup plus importante (K^2) : la dégradation des monomères libres B sera rapide. Source : [VEITIA et collab., 2008]

(régions cis-régulatrices) et d'en calculer les rapports entre les mutants avec des variations de dosage et le cas normal. La figure 5.31 montre le résultat principal de la compensation possible à des variations de dosage en fonction des constantes de dissociations K_{ns} et K_s respectives des liaisons non spécifiques et spécifiques. En absence de tout effet de tampon on s'attendrait à un changement d'un facteur 0,5 avec la perte d'un chromosome et d'un facteur 1,5 en cas de chromosome surnuméraire. Cette figure illustre comment en fonction de quantités données d' ADN_s et d' ADN_{ns} il existe des zones de combinaisons K_{ns} et K_s avec un tampon maximal. Un tel effet tampon impliquerait une co-évolution des concentrations des facteurs de transcriptions intracellulaires, ou de leurs constantes de dissociation, en lien avec les nombres de sites spécifiques et non spécifiques dans le génome.

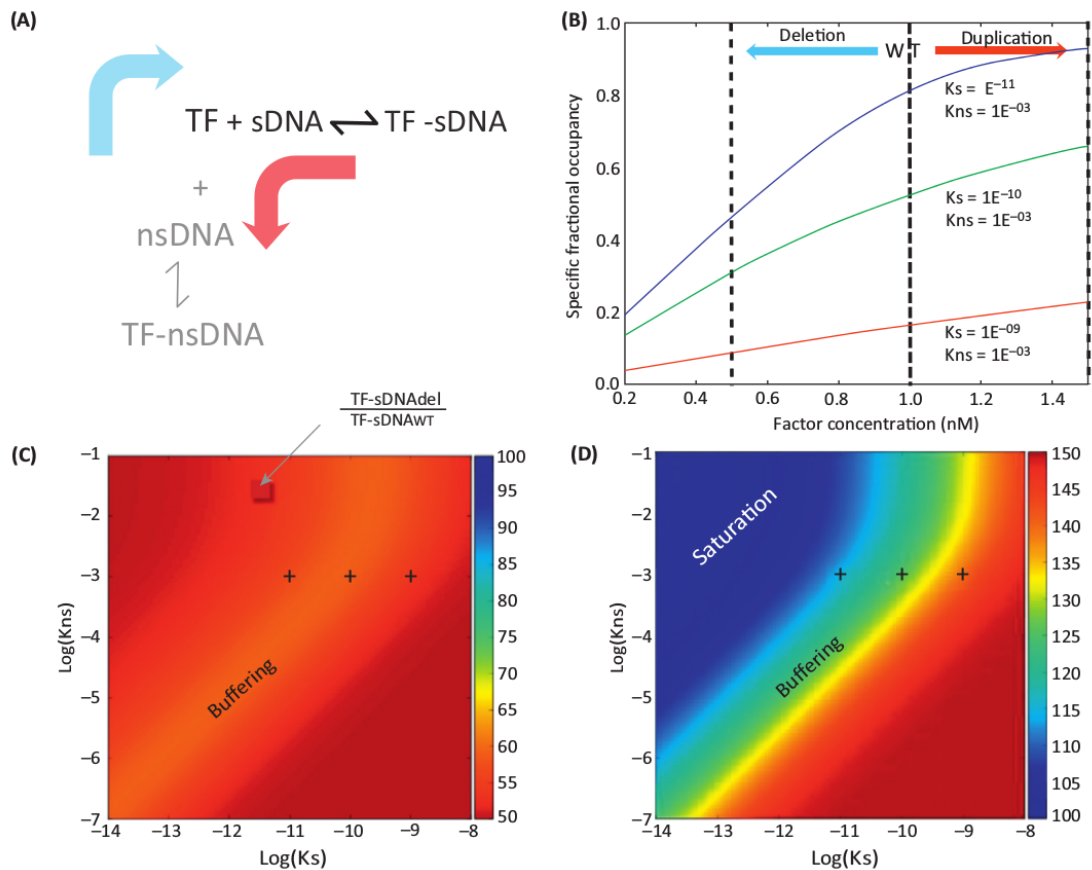


FIGURE 5.31 – Effet tampon sur les effets de dosage dans la transcription dû aux liaisons non spécifiques. A) réactions représentant la liaison d'un facteur de transcription sur un site de régulation spécifique (sDNA), donc une liaison forte de constante de dissociation K_s faible, ou sur un site de liaison non spécifique (nsDNA). La flèche bleue représente un décalage de l'équilibre en faveur des liaisons spécifiques en réponse par exemple à une délétion hétérozygote du gène codant pour le facteur de transcription. Le décalage contraire vers les liaisons non spécifiques est indiqué en rouge et survient par exemple lors de duplication du gène codant le facteur de transcription. B) courbes représentant l'occupation fractionnée d'un site spécifique (fraction des sites occupés sur nombre total de sites spécifiques) pour différentes concentrations de facteur de transcription et trois différentes combinaisons de K_s et K_{ns} correspondant aux croix indiquées sur les cartes d'intensités en C) et D). L'occupation fractionnée a été évaluée dans des conditions d'équilibre supposant 300 sites spécifiques et 10^9 sites non spécifiques dans un volume nucléaire de $5 \cdot 10^{-13}$ l. Pour une constante de dissociation spécifique élevée $K_s = 10^{-9}$ M (plus faible liaison spécifique), il n'y a pas d'effet tampon : en cas de délétion ou duplication du régulateur l'occupation du promoteur change proportionnellement. Pour K_s plus faible (10^{-10} M, affinité plus forte) on constate un effet tampon : quand la concentration du régulateur est divisée par deux (de 1 nM à 0.5 nM) : la fraction de sites spécifiques occupés d'un facteur *plus grand* que la valeur attendue 0,5 ; quand la concentration du régulateur augmente d'un facteur 1,5 comme sous l'effet d'une duplication d'une copie, on a aussi un effet tampon car la fraction de sites spécifiques occupés augmente d'un facteur *plus petit* que 1,5. C) Carte d'intensité représentant l'effet d'une diminution par deux de la concentration du facteur de régulation représenté en pourcentage de variation en fonction de K_s et de K_{ns} : rouge (50%) signifie aucun tampon (c'est-à-dire une diminution de moitié du régulateur amène une diminution de moitié de la régulation d'un gène cible). Les régions oranges montrent différents degrés de tampon : en dépit de la réduction de moitié du facteur de transcription l'occupation spécifique d'un régulateur cible ne diminue pas de moitié par rapport à la normale. D) Carte d'intensité de l'occupation spécifique pour une augmentation d'un facteur 1,5 du régulateur (comme dans le cas d'une duplication d'un allèle codant pour le facteur de transcription). Ici le rouge (150%) signifie absence de tampon. Le bleu foncé (100%) pourrait faussement être interprété comme tamponnage parfait, mais correspond en réalité à la saturation des sites spécifiques dans la condition normale : tous les sites sont déjà occupés par un régulateur et rien ne change en cas de surexpression. Source : [VEITIA et collab., 2013]

5.5.3 Un lien entre robustesse au dosage et AND « poubelle » ?

Dans le cadre de notre intérêt pour le dosage génétique et les contraintes d'équilibre qui s'y rapportent dans un génome, ma publication avec R. Veitia « Whole genome duplications and a 'function' for junk DNA? Facts and hypotheses » [VEITIA et BOTTANI, 2009] propose l'existence d'un lien entre le contenu global d'ADN dans le génome d'un organisme et la robustesse à des variations de dosage.

Contrairement à ce que l'on avait cru longtemps, la quantité d'ADN portant l'information héréditaire d'un organisme, la taille de son génome, apparaît n'avoir aucun rapport avec ce que l'on pourrait qualifier comme sa complexité, à savoir notre interprétation de la sophistication de l'ensemble de ses fonctions. Les résultats du séquençage de génomes entier depuis le tournant du millénaire ont clairement mis en évidence, avec surprise de la communauté scientifique, l'absence de lien entre dimension du génome et propriétés phénoménologiques globales des êtres vivants. Des organismes considérés comme globalement plus simples que d'autres ont pourtant des génomes de dimensions comparables ou plus grandes que d'autres. Cette constatation porte le nom de *paradoxe de la valeur C* et est encore mal comprise. Cette valeur correspond à la taille totale en ADN du génome unique (sans tenir compte de la ploïdie). Quels facteurs contribuent à la taille du génome si elle n'est pas corrélée avec la complexité apparente des organismes hôtes ?

Les génomes portent de nombreuses fractions de séquences d'ADN non codantes pour des gènes, qui contribuent de manière très significative à leur volume total [LANDER et collab., 2001; PIEGU et collab., 2006], sans rôle fonctionnel. L'existence de ces parties répétitives non fonctionnelles du génome suscite beaucoup d'interrogations sur leur raison d'être. Comment expliquer leur persistance, leur prolifération même, d'un point de vue évolutif, en absence de rôle fonctionnel ? L'incompréhension concernant leur rôle a même conduit à les qualifier d'*ADN poubelle* (« junk DNA »). Ou à les considérer comme des éléments de séquences presque autonomes, « égoïstes » se propageant dans les organismes sans contribution à la physiologie.

Cependant de nombreux travaux et des résultats qui se sont multipliés ces dernières années découvrent des fonctions et des effets phénotypiques de ces séquences non codantes [CAREY, 2015; ENCODE PROJECT CONSORTIUM, 2012]. Dans notre publication on propose un rôle de l'ADN non fonctionnel justifiant sa présence, lié à des propriétés globales physique de la cellules imposées par l'équilibre de dosage nécessaire entre produits de gènes en interaction après des événements massifs de duplication du génome. En effet la plupart des espèces ont connu durant leur évolution des événements évolutifs majeurs de duplication de leur génome entier, source de la polyploïdie. Ceci est le cas notamment des végétaux dont le degré de ploïdie est souvent élevé. On estime que 30 à 80% des plantes sont polyploïdes [MEYERS et LEVIN, 2006; OTTO, 2007; RIESEBERG et WILLIS, 2007] et ont connu plusieurs événements de duplication de leur génome entier dans leur histoire évolutive [LYONS et SCHNABLE, 2015]. Des événements de duplication du génome entier on également eu lieu dans les animaux et au moins deux duplications ont été identifiées dans la lignée de vertébrés ayant conduit aux humains [DEHAL et BOORE, 2005].

Le volume nucléaire est proportionnel au contenu d'ADN Notre proposition s'appuie essentiellement sur l'observation dans de nombreux travaux d'une corrélation positive entre la quantité d'ADN génomique et les volumes nucléaires et cellulaires [CAVALIER-SMITH, 1982; MIRSKY, 1951]. La quantité brute d'ADN, indépendamment de sa séquence semble déterminer directement le volume aussi bien du noyau que de la cellule entière. Ceci pourrait s'expliquer par la nécessité d'une densité optimale dans le noyau et dans le cytoplasme afin d'assurer les différentes fonctions moléculaires et les échanges entre les compartiments. L'ADN et ses protéines associées jouent alors un rôle structurel pour le volume du noyau, qui se répercute sur celui de la cellule [CAVALIER-SMITH et BEATON, 1999]. Un résultat direct et très significatif pour notre discussion vient d'observations sur des séries de mutants de ploïdies croissante, dont le volume varie de la même manière [GREGORY, 2001].

Duplication de génome entier et disruption des équilibres de dosage En raison du lien entre volume et contenu d'ADN nous nous attendons donc une augmentation proportionnelle, doublement, du volume cellulaire des mutants à la suite d'un événement de duplication du génome entier. En supposant, comme dans nos autres études, un taux de production des produits des gènes proportionnel à leur dosage, la quantité de protéine dans le mutant polyploïde est double de celle chez son progéniteur. En raison de l'augmentation du volume nucléaire dans le même rapport, la concentration dans le noyau restera cependant identique. En conséquence tous les équilibres d'interactions pour la formation de complexes macromoléculaires, ou de régulation dans des voies métaboliques ou de régulation transcriptionnelle restent les mêmes. En effet, tous les gènes sont dupliqués, ainsi que tous leurs partenaires d'interactions et éventuels promoteurs cibles. Un événement de duplication de génome entier n'aura pas de conséquences en termes de déséquilibre de dosage.

Cependant, cette situation n'est probablement pas stable. Des analyses génomiques et des expériences montrent qu'après une duplication de génome il existe une tendance marquée pour revenir à un état diploïde [GERSTEIN et collab., 2006] qui se manifeste par des suppressions d'ADN, des réarrangements chromosomiques multiples et la perte de chromosomes. Ces réarrangements peuvent bien se comprendre car la suppression de gènes et d'ADN non nécessaires en de multiples copies devrait en effet être avantageuse en permettant de gagner sur les coûts d'expression et réplification de ces éléments inutiles. La prolifération de l'ADN non fonctionnel montre cependant que cette réorganisation vers l'état diploïde ne se fait pas rapidement ni même avec un retour au génome initial.

Nous suggérons qu'une raison contribuant à la préservation de la quantité d'ADN augmentée suite à une duplication de génome vient de la nécessité immédiate de la cellule à préserver les bons équilibres de dosage à son fonctionnement que des suppressions massives viendraient perturber. Nous allons comparer les conséquences cinétiques sur deux exemples simples d'un scénario de duplication de génome avec préservation de l'ADN dupliqué (WGD « whole genome duplication ») et d'un scénario de duplication du génome suivi de réarrangements avec suppression massive d'ADN (WGD + Δ). Sous l'effet de la duplication du génome, gènes et partenaires d'interactions de leurs produits sont doublés ainsi que le volume : les concentrations sont préservées.

En cas de réarrangements suivant une duplication on considère que l'ADN concerné est celui non codant, les séquences fonctionnelles, actives dans la transcription sont préservées des suppressions. Ainsi le nombre de copies, doublé, des gènes après l'événement de duplication reste double, mais la quantité totale d'ADN diminue. En conséquence le volume nucléaire diminue lui aussi. Il en résulte que la concentration des produits génétiques va donc augmenter. Comme on va le voir sur des modèles simples de réactions ceci amènera à des changements de cinétiques pouvant avoir des répercussions fonctionnelles.

Modèle simple de cinétique de dimerisation Des conséquences que l'on peut déjà voir dans le cas simple de la cinétique de formation d'un dimère MM à partir d'une protéine monomère M. Un système décrit par les équations de la cinétique chimique :

$$\frac{d[M]}{dt} = S - D[M] - k_+[M]^2 + 2 * k_-[MM] \quad (5.42)$$

$$\frac{d[MM]}{dt} = k_+[M]^2 - k_-[MM] \quad (5.43)$$

où S est le taux de synthèse du monomère M, proportionnel à la concentration de l'ARN message, D le taux de dégradation, proportionnel à la concentration des enzymes de dégradation, k_+ et k_- les taux d'association et de dissociation du monomère.

Si l'on suppose simplement une duplication du génome entier, il ne devrait pas y avoir de changement de la cinétique ci-dessus. En effet les concentrations des produits ne changent pas

en raison du doublement conjoint de la synthèse des protéines et du volume nucléaire. Par contre, si la duplication était suivie de suppression d'ADN, *mais avec préservation des séquences codantes*, les concentrations des protéines augmenteraient, car la synthèse serait double alors que le volume nucléaire ne le serait pas. Pour examiner ces effets on considère dans le modèle le cas limite correspondant à un retour du volume à sa valeur initiale avant duplication. Dans ce cas, la concentration d'ARNm ayant doublé, le taux S est double. C'est aussi le cas pour le taux de dégradation car la concentration d'enzyme double pour la même raison : $S \rightarrow 2 * S$, $D \rightarrow 2 * D$. Les taux d'association et dissociation k_+ and k_- restent cependant identiques car ils ne sont pas affectés par ces variations.

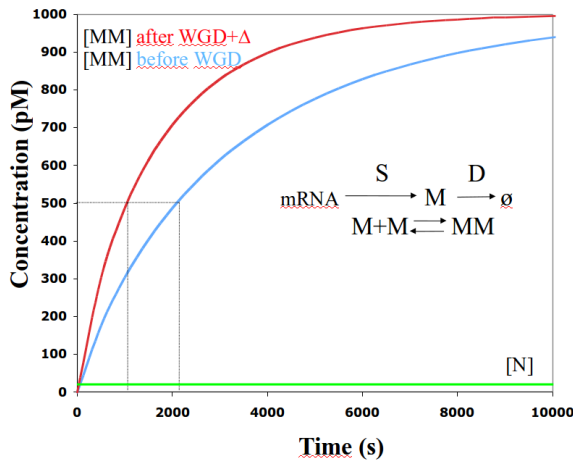


FIGURE 5.32 – **Dynamique de formation d'un dimère MM et duplication de génome.** La formation de MM dépend du taux de synthèse S, de la dégradation du monomère M avec taux D et des taux d'association et dissociation du complexe. Courbe bleue : formation du dimère après duplication du génome (paramètres $S = D = 1$ nM/s). Courbe rouge : après duplication du génome et suppression d'ADN (laissant les gènes codant comme dupliqués et retour au volume initial, $S = D = 2$ nM/s). L'état stationnaire est atteint plus rapidement dans ce cas (courbe rouge) que dans le cas original (original dupliqué avant suppression, identique au diploïde). Source : [VEITIA et BOTTANI, 2009]

Les concentrations stationnaires $[M]_{stat} = \frac{S}{D}$ et $[MM]_{stat} = \frac{k_+}{k_-}$ sont les mêmes avant et après suppression. Mais comme le montre l'intégration numérique du système sur la figure 5.32 la cinétique avec la suppression est différente de celle sans la suppression (identique à celle de l'organisme avant une duplication), même si les états stationnaires sont identiques. Cette différence cinétique peut être cruciale pour des processus s'échelonnant dans le temps. De même si le dimère MM est dans une relation en équilibre avec un autre monomère N qui possède une autre cinétique il peut y avoir un problème avant d'atteindre l'état stationnaire.

Un telle différence du changement de concentration qui suivrait une suppression d'ADN a des conséquences encore plus visibles dans un modèle plus complexe d'un processus dynamique comme l'horloge du cycle de la mitose dont je ne détaillerais pas les équations ici. La figure 5.33 montre comment l'élévation des concentrations d'enzymes qui suivraient une diminution de volume après des suppression d'ADN et un duplication de génome changent radicalement la cinétique de ce phénomène périodique.

Suppression d'ADN et liaison spécifique/non spécifique de facteurs régulateurs La suppression d'ADN qui interviendrait à la suite d'une duplication affecte aussi le rapport entre le nombre de sites spécifiques de régulation, préservé, pour un facteur de transcription et le nombre de sites non spécifique sur le reste du génome. Ici aussi des modèles simples d'équilibres chimique basés sur la cinétique chimique permettent d'extrapoler des comportements possibles.

Pour calculer l'effet des liaisons non spécifiques dans ces processus nous avons considéré un modèle simple dans lequel une protéine de transcription P peut se lier spécifiquement sur des sites S avec une constante de dissociation K_S et de manière non spécifique à d'autres sites D avec une constante K_{ns} ($K_S \ll K_{ns}$) :



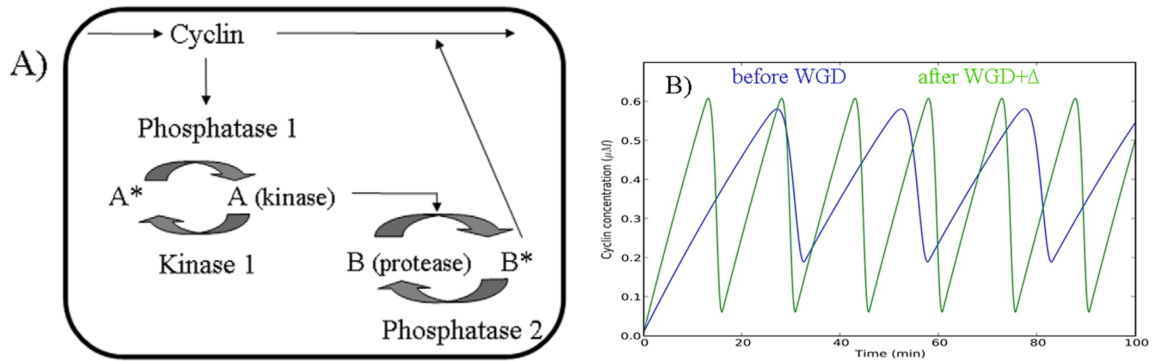


FIGURE 5.33 – Conséquences sur une cinétique biochimique d’une duplication de génome suivie de suppression massive d’ADN. A) Modèle minimal de cycle cellulaire de la mitose d’après Goldbeter [GOLDBETER, 1991] basé sur une cascade de modification post-traductionnelles qui modulent la dégradation d’une cycline par une protéase. Le circuit de rétroaction négatif génère des oscillations. B) La courbe bleue est la variation périodique de la cycline avec le jeu de paramètres standard de Goldbeter. La courbe verte (à plus haute fréquence) correspond à des paramètres de concentration double qui résulteraient d’une duplication de génome entier suivie de suppression massive d’ADN. Source : [VEITIA et BOTTANI, 2009]

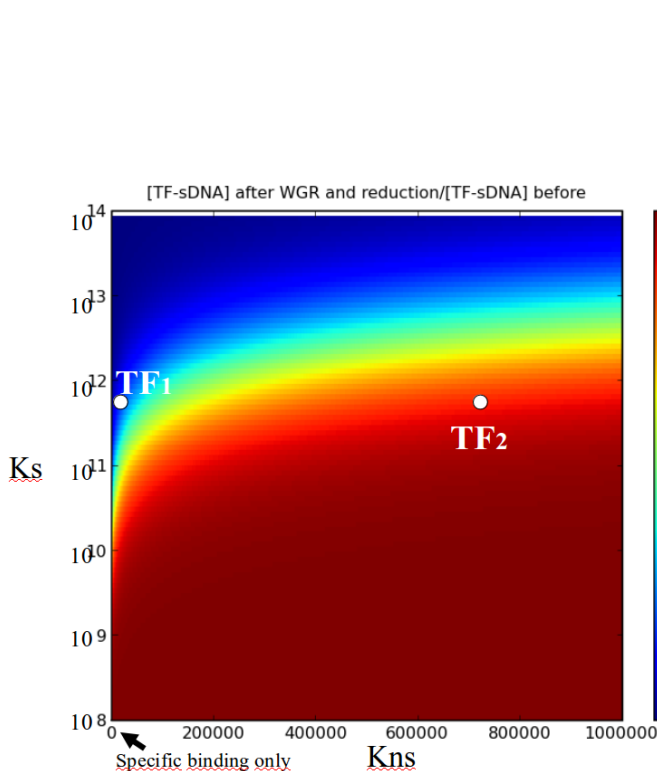


FIGURE 5.34 – Liaison spécifique de facteurs de régulation en présence de différents niveaux de liaison non spécifique. Exploration quantitative de la liaison spécifique de facteurs de régulation en présence de différents niveaux de liaison non spécifique suivant duplication de génome avec suppression d’ADN suivant le modèle du paragraphe ???. Un facteur [TF]= 1 nM reconnaît ici spécifiquement 10 sites de liaison dans le noyau avec une constante spécifique K_s entre 10^8 et 10^{14} M) et une liaison non spécifique K_{ns} plus faible 0 à 10^6 . La concentration intranucléaire des sites spécifiques est de l’ordre de $3 \cdot 10^{11}$ M (supposant un volume nucléaire de $5 \cdot 10^{-13}$ L). La concentration des sites non spécifiques est supposée 7 ordres de magnitude plus grande que celle des sites spécifiques. L’échelle de couleurs représente le rapport de la concentration des facteurs de transcription liés spécifiquement dans le cas d’une duplication suivie de suppression (WGD + Δ , qui laisse le gène du facteur plus de ses cibles dupliqué), sur la concentration du facteur lié aux sites spécifiques avant duplication du génome. Source : [VEITIA et BOTTANI, 2009]

À l'équilibre on a

$$\frac{[P][S]}{[PS]} = K_S \quad (5.45a)$$

$$\frac{[P][D]}{[PD]} = K_{ns} \quad (5.45b)$$

et la conservation

$$[P]_{\text{free}} + [PS] + [PD] = [P]_{\text{tot}} \quad (5.46)$$

En utilisant ces relations il est simple d'exprimer la concentration d'un complexe spécifique de la protéine lié à l'ADN [PD] comme une fonction de la concentration de protéine totale $[P]_{\text{tot}}$:

$$[PS] = \frac{[P]_{\text{tot}}[S]K_{ns}}{K_S K_{ns} + K_S [D] + K_{ns} [S]} \quad (5.47)$$

On peut alors évaluer la duplication d'un génome entier suivi par la suppression massive d'ADN qu'accompagne une diminution du volume. Pour illustrer les effets on considère le cas limite où les concentrations du gène et des promoteurs spécifiques sont doublées et la suppression d'ADN conduit à la même quantité d'ADN no spécifique qu'avant duplication :

$$[P]_{\text{tot}} \rightarrow 2 * [P]_{\text{tot}}$$

$$[S] \rightarrow 2 * [S]$$

$$[D] \rightarrow [D]$$

Avec ces changements de paramètres on obtient immédiatement la concentration des facteurs liés spécifiquement après duplication du génome et réduction de l'ADN (WGD+ Δ) :

$$[PS]_{\text{WGD}+\Delta} = \frac{4[P]_{\text{tot}}[S]K_{ns}}{K_S K_{ns} + K_S [D] + 2K_{ns} [S]} \quad (5.48)$$

La figure 5.34 présente le rapport $\frac{[PS]_{\text{WGD}+\Delta}}{[PS]}$ des concentrations de facteurs liés spécifiquement après duplication du génome et réduction sur la concentration avant suppression (identique à celle WGD sans suppression). Pour une liaison non spécifique faible, la concentration de cibles spécifiques liées en cas de WGD + Δ est le double de celle sans WGB (zone bleue). En présence de liaison non spécifique significative, la concentration de sites spécifiques liés se trouver jusqu'à valoir quatre fois celle sans duplications. Il en serait ainsi car la synthèse de facteurs serait double, alors que les sites non spécifiques disponibles pour séquestration seraient toujours de concentration identique. TF₁ et TF₂ sont un exemple de deux facteurs de même concentration globale, mais le premier se lie seulement spécifiquement, et le second possédant des liaisons non spécifiques significatives. Dans le scénario WGD + Δ , le facteur TF₂ serait deux fois plus associés à ses cibles que TF₁. Si les deux facteurs agissaient initialement en relation équilibrée, la duplication suivie d'une suppression massive d'ADN viendrait donc perturber leur équilibre d'action.

Effet de dosage et duplication de génome, conclusion Les idées présentées dans cette publication sont fondées sur des modèles simples d'activité génétique d'effets globaux possibles lors d'événements de duplication de génomes qui favoriseraient le maintien d'ADN non fonctionnel, plutôt que des suppressions massives de celui-ci pour un retour à un génome plus petit. Cette conclusion tient au lien entre taille du génome (en termes de quantité totale d'ADN) et volume nucléaire. Une simple caractéristique géométrique et physique qui impacte globalement les concentrations des protéines. Un phénomène comme la suppression massive d'ADN qui viendrait diminuer le volume se répercute automatiquement à l'ensemble de la dynamique fonctionnelle

du génome. En conclusion, une partie de l'ADN non codant, incluant les répétitions de séquences, pourrait être justifié par la maintenance d'une régulation génétique optimale après une duplication de génome. L'évolution du génome et des séquences non codantes est ainsi liée à la maintenance des volumes nucléaires et cellulaires pour respecter les équilibres de concentrations des protéines compatibles avec le fonctionnement génétique et métabolique.

Une autre forme de polyploïdisation que la duplication de génomes entiers consiste dans la fusion de deux génomes d'organismes différents en cas d'hybridation. On parle dans ce cas de *allopolyploïdisation*, un phénomène surtout présent chez les plantes. Dans ce cas, si la fusion implique des organismes de génomes de tailles différentes (différentes valeurs C) on peut s'attendre à ce que l'augmentation de volume du la quantité totale d'ADN du génome fusionné affectera de manière différente les facteurs et cibles issus de chaque sous-génome. L'exploration de cette idée d'un autre impact global du lien entre ADN, géométrie et fonction fait l'objet de notre article récent [BOTTANI et collab., 2018] discuté plus bas 5.5.4.

5.5.4 Dosage et dominance génétique

L'importance des rapports d'équilibres plutôt que des niveaux d'abondance individuels eux même se retrouve également dans des questions classiques de génétique comme l'origine de la **dominance** et de la **récessivité** des allèles. En effet le nombre de copies d'un gène correspond au nombre de ses allèles présents dans le génome, identiques ou différents. D'autre part, la génétique classique nous a appris que les différents allèles (deux pour les humains qui sont diploïdes) ne se comportent pas de manière identique. Prenons le cas d'un diploïde, dans les combinaisons des deux types d'allèles possibles pour un locus génétique, si les deux allèles sont différents le phénotype peut dépendre intégralement de l'un des deux allèles, dit alors allèle dominant, le second allèle est alors récessif. L'existence d'effets de dosage de gènes, donc de conséquences de différents nombres d'allèles est donc liée à la question classique de génétique des mécanismes et des raisons de la dominance de certains allèles sur d'autres.

De la non linéarité de la dominance

Comme cela avait été remarqué par Reiner Veitia [VEITIA, 2002b], la thématique du dosage de gènes et ses conséquences est étroitement reliée aux propriétés de la fonction de réponse transcriptomique que nous avons discuté précédemment. Le nombre de copies d'un gène sera naturellement un paramètre supplémentaire, probablement multiplicatif, de cette relation quantitative entre le niveau d'expression d'un gène et les concentrations des facteurs de transcriptions qui le régulent.

Typiquement, on l'a vu plus haut, l'activation d'un promoteur implique la synergie ou même la coopérativité de plusieurs facteurs causant des courbes de réponse sigmoïdes, typiquement de la forme de fonctions de Hill. La non linéarité a pour effet d'amplifier les conséquences d'une variation de dosage, voir figure 5.35. Si le gène correspondant au facteur de transcription est l'objet d'un dosage anormal, sa concentration de produit en entrée des cibles régulées sera affectée, du même multiple a priori que la variation de dosage, soit un facteur 0,5 si le gène se retrouve en copie unique. Du fait de cette variation de l'abondance d'un facteur de transcription l'activation d'un promoteur qu'il régule peut basculer de part et d'autre d'un niveau élevé à faible vers une inactivation ou une pleine activation et subir un changement plus radical que celle d'un facteur du 0,5 du régulateur. Cette sensibilité sera d'autant plus grande que le degré de synergie pour l'activation de la transcription est élevé. C'est-à-dire que le facteur de transcription agit sous forme multimérisée dans un complexe de nombreux monomères (dimers, quadrimers, octomères).

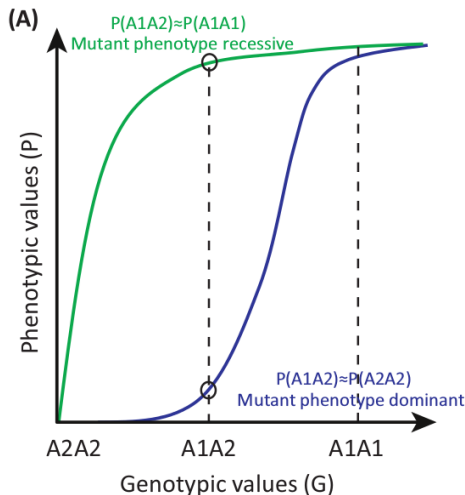


FIGURE 5.35 – **Relations non linéaire entre des valeurs représentant le génotype (G) et le phénotype (P).** Représentation de deux relations non linéaires entre des valeurs représentant le génotype (G) et le phénotype (P) pour illustrer le concept de dominance Mendélienne. La courbe verte représente une fonction hyperbolique reliant G (par exemple une concentration d'enzyme) à P, qui est la forme typique attendue pour la plupart des réactions enzymatique. Dans ce cas le phénotype mutant est récessif, car l'hétérozygote A1A2 sera associé à un phénotype plus proche de l'homozygote original A1A1. Dans un autre système où G et P sont liés par une fonction sigmoïde, en raison de la non-linéarité le phénotype mutant sera plus probablement dominant car la valeur associée à A1A2 sera plus proche de A2A2 que de l'homozygote original A1A1. Comme discuté ailleurs, ces courbes sigmoïde apparaissent en cas de synergie et de coopérativité. Source : [VEITIA, 2003]

Dominance génétique dans des allopolyploïdes : hypothèses et modèles

L'hybridation est le processus de combinaison de deux espèces différentes compatibles par reproduction sexuée. Les individus hybrides peuvent être stériles ou non, conduisant à l'évolution de nouvelles espèces dites hybrides doté d'un matériel génétique initialement la combinaison des deux espèces génitrices. Chez les plantes l'hybridation est très commune et combinée avec la propriété des végétaux d'admettre des degrés de polyploïdies élevés c'est un processus important de création de nouvelles espèces (spéciation) [RIESEBERG, 1997]. Résultat de l'hybridation, l'allopolyploïdie est l'état polyploïde combinant des chromosomes provenant de deux espèces différentes. Une situation donc surtout commune chez les plantes. Dans l'article « Gene expression dominance in allopolyploids : hypothèses and models » [BOTTANI et collab., 2018] avec R. Zabet, JF Wendel et R. Veitia nous étudions comment suite à une hybridation les gènes porté par un sous-génome de l'hybride issus d'un progéniteur peuvent globalement se retrouver dans un état dominant d'expression favorisée par rapport à la copie de l'autre sous-génome.

De nombreuses observations indiquent que lors de la formation de l'hybride les copies des gènes en provenance des deux parents ne contribuent pas de manière égale et additive au phénotype. Plus particulièrement l'une des copies issue d'un progéniteur est exprimée préférentiellement. L'exemple le plus connu concerne les gènes de l'ARN ribosomal dont l'ensemble d'un ancêtre est rendu silencieux. Un phénomène de dominance de gènes issus d'un progéniteur, voir même de tout son génome, semble être caractéristique de nombreux hybrides [STEIGE et SLOTTE, 2016; WENDEL et collab., 2018].

Les mécanismes conduisant à cet effet de dominance d'une copie ou de l'autre, c'est-à-dire de l'expression préférentielle des gènes dupliqués d'un seul génome progéniteur sont encore mal connus. L'explication la plus commune s'appuie sur la distribution différentes dans les génomes d'éléments transposables. La proximité de tels éléments à un gène est supposée faciliter sa répression car la fermeture de la chromatine serait favorisée. Dans l'hybride le génome issu du progéniteur avec la plus grande densité d'élément transposables serait ainsi davantage réprimé [CHENG et collab., 2016; HOLLISTER et GAUT, 2009; RENNY-BYFIELD et collab., 2015]. Bien que cette explication de la dominance d'expression génétique dans les allopolyploïdes est possible, d'autres facteurs peuvent également intervenir.

Dans notre publication nous proposons une explication pour les allopolyploïdes basée sur les différences d'équilibres de liaisons entre sous génomes combinés de tailles différentes et ayant évolué pour des propriétés chimiques différentes (affinités, concentrations un niveau de liaison etc...)

Modèle simple de dominance dans un hybride Comme d'habitude on considère que la transcription est proportionnelle à la fraction d'occupation d'un promoteur et la fonction de réponse transcriptionnelle est donc représentée par la fonction de Hill :

$$TR = \frac{[TF]^n}{K^n + [TF]^n} \quad (5.49)$$

Ici [TF] représente la concentration libre de facteur de transcription que l'on suppose en excès par rapport à la concentration de promoteur de telle manière que les concentrations libres et liées sont pratiquement équivalentes. L'exposant n est donc lié au nombre de sites de liaison dans le promoteur et K la constante de dissociation microscopique correspond à la concentration de facteur de transcription pour 50% de l'occupation maximale.

Ceci signifie qu'il existe un domaine dynamique de réponse utile pour la régulation autour de $[TF]_{TR0,5} = K$ permettant le contrôle des gènes cibles. On va considérer un gène homologue dans les deux génomes initiaux des espèces progénitrices d'un hybride. Ce gène est régulé de la même manière par la fonction de réponse TR précédente. Cependant dans chaque organisme l'évolution a dû ajuster pour ce même gène la valeur utile de concentration de son facteur de transcription ou bien également la valeur de la constante K sur le promoteur en fonction de l'environnement génétique et cellulaire propre à chaque organisme. Avant l'hybridation, les facteurs de transcriptions homologues TF_1 et TF_2 qui régulent le gène cible homologue sont exprimés dans chaque progéniteur à des niveaux permettant la régulation adéquate.

Après l'hybridation, les concentrations $[TF_1]$ et $[TF_2]$ peuvent être différentes que dans les parents. En outre les promoteurs cibles de ces facteurs sont dans le contexte génomique de l'hybride différent. Régulateurs d'un même gène homologue, les facteurs TF_1 et TF_2 le sont aussi et contribuent à une même concentration du facteur régulant chaque copie de la cible issue de chaque progéniteur. La nouvelle concentration du facteur $TF (= TF_1 = TF_2)$ dans l'hybride pourra suffire pour activer le promoteur avec le plus petit seuil d'activation (celui issu du progéniteur avec $[TF]_{TR0,5}$ le plus bas), l'autre promoteur sera moins actif, voir silencieux. Ainsi le sous-génome portant les seuils d'activation mes plus bas sera plus probablement le plus actif après l'hybridation, voir figure 5.36.

Modèle plus sophistiqué de dominance génétique après hybridation Le modèle précédent suppose une différence des constantes K dans les progéniteurs. Pour contextualiser mieux les différences de fonctionnement de deux gènes homologues dans les espèces initiales puis dans le génome fusionné on propose un modèle plus avancé qui tient compte des tailles respectives des deux génomes initiaux affectant les liaisons non spécifiques des facteurs de transcription. Un facteur de transcription peut en effet être impliqué dans deux types d'association : spécifique avec un rôle fonctionnel sur des sites de régions ci-s-régulatrices de gènes cibles et non spécifiques à tout autre endroit de l'ADN, sans impact transcriptionnel.

Plus spécifiquement nous supposons deux génomes qui vont fusionner de tailles différentes. Deux cibles homologues G_1 et G_2 dans chaque génome sont régulé par de facteurs de transcription paralogues TF_1 et TF_2 avec des affinités différentes (TF_i régule le gène G_i du sous-génome i avec une énergie de liaisons E_i). Deux scénarios sont possibles, soit avec régulation croisée, TF_1 régule aussi le gène G_2 dans l'hybride (et respectivement TF_2 peut agir sur G_1), soit avec une régulation uniquement spécifique. On considère aussi pour simplicité $V_3 = V_1 + V_2$ ainsi que des promoteurs avec un seul site de liaisons.

La fraction d'occupation du promoteur se calcule comme :

$$f_i = \frac{[TF_i]e^{-E_i/k_B T}}{[TF_i]e^{-E_i/k_B T} + [NS_i]e^{-E_{NS}/k_B T}} = \frac{[TF_i]}{[TF_i] + [NS_i]K_i} \quad (5.50)$$

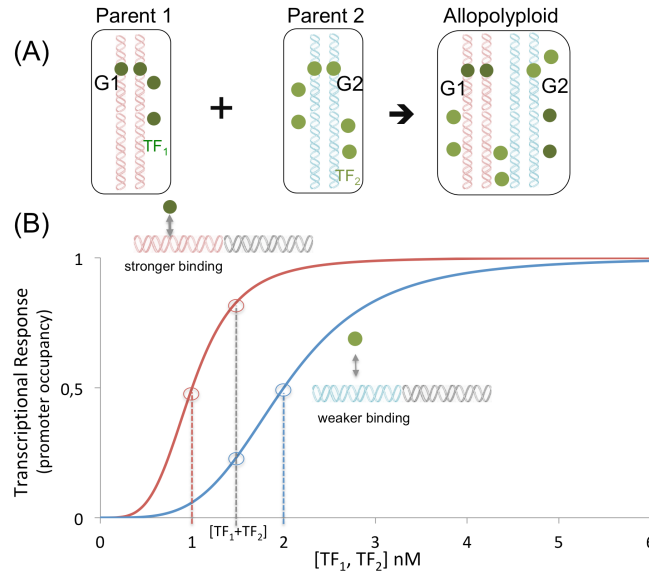


FIGURE 5.36 – **Modèle simple de dominance génomique lors d'hybridation.** A) effet d'une hybridation (allopolyploïdisation) impliquant deux cellules génitrices 1 et 2. On suit la transcription de deux promoteur homologues G_1 et G_2 répondant à des activateurs TF_1 (ou TF_2 , l'homologue). Les cercles verts représentent TF_1 ou TF_2 , qui ont des affinités différentes pour leurs séquences cis-régulatrices. B) On suppose que les cellules parents initiales ont le même volume et que $K_1 = 1$ nM, $K_2 = 2$ nM et que le niveau de transcription initiales des parents TR vaut un demi. Juste après la fusion de l'hybridation on suppose un volume cellulaire double. Dans ce cas le génome fusionné hybride conduira à des fonctions de réponse TR dans lequel les allèles du progéniteur 1 seront exprimés à 80% de la saturation, alors que ceux du progéniteur 2 fonctionneront à 25%, c'est-à-dire les parent 1 sera dominant.

où $[NS_i]$ est la concentration des sites non spécifiques du génome i dans la cellule initiale. Les constantes de dissociation K_i peuvent être différentes pour les deux facteurs paralogues.

Comme les gènes cibles dans les génomes qui vont fusionner sont homologues et probablement avec les mêmes rôles fonctionnels, on va supposer pour simplicité des occupations fractionnées similaires des promoteurs homologues avant la fusion $f_1 = f_2 = f$. Les concentrations de facteurs correspondant pour cette valeur de l'occupation peuvent être déduites de l'équation précédente en terme de nombre de molécules TF_i :

$$[TF_i] = \frac{[NS_i]K_iP}{1 - P}$$

Dans l'allopolyploïde, on doit considérer un volume plus large et la possibilité de régulation croisée. Le facteur TF_1 peut se lier dans ce cas avec une énergie E_{12} sur la cible paralogue du second génome (et symétriquement TF_2 avec l'énergie E_{21} sur le premier).

Les occupations fractionnées seront données par

$$f'_1 = \frac{[TF_1]'e^{-E_1/k_B T} + [TF_2]'e^{-E_{21}/k_B T}}{[TF_1]'e^{-E_1/k_B T} + [TF_2]'e^{-E_{21}/k_B T} + [NS]e^{-E_{NS}/k_B T}}$$

(et le symétrique pour f'_2).

Pour décrire la régulation croisée on peut introduire un coefficient C_1 , tel que $e^{-E_{21}/k_B T} = C_1 e^{-E_1/k_B T}$. Ce coefficient vaut $C_1 = 1$ pour une régulation croisée pleine et $C_1 = 0$ en absence. De cette manière on a

$$f'_i = \frac{[TF_i]' + [TF_j]'C_i}{[TF_i]' + [TF_j]'C_i + [NS]K_i}$$

avec $j = 2$ si $i = 1$, et le symétrique pour $i \leftrightarrow j$. Les concentrations sont calculées dans le volume V_3 : $[TF_i] = \frac{TF_i}{V_3}$ et $[NS] = (NS_1 + NS_2)/V_3$.

On peut maintenant exprimer f'_1 et f'_2 en fonction de la probabilité d'occupation f du promoteur dans les progéniteurs, qui amène à :

$$f'_1 = \frac{1 + \frac{C_1 NS_2 K_2}{NS_1 K_1}}{1 + \frac{C_1 NS_2 K_2}{NS_1 K_1} + (1 + \frac{C_1 NS_2}{NS_1}) \frac{1-f}{f}}$$

et le symétrique pour f'_2 .

On peut alors considérer trois cas : A) dans les génomes originaux la probabilité d'occupation des promoteurs homologues *et* les concentrations de facteurs sont identiques ($f = f_1 = f_2$ et $[TF_1] = [TF_2]$), une situation qui résulte de la co-adaptation de chaque constante de dissociation et la taille de son génome non spécifique. B) les constantes de dissociation sont identiques et les concentrations des facteurs TF_1 et TF_2 sont alors proportionnelles aux tailles des génomes respectifs, ceci pourrait être le cas de facteurs très conservés se liant à des séquences de promoteurs très conservés et implique alors nécessairement une pleine régulation croisée. C) un mélange des deux situations précédentes.

Dans le cas A) la condition $[TF_1] = [TF_2]$ amène à la relation suivante entre les constantes de dissociations et les tailles des génomes : $NS_1 \times K_1 = NS_2 \times K_2$ d'où vient

$$f'_1 = \frac{1 + C_1}{1 + C_1 + (1 + \frac{K_1}{K_2}) \frac{1-f}{f}}$$

et l'équivalent pour f'_2 . Dans ce cas il peut y avoir ou non régulation croisée, le facteur de transcription de chaque génome capable de se lier ou non sur la cible de l'autre génome.

La figure 5.37 issue de ce modèle montre systématiquement la domination dans l'allopolyploïde des promoteurs issus du génome le plus grand. La figure illustre la situation pour $N_2 = 5xN_1$ mais le résultat reste valable pour des facteurs plus faibles.

Dans le cas B) les promoteurs cibles dans l'allopolyploïde sont occupés de la même manière indistinctement et il ne peut y avoir de dominance. Enfin le cas C) considère une situation intermédiaire où $K_2 = \beta K_1$ avec $0 \leq \beta \leq \frac{N_1}{N_2}$. L'analyse de cette extension confirme que même dans cette situation moins extrême l'expression des gènes issus du génome de taille la plus grande est privilégiée.

Le scénario décrit correspond à la situation juste après l'hybridation. Dans les premiers stades de l'hybride les facteurs de transcription tels que pris en compte dans le modèle seront eux-même éventuellement sujet à un effet de dominance car leurs propres gènes sont des cibles de facteurs de transcription. Ainsi par un effet de cascade on s'attend à une rétroaction positive qui renforce une situation initiale et la domination d'un des gènes homologues.

Une autre situation de dosage génétique : la compensation de dosage entre gènes du chromosome X et les autosomes.

L'équilibre nécessaire de dosage se manifeste lors du développement notamment dans le contrôle chez les mammifères de l'expression du gonosome X. L'article « X chromosome inactivation and active X upregulation in therian mammals : facts, questions, and hypotheses. » avec Reiner Veitia, F. Veyrunes F et J Birchler [VEITIA et collab., 2015] traite de la problématique du dosage génétique

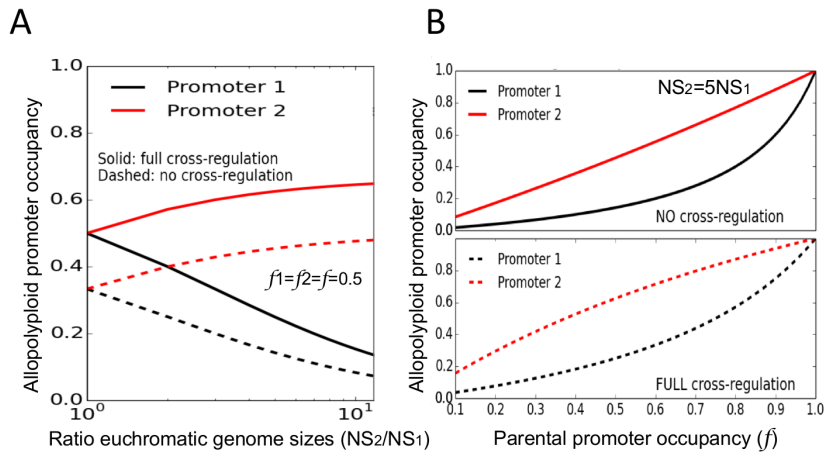


FIGURE 5.37 – **Conséquences sur l’occupation d’un promoteur suite à l’hybridation de deux organismes.** A) probabilité d’occupation du promoteur dans l’allopolyploïde en fonction du rapport des génomes progéniteurs pour une probabilité d’occupation identique des promoteurs avant la fusion identique $f = f_1 = f_2 = 0,5$. Dans les deux cas avec et sans régulation croisée c’est toujours le promoteur issu du génome le plus grand qui domine. B) probabilité d’occupation du promoteur dans l’allopolyploïde en fonction de l’occupation du génome progéniteur (supposé identique dans les deux) pour $NS_2=5 \times NS_1$. Panneau du haut ; pas de régulation croisée. Panneau du bas : régulation croisée, . La différence entre les occupations des promoteurs homologues dans l’allopolyploïde est maximale pour une valeur particulière f . Ici aussi, avec ou sans régulation croisée c’est le promoteur issu du génome qui domine.

lié aux chromosomes sexuels des mammifères euthériens. Les femelles possèdent deux copies du chromosome X et les mâles une seule copie en plus du chromosome Y dont ils sont seuls porteurs. La question du dosage concerne les niveaux d’expression des gènes du chromosome X entre mâles et femelles.

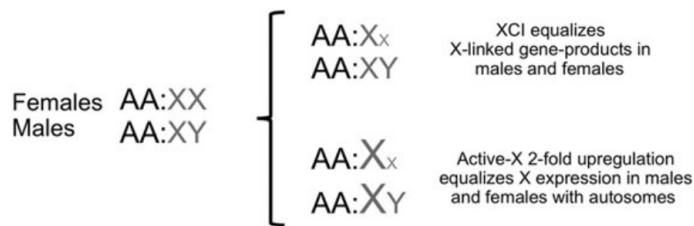


FIGURE 5.38 – **Mécanismes de compensation de dosage du chromosome X** : inactivation d’un chromosome X (femelle) ou sur-régulation active du chromosome X (mâle). L’inactivation d’un chromosome X (XCI) chez la femelle égalise l’expression des gènes liés au chromosome X entre femelles et mâles euthériens. La sur-régulation (Active Xa) des gènes de l’exemplaire chez les mâle permettrait aussi d’équilibrer les niveaux entre les autosomes et les produits des gènes X.

Des observations de nature différentes montrent en effet l’existence de mécanismes pour équilibrer l’expression des gènes de ce chromosome entre mâles et femelles qui en possèdent deux copies, figure 5.38. Ainsi chez la Drosophile l’absence d’un chromosome X chez le mâle est compensée par des gènes sur le seul X présent deux fois plus actifs que chacune des copies sur les deux chromosomes X de la femelle. Chez les mammifères, la compensation a lieu chez les femelles avec l’inactivation globale d’une des deux copies de l’autosome X aboutissant des gènes du X exprimés de la même manière dans les deux sexes, figure 5.38. Alternativement la compensation peut impliquer la sur-expression des gènes du chromosome X mâle. La présence de ces systèmes de compensation est un signe de la nécessité physiologique du respect des équilibres de dosage.

Alterations épigénétiques et déséquilibres de dosages : mécanismes du vieillissement.

Dans l'article « Aging : Somatic Mutations, Epigenetic Drift and Gene Dosage Imbalance » [VEITIA et collab., 2017] avec R. Veitia, DR, Govindaraju et JA Birchler nous illustrons comment la question du dosage génétique et du respect de l'équilibre de dosage intervient dans le contexte du vieillissement. Nous présentons un modèle qui relie des changements au niveau génomique durant la vie d'un individu aux conséquences liées à l'âge sur phénotype.

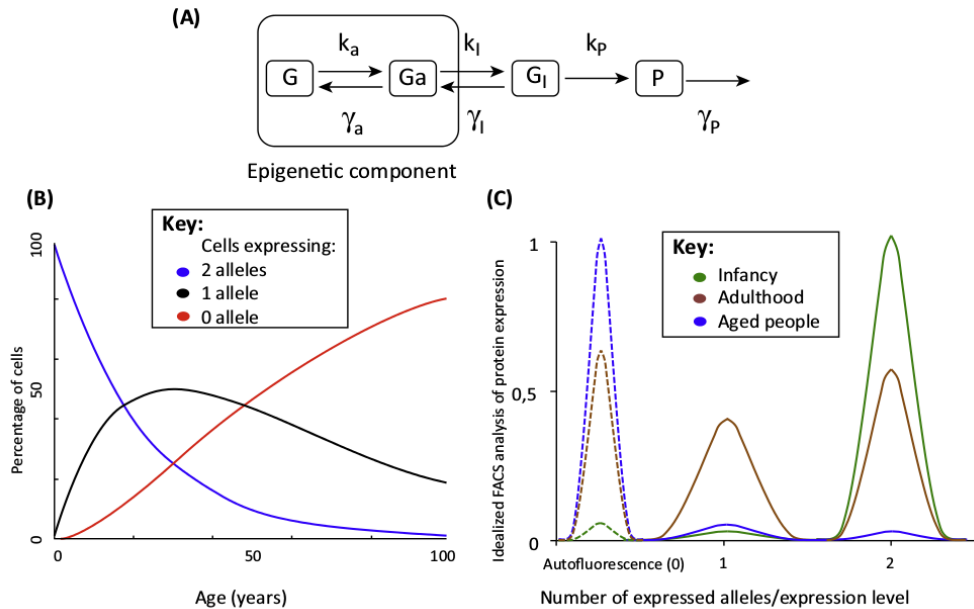


FIGURE 5.39 – **Expression de gène avec décroissance exponentielle d'accessibilité de la chromatine.** A) modèle où G indique le gène d'intérêt dans un état inaccessible en raison, par exemple d'ADN méthylé ou de parqueurs répressifs pour des histones; G_a représente le gène dans un état accessible et G_i est le gène induit pour la transcription; k_a et γ_a contrôlent la décroissance exponentielle de l'accessibilité; k_i et γ_i déterminent la transcription du gène dans la chromatine ouverte; k_p et γ_p sont enfin les taux de production et de dégradation de la protéine. B) Représentation du pourcentage de cellules dans la population exprimant deux, un seul ou aucun allèle en fonction de l'âge de l'individu. On suppose que les proportions des sous populations de chaque type n'affecte pas leur prolifération. Le taux d'inactivation d'un allèle a été fixé à 1% des cellules avec deux allèles pour un âge de 100 ans. C) Idéalisations d'un profil FACS représentant l'évolution de l'abondance de protéines avec le temps pour trois âges et d'après l'évolution en B). Source : [VEITIA et collab., 2017].

Le processus de vieillissement est une propriété universelle de la plupart des organismes se traduisant par une diminution progressive et souvent irréversible de fonctions physiologiques et reproductives. L'accumulation dans le temps de dommages cellulaires semble être la cause du vieillissement mais les mécanismes fondamentaux restent mal compris. Les changements liés à l'âge sur le génome amènent à des dérégulations de l'accessibilité des gènes et de leur expression. L'accumulation de telles altérations dans les cellules somatiques avec l'âge conduit à une diminution des fonctions cellulaires que manifestent les effets du vieillissement.

De nombreuses altérations du génome de cellules somatiques peuvent intervenir durant la vie, parmi lesquelles des variations de nucléotides individuelles (single nucleotide variants –SNVs–), des variations du nombre de copies de gènes (copy number variants –CNVs–), de l'aneuploïdie de chromosomes, et des changements épigénétiques. Ces changements viennent d'erreurs dans la réplication de l'ADN ou dans les processus de réparation de dommages sur l'ADN ainsi que modifications épigénétiques. Les modifications épigénétiques liées à l'âge font intervenir des changements des patrons de méthylation de l'ADN, des modifications des histones et plus largement le remodelage de la chromatine. Une des conséquences majeures des changements épigénétiques est leur transmission lors des divisions cellulaires. Ces processus d'accumulation de mutations dé-

létaires dans des cellules somatiques correspondent au scénario bien connu du cliquet de Muller (Muller's ratchet) par lequel un génome accumule des mutations de manière irréversible [LYNCH, 2016; MULLER, 1964].

Dans notre article [VEITIA et collab., 2007] nous proposons notamment un modèle de dérive épigénétique de l'inactivation de gènes par un processus stochastique d'extinction. Dans ce scénario de type « utilise ou perd le » (use it or lose it) un gène induit sollicité pour être exprimé est protégé de modification épigénétiques. Mais durant une absence prolongée d'induction de la transcription, le gène non induit mais toujours accessible, risque à chaque instant de perdre son accessibilité. Ceci accroît la probabilité de l'état inaccessible G manifestant ainsi une extinction dans le temps de l'expression génétique, voir figure 5.39. Ce changement est stochastique et indépendant pour chaque gène et chaque allèle. Avant même les conséquences d'une extinction complète l'épigénétique se manifestera par des manifestations de déséquilibre de dosage.

5.5.5 Conclusion dosage génétique

Reprenant une conclusion de James Brichler [CONANT et collab., 2014], comme le suggèrent les différents mécanismes abordés dans les études ci-dessus, l'équilibre de dosage n'est pas une simple force agissant sur le génome et dont la présence ou absence peut-être détectée. C'est un phénomène à multiples facettes se manifestant lors de multiples événements de plasticité génomique, allant de la duplications du génome entier, dans le cas d'aneuploïdie, de duplications partielles de segments chromosomiques, de répétitions tandem, de simples variations du nombre de copies de gènes ou encore d'extinctions épigénétiques de gènes.

Les modèles, techniquement simples, avec lesquels j'ai pu contribuer dans ce contexte suggèrent comment des relations générales mathématiques et physiques de rapports de nombres, de volumes, de vitesses influencent des propriétés génétiques, sans devoir faire appel à des mécanismes moléculaires spécifiques.

5.6 Variabilité phénotypique et relation génotype/phénotype

« To understand the genetic basis of disease will require not just mapping these genes but an understanding of how the phenotype is created in the first place and the messy interactions between genetic variation and environmental variation. »

Marc Kirschner, « The meaning of systems biology », Cell 121 :503-4 2005.

Les travaux précédents concernaient l'organisation des circuits de régulation génétiques ainsi que du fonctionnement dynamique ou d'équilibre manifestant l'activité des réseaux. On a vu une analyse de l'organisation des réseaux transcriptionnels et d'interaction protéiques, des modélisations mathématiques déterministes et stochastiques d'un exemple de circuit pour des oscillations génétiques. Les comportements non linéaires relevés dans ce type d'étude et traités par l'analyse mathématique sont déterminants dans l'activité des gènes et protéines en interaction, comme tout particulièrement les effets quantitatifs du dosage génétique que nous venons de discuter. En dehors de l'étude de l'organisation des réseaux, ces thèmes aux conséquences nombreuses pour le vivant et les biotechnologies restent cependant encore de nature relativement mécanique.

Le travail que j'introduis ici est une activité de longue haleine résultant d'une collaboration avec Suzannah Rutherford du Fred Hutchinson Cancer Research Center, Seattle basé dont la publication visant un journal de premier plan est encore en cours. Il s'agit d'un ensemble de données génomiques original et de son analyse concernant des questions fondamentales de variabilité phénotypique et de leur relation avec le génotype. Un projet que j'ai rejoint après sa phase de construction génétique et d'acquisition de données pour en faire l'analyse statistique et contri-

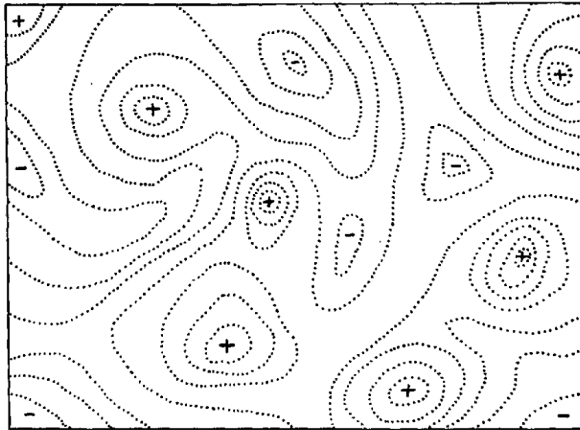


FIGURE 2.—Diagrammatic representation of the field of gene combinations in two dimensions instead of many thousands. Dotted lines represent contours with respect to adaptiveness.

FIGURE 5.40 – **Première représentation conceptuelle d'un paysage adaptatif** [WRIGHT, 1932]. Dans un espace virtuel à deux loci génétiques, chacun suivant les variations alléliques contribuant différemment à la valeur adaptative. Des combinaisons particulières correspondent à des maxima locaux de la valeur adaptative. Suivant la régularité des contours les espèces vont évoluer plus ou moins brusquement dans le paysage sous l'effet de mutations aléatoires. Source : [WRIGHT, 1932];

buer à la modélisation. Le projet de Suzannah Rutherford consistait dans la production d'une ressource biologique de cinquante souches de levures *S. cerevisiae* issues d'habitats différents et modifiées pour permettre le suivi de leur réponse à un stress (choisi pour être la réponse au stress hyper-osmotique). Cette variété d'organismes génétiquement très proches doté d'un système rapporteur sur une voie de signalisation critique permet l'exploration des relations générales Génomique vers Phénotype (G→P) au niveau de complexité de la cellule de levure.

Introduction à la problématique

Outre la compréhension des mécanismes de système de régulation et circuits biochimiques et génétique on voudrait en effet aussi progresser dans la compréhension des principes plus globaux de génétique abordant la question de la relation entre l'information héréditaire dans la séquence d'ADN des organismes vivants et sa transformation dans un organisme biologique vivant dotée de traits particuliers avec une certaine valeur adaptative. À l'échelle plus large incluant l'évolution naturelle des espèces la question la plus pertinente est d'ailleurs la relation entre un génotype et la valeur adaptative de son phénotype. Une relation transitive du génotype au phénotype, et du phénotype à la valeur adaptative. Les propriétés générales de ces relations, au sens mathématique, de monotonie ou discontinuité, de non-linéarité, dérivabilité affectent directement le déroulement de l'évolution naturelle.

Via les relations $G \rightarrow P$ les différences dans l'information (au niveau du génotype G) donneront en effet des différences de phénotypes avec en conséquences des différences dans les valeurs adaptatives des individus. Dans la mesure où les phénotypes sont reliés à la valeur adaptative, les propriétés de la relation $G \rightarrow P$ affectent les histoires évolutives possibles soumises à la sélection naturelle. Elles affectent la forme du paysage adaptatif d'une espèce qui représente métaphoriquement la valeur adaptative, c'est-à-dire le taux de réplication, associée à tout génotype possible par l'entremise du phénotype [WRIGHT, 1932]. Définissant une topographie abstraite, sur un espace à très haute dimension, le paysage adaptatif permet de représenter un chemin évolutif d'une population vers des niveaux plus élevés d'adaptation en fonction des changements aléatoires des génotypes. Cette métaphore permet de se représenter des contraintes et des facilitations à l'évolution en fonction de vallées, de cols, d'aspérités que la valeur adaptative de l'espèce rencontrerait quand les génotypes présents dans une population de ses individus varient. Des différences héréditaires de valeur adaptative au sein d'une population et leur position sur le paysage adaptatif définissent ainsi les taux et directions des changements évolutifs.

La biologie de l'évolution explique comment la sélection naturelle et donc l'adaptation des espèces dans un environnement n'agit qu'en présence de différences de valeurs adaptatives [ORR, 2009]. Valeurs qui manifestent le succès reproductif d'un phénotype dans un contexte donné et

sont fonction du génotype de l'organisme. Le processus de l'évolution naturelle suppose que des différences d'information codées au niveau génétique se traduisent en des différences de valeur adaptative entre individus, c'est-à-dire de différences dans les taux de reproduction des individus porteurs de variantes de génotype. La vitesse et l'étendue des changements évolutifs dépendront du rapport entre stabilité des traits hérités et leur variabilité. Pour une population d'individus, dont les génotypes sont distribués dans un sous ensemble de l'espace des génotypes possibles, quelles sont les variations attendues de la valeur adaptative et en conséquence le sens et la vitesse du changement évolutif de cette population ?

Alberch a introduit le concept d'une relation du génotype vers le phénotype (relation $G \rightarrow P$) en 1991 en contraste de la vision standard d'un plan génétique (genetic blueprint) déterminant les organismes. Dans sa vision les gènes ne spécifient pas le développement et la forme d'un organisme, ce sont des facteurs critiques, mais parmi d'autres, qui déterminent le phénotype via le développement. Le processus de croissance de l'organisme affecté par les gènes, mais qui durant son déroulement affecte à son tour l'expression génétique. Comme le rappelle [PIGLIUCCI, 2010], plutôt qu'un plan de construction (blueprint) ALBERCH [1991] propose la notion d'une relation non bijective $G \rightarrow P$ entre un espace des paramètres en rapport avec le développement et liés à l'expression génétique et des manifestations observables pouvant être de nature différentes. Les propriétés de l'application pour une espèce déterminent ses caractéristiques évolutives. Ainsi la relation $G \rightarrow P$ est approximativement injective, un ensemble de paramètres génotypiques donnant un phénotype particulier. L'étendue de l'injection, qui correspond à la surface des paramètres transformée dans un même phénotype, détermine la stabilité de ce phénotype face à des variations et mutations possibles des paramètres génétiques de développement. Les zones de stabilité phénotypiques admettent des frontières de transformation sur lesquelles de petits changements des paramètres génétiques de développement causent la transition d'un état du phénotype à un autre. Pour une population, sa stabilité phénotypique sera déterminée par la zone de l'espace des paramètres qu'elle occupe et la proximité avec des frontières de transformation.

Malheureusement les structures de paysages adaptatifs sont inconnues. Hors d'expériences numériques [FORTUNA et collab., 2017] il ne semble pas y avoir d'exemples vivants pour lesquels de tels paysages soient caractérisés. La notion même de pays de carte ou paysage adaptatif reste largement une métaphore pour aider la compréhension de l'évolution naturelle. En effet, d'après PIGLIUCCI [2013] établir un paysage adaptatif complet est une opération jamais complètement entreprise car nécessitant de caractériser à la fois la relation du génotype au phénotype (en lien donc avec la biologie du développement) puis du phénotype à la valeur adaptative pour un organisme donné en explorant un ensemble suffisamment varié de génotypes possibles. Seuls quelques cas particuliers sont accessibles que Pigliucci passe en revue [PIGLIUCCI, 2010]. Le cas le mieux caractérisé est celui du repliement de l'ARN qui constitue l'exemple le plus simple et peut-être rappelant une situation primordiale. Les fonctions des protéines sont le second exemple bien caractérisé car accessible à des manipulations expérimentales directes. Puis quelques situations partielles ont pu être caractérisées au niveau de réseaux de régulation génétique.

Les cartes de génotype-phénotype ont été étudiées dans de multiples systèmes biologiques récemment analysés par FORTUNA et collab. [2017] comme des structures de protéines et de molécules d'ARN, le métabolisme et des circuits de régulation génétiques. Ces systèmes très différents semblent cependant partager un certain nombre de points communs dont notamment une robustesse dans une certaine mesure à des changements génotypiques. Ainsi que la possibilité d'explorer graduellement par des mutations ponctuelles des variations de phénotype.

Les déterminations sur des données expérimentales de cartes $G \rightarrow P$ au niveau d'un organisme restent très limitées en raison de la difficulté à disposer de panels variés de génotypes avec leurs phénotypes complexes caractérisés. Les relations $G \rightarrow P$ et paysages adaptatifs demeurent donc essentiellement du domaine de considérations théoriques en biologie de l'évolution [ALBERCH, 1991; CATALÁN et collab., 2017; LEWONTIN, 1974; PIGLIUCCI, 2013; WRIGHT, 1932]. En outre on ne pourra au mieux connaître qu'une petite partie d'un tel paysage étant donné l'immensité de la

diversité génétique possible. Les populations existantes ne représentant qu'une infime proportion du champ de toutes les combinaisons et mutations génétiques [GAVRILETS, 1997; WRIGHT, 1932]. Des approches théoriques basées sur des systèmes évolutifs numériques simulés sur ordinateur permettent également d'explorer des relations génotype-phénotype abstraites, inspirer l'étude des cas naturels et rechercher des similitudes. Si les structures des paysages adaptatifs sont inconnues, on peut cependant espérer en déduire des propriétés à partir de celles des relations $G \rightarrow P$ qui en sont une composante et peut-être plus directe à caractériser. Le nombre limité de souches de laboratoires et l'environnement génétique limité qui en découle pour la plupart des études des déterminants génétiques de la physiologie favorise une vue étroite et probablement faussement déterministe des fonctions et circuits génétiques qui ne peut rendre compte de la variabilité des génotypes et des environnements naturels réels. Une limitation qui se manifeste aussi dans les difficultés à prédire la sensibilité de traitements avec de nouvelles molécules.

Le projet initié par Suzannah Rutherford que j'ai rejoint pour l'analyse des données et leur modélisation visait à s'appuyer sur des technologies génomiques récentes pour contribuer à déterminer des variations de phénotypes fonctionnels dans des populations se reliant à des variations naturelles de génotype.

5.6.1 Notre projet de caractérisation de la variabilité phénotypique et de la relation $G \rightarrow P$

Le projet a pris comme objet d'étude pour les variations phénotypiques la réponse au stress hyper-osmotique de cinquante génotypes de la levure bourgeonnante (*Saccharomyces*). Cette collection a été construite à partir de populations isolées d'habitats différents, sauvages ou industriels, et modifiées pour permettre l'observation d'une réponse génétique à une stimulation contrôlable par l'expérimentateur. Pour cela les prélèvements d'origine de cellules haploïdes ont été accouplés avec une souche de laboratoire porteuse du gène *GPD1* en sortie de la voie de signalisation *HOG* (High Osmotic Stress) fusionné avec le rapporteur fluorescent *GFP*. La bibliothèque est ainsi constituée de cellules hybrides de diploïdes synthétiques implémentant des variantes de génotypes et permettant la mesure d'activité génétique en réponse à des conditions contrôlables. Le choix de la voie de réponse au stress hyper-osmotique comme système génétique contrôlable, vient de deux raisons. L'une biologique, l'importance de ce système intervenant lors de chocs hydriques, communs dans de nombreux habitats, pour la survie des cellules. Donc un système sur lequel l'adaptation naturelle a dû jouer significativement. L'autre raison est pratique : cette voie est relativement facile à contrôler en laboratoire par l'application simple de concentrations salines différentes dans le milieu de culture.

La construction de cette collection de souches ouvre la voie à des explorations de la variabilité phénotypique et d'approcher expérimentalement des propriétés de la relation entre génotypes et phénotypes. D'où on peut espérer déduire des indications sur le paysage évolutif dans le cas de cet organisme modèle standard.

J'ai eu l'opportunité de rejoindre ce projet pour contribuer à l'analyse statistique des résultats et leur interprétation théorique me permettant de compléter mes intérêts dans la biologie des systèmes avec un travail sur des données d'expérience directement lié à des interrogations sur l'évolution. Le travail auquel j'ai participé et illustré ici est une première application de cette ressource, établissant une preuve de principe de la démarche et de son intérêt pour ce questionnement. Les expériences sur cette collection menées par Suzannah Rutherford jusqu'à présent relient les réponses transcriptionnelles des 50 variantes de génotype exposés à des stress hyper-osmotiques aux variations de leur adaptation au stress et à leur survie. Bien que à ces stades les variations entre les génotypes nous sont encore inconnues (les souches n'ont pas été séquencées) nous avons déjà pour ces 50 variants la caractérisation de 50 phénotypes à un niveau de détails dévoilant leur dynamique d'expression génétique d'adaptation au milieu imposé, ainsi que le résultat en termes de succès de l'adaptation ou de stratégie comportementale à long terme pour chaque souche. Ces ré-

sultats sont une première caractérisation de la variabilité phénotypique que l'on voudrait observer dans l'étude de relations $G \rightarrow P$.

Je présenterai ici seulement un résultat particulièrement significatif de ce travail dont une version préliminaire est disponible sous forme de préprint [HIRATE et collab., 2016]. La figure 5.43 présente le résultat de clustering des profils temporels des distributions de l'expression du gène $GPD1 : :GFP$ marqueur de la réponse transcriptionnelle à un stress hyper-osmotique moyen appliqué à un instant initial pour les 50 souches de la collection dans des conditions après une transition diauxique.

Hétérogénéités de réponses transcriptionnelles au sein de chaque population La distribution des niveaux d'expression au sein des cellules individuelles d'une population de 10000 individus a été mesurée par FACS à 7 intervalles de temps 20, 48, 72, 96, 120, 144, 168 heures suivant l'application du stress à $t = 0$ heure. Chaque condition a fait l'objet de 4 à 15 répliques, ce qui fournit un ensemble de données étendu sur l'évolution dans le temps de réponse transcriptionnelle au stress des cellules individuelles au sein des différentes populations. Un exemple représentatif pour chaque souche, de l'évolution temporelle de ces distributions est donné dans la colonne de droite de la figure 5.43 : chaque distribution représentée est un exemple caractéristique parmi les 4 à 15 répliques disponibles suivant les cas.

On observe différentes trajectoires d'accumulation de distributions de $GPD1 : :GFP$ multimodales qui ne sont pas bien décrites par une expression moyenne alors même que les profils sont bien reproductibles pour chaque souche et instant (répliques non montrées). Ces différentes trajectoires représentent des différences dans l'activation de la voie HOG et l'induction de gènes de réponse au stress comme $GPD1$. Ni la fluorescence totale $GPD1 : :GFP$ ni les taux de changement de la fluorescence corréleront bien avec la viabilité. Il en est de même pour toutes les composantes de la décomposition des distributions en 4 Gaussiennes : aucune des 4 Gaussiennes ($G0$ - $G3$) numérotés suivant leur niveau croissant de fluorescence moyenne (voir figure 5.41), à aucun instant durant le stress ne prédit le niveau de survie d'une souche.

Variation continue des types de réponses Les formes particulières, et hérissables, des distributions sont reliées aux (prédisent les) différences de viabilité entre les différentes souches qui ne se réduisent pas à des caractéristiques simples (expressions moyenne, expression d'une ou plusieurs composantes, Gaussienne, des distributions multimodales). Pour rechercher des types de relations entre la réponse transcriptionnelle au stress et la viabilité nous avons procédé à des regroupements statistiques des trajectoires dans le temps des distributions d'expression de chaque souche sans biais éventuel sur la forme et les composantes Gaussiennes éventuelles des distributions.

En fonction de leurs comportements spécifiques (hérissable) de signalisation HOG, les 50 souches se répartissent entre six groupes de comportement moyens, figure 5.43 à gauche. Le regroupement (clustering) qui apparaît est un groupement moyen d'histoires reconstituées de l'évolution des distributions de $GPD1 : :GFP$ sur les 8 temps mesurés. Des historiques différents de l'évolution des distributions sont reconstitués par des permutations aléatoires entre les répliques disponibles à chaque instant. Chacune de ces permutations pour les cinquante souches est ensuite utilisée pour calculer une matrice des distances moyennes (suivant une certaine métrique), entre les évolutions des distributions de toutes les souches. Le clustering présenté est ensuite issu de cette matrice de distances.

Les souches ont été ensuite ordonnées au sein de chaque groupe moyen, et les groupes entre eux, en fonction de la force de leur assignation au sein de chaque groupe dans le calcul de regroupement avec chaque permutation de répliques. (Les souches apparaissant dans le regroupement de la figure 5.43 aux extrêmes de chaque groupe moyen ont tendance à se retrouver associées avec les groupes voisins au-dessus ou en dessous dans des réalisations individuelles de regroupement

avec des permutations données de répliques). Cette analyses numériquement exigeantes sur un volume de données originales important donnent un image inégalée de la variabilité sur des génotypes différents mais proches de la variation d'expression transcriptionnelle au sein d'une population et de son évolution au cours du temps. Cette variation d'activité génétique en bout de la voie de signalisation HOG est ensuite mise en rapport avec l'adaptation et la survie de chaque souche, la viabilité ayant été mesurée par exclusion de iodure de propidium durant les mesures par cytométrie de flux.

Ces regroupements et leur classement font apparaître des types de comportements différents. Au sein de chaque population isogénique (une même souche) on voit apparaître une distribution de vitesses des réponses de la signalisation du stress hyper-osmotique : les cellules, de génotype identique, ne répondent pas toutes de la même manière à la même vitesse. Deux types de comportement semblent composer et être présents dans chaque population. On retrouve des cellules « passives » de faible activation de la voie HOG; alors que d'autres cellules « réactive » réagissent rapidement en exprimant fortement GPD1. Les différentes souches montrent des proportions différentes, et héréditaires (reproductibles et propres à chaque souche), des deux types de comportements cellulaires. Selon ces proportions les souches sont ainsi plutôt « passives » ou plutôt « réactives ». Les cellules plus réactives, et à fortiori les souches qui le sont davantage, réagissent agressivement au stress et semblent aussi reprendre leur division plus rapidement, avec le risque d'échouer et de tuer mère et fille. Les cellules passives, et donc surtout les souches qui en ont une fraction plus grande, réagissent moins au stress et sont initialement plus vulnérables, mais semblent davantage viables sur le long terme.

En effet, après 168 heures dans 2,4 M KCl les cellules les plus réactives dans les souches de rang le plus élevé commencent à disparaître. Comme vu sur des enregistrements vidéos, ces cellules, les plus agressives dans leur réponse, commencent à se diviser après 168 h et meurent. La signalisation rapide et un rétablissement de la division, qui serait un avantage dans des conditions modérées de stress semble devenir un inconvénient en cas de stress plus sévère ou prolongé. Les souches plus passives, de réponse transcriptionnelle modérée ont une meilleure viabilité sur le long terme.

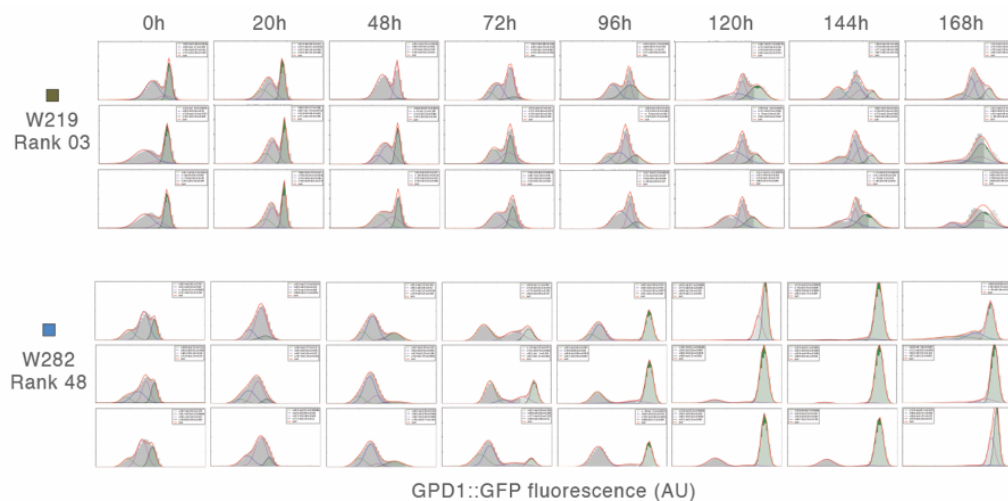


FIGURE 5.41 – **Reconstructions de séquences temporelles de profils d'expression génétique dans la réponse au stress hyper-osmotique.** Exemples de reconstructions à partir de permutations de répliques de séquences temporelles des profils d'expression de la fluorescence du gène GPD1 : :GFP en bout de la voie HOG de réponse au stress hyper-osmotique fusionné au rapporteur fluorescent GFP pour deux souches classées aux extrêmes dans le classement issu du clustering. Les distributions aux instants successifs représentent la réponse transcriptionnelle de cultures post-diauxiques suite à un stress de 2.5 M KCl à $t = 0$ h. Les figures montrent la décomposition en quatre Gaussiennes des distributions avec en rouge la somme des 4 décompositions. Trois exemples de permutations par souche sont indiquées pour deux souches, W219 de rang 3 (plus prudente), et W282 de rang 48 (plus intrépide).

Le classement des souches de la figure 5.43 à partir de l'algorithme de regroupement ordonne les souches selon leur tendance de plus en plus réactive (c'est-à-dire possédant une fraction croissante des cellules de la population avec une signalisation rapide et forte). Le rang des souches dans ce classement, les groupements ordonnés de la même façon, prédisent la viabilité aux différents instants et différentes conditions de stress (figure 5.42). C'est-à-dire pour un niveau de stress donné et de durée d'exposition il existe une relation régulière entre le rang d'une souches et sa viabilité. Cette relation change suivant les conditions osmotiques (intensité ou durée du stress). En conséquence, en fonction de la sévérité du stress ou de la durée d'exposition la stratégie de survie (c'est-à-dire le rang, ou le regroupement moyen) optimale (celle de viabilité la plus élevée) est différente.

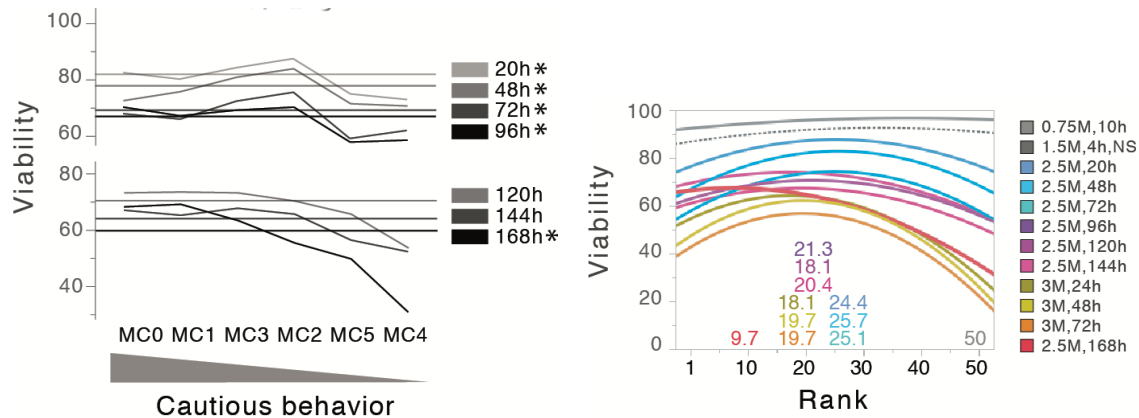


FIGURE 5.42 – **Viabilités des différentes souches à un stress intense et rang issu du clustering.** A. Viabilités moyennes en fonction des regroupements moyens successifs classés des plus « lents » aux plus « rapides », voir figure 5.43 MC0 (2 souches), MC1 (15), MC3 (12), MC2 (12, MC5 (6), MC4 (3) dans des cultures à 2,5 MKCl aux différents temps indiqués (du haut vers le bas). Les lignes horizontales indiquent la viabilité moyenne des courbes (moyenne sur tous les regroupements, c'est-à-dire sur toutes les 50 souches). Aux temps courts après l'application du stress la viabilité de tous les regroupements est proche, sauf le plus rapide MC4. Aux temps longs sous stress, la viabilité diminue fortement des regroupements plus prudents aux plus intrépides. B. Le rang prédit la viabilité dans différents environnements osmotiques. Les courbes indiqués sont les meilleurs ajustements d'un polysome quadratique pour chaque environnement indiqué. La stratégie optimale de survie se décale du rang le plus élevé (50, le plus rapide) au rang le plus bas (plus lent) dans des environnement de plus en plus sévères.

Il existe donc une signification biologique à la présence au sein de chaque population isogénique d'un mélange de cellules avec des comportements de signalisation différents. Une fraction de cellules « passives » et une fraction d'autres « réactives ». La proportion relative des deux types de réponses détermine le rang dans notre ordonnancement et correspond à des stratégies de survie différentes, adaptées à des environnements différents. Au sein de chaque souche les cellules réactives ont une signalisation rapide et forte au stress osmotique, qui les amènent à tenter une reprise plus rapide de leur cycle cellulaire. Peut-être avant le retour à un environnement standard et compromet la survie des cellules filles. Une stratégie bien adaptée à un changement osmotique modéré pour lequel les mécanismes de résistance à la pression osmotique activés rapidement permettent une reprise de la vie cellulaire. La fraction des cellules passive sont plus lentes à réagir et activent moins fortement les gènes HOG et les mécanismes d'adaptation au stress qu'ils contrôlent. Une stratégie modérée est a priori inadaptée à court terme, mais s'avère favorable sur le long car elle ne précipite pas la cellule dans une reprise prématurée de la division.

La relation entre les différents génotypes des souches, le phénotype transcriptionnel de chacune et sa valeur adaptative, approximée ici par la viabilité, semble ainsi varier continument entre les génotypes par le biais de l'ajustement des fractions des cellules passives et réactives dans chaque population de réponse rapide.

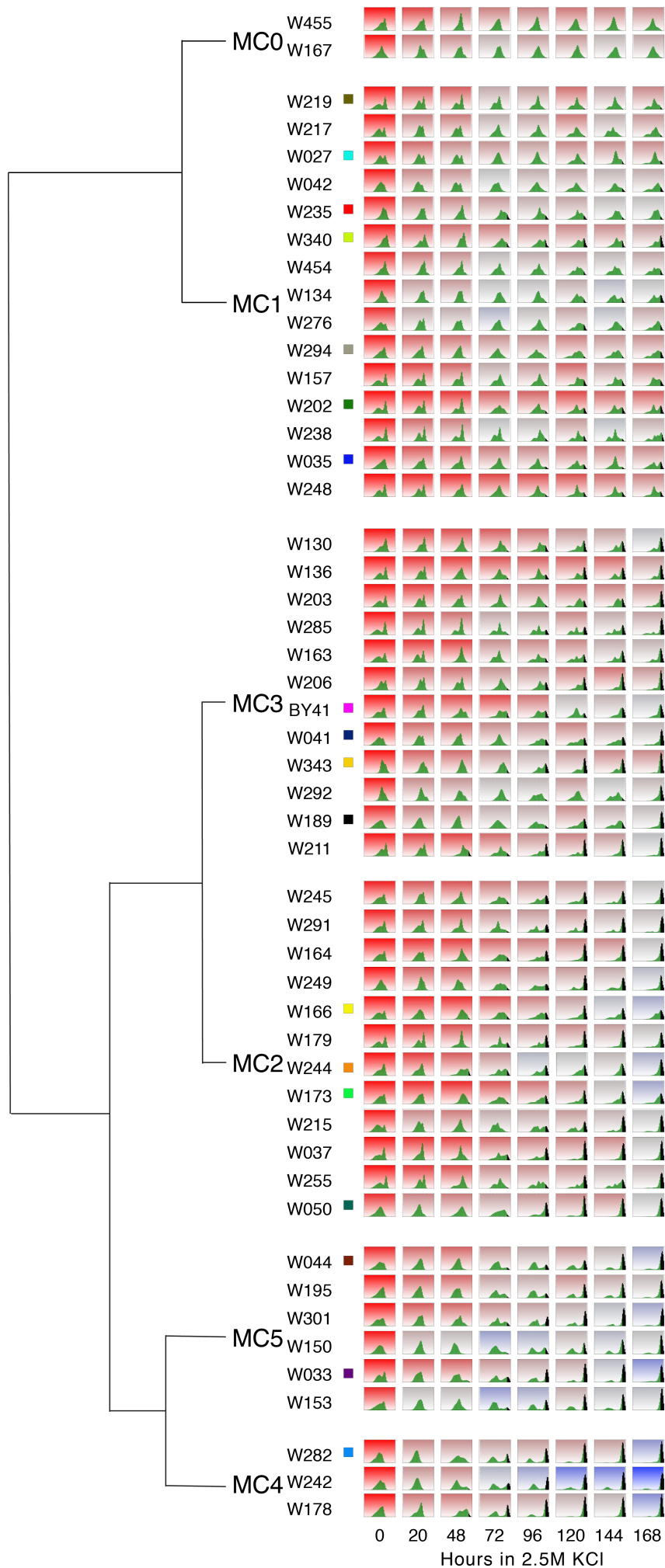


FIGURE 5.43 – (Page précédente.) **Variation continue de la survie durant un stress hyper-osmotique sévère activité de la voie de réponse de cultures post-diauxiques.** Variation continue de la survie et des taux d'accumulation de fluorescence GPD1 : :GFP de cultures post-diauxiques durant un stress hyper-osmotique sévère. Clustering moyen des différentes souches en fonction des profils temporels des distribution de fluorescence du gène GPD1 marqué à la GFP mesurés sur cellules uniques par cytométrie de flux activé par fluorescence FACS (4 à 15 réplifications par condition) et classement des souches de « plus prudente » à « plus intrépide » en fonction du pourcentage de regroupement de chaque souche à la caractéristique moyenne de chaque cluster. Les histogrammes (en vert) sont pour chaque temps une distribution caractéristique de la fluorescence GPD1 : :GFP à chaque instant suivant le stress (de gauche à droite 0, 20, 48, 72, 96, 120, 144, 168 heures). Fond coloré indique la survie relative, rouge 99,7% de viabilité à bleu 11,7% de viabilité. La numérotation du classement séquentiel selon la description dans le texte va de 1 en haut à 50 en bas. Avant le stress osmotique 2,5 M KCl (fort) à 0h les viabilités stationnaires étaient uniformément élevée (entre 93 et 99,6%)

Un modèle d'évolution et ajustement des stratégies Nous avons pu capturer les différentes stratégies fonction des proportions de cellules de type différents dans la population isogénique dans un modèle simple à deux états qui inclut des proportions hérissables de cellules lentes et rapides. Ce modèle reproduit les observations des variations de signalisation et des différences de survie suivant les changements d'environnements osmotiques et explique la relation entre rang et viabilité. Pour cela on suppose que le comportement par défaut, ancestral est une réponse au stress osmotique agressive et rapide. On suppose qu'en raison de l'imprédictibilité, de sévérité et de durée, des changements des conditions osmotiques de l'environnement les souches ont évolué une probabilité hérissable de produire un comportement passif de signalisation et réponse au stress.

Le modèle considère 3 environnements de stress osmotiques. E0 : un environnement permissif où les deux types de cellules croissent aussi bien. E1 : un environnement de stress modéré dans lequel les cellules réactives se divisent mais pas les passives. E2 : un environnement de stress osmotique sévère dans lequel les cellules réactives (rapides) se divisent mais meurent et les cellules passives (lentes) survivent sans se diviser. À chaque génération la destinée d'une cellule dépend de l'environnement dans lequel elle se trouve, de son type, ainsi que d'une probabilité p pour une cellule réactive de devenir passive. En désignant par P et R les nombres de cellules de chaque type, les règles d'évolution dans chaque environnement à chaque génération sont :

<p>E0 : les deux types de cellules (R et P) se divisent $R \rightarrow R + (1 - p)(R + P) (= R + (1 - p)R + p(R + P))$ $P \rightarrow P + p(R + P) (= P + pR + pP)$ (Donc population totale $P_{tot} = R + P \rightarrow P_{tot} + p(R + P) > 2P_{tot}$)</p>
<p>E1 : seulement les cellules R se divisent $R \rightarrow R + (1 - p)R$ $P \rightarrow P + pR$ (Donc population totale $P_{tot} = R + P \rightarrow P_{tot} + (1 - p)R + pR = P_{tot} + R$)</p>
<p>E2 : les cellules R meurent; les cellules P restent (sans division) $R \rightarrow 0$ $P \rightarrow P$ (Donc population totale $P_{tot} = R + P \rightarrow P$)</p>

On détermine la valeur adaptative relative en fonction de p pour les 9 combinaisons de transition possibles (3 combinaisons correspondant à aucun changement). La plupart des combinaisons d'environnements favorisent une stratégie optimale soit entièrement prudente ($P = 1$) ou entièrement réactive ($P = 0$). Des stratégies intermédiaires sont optimales seulement dans le cas de changements d'environnement osmotique de modéré à sévère $E_1 \rightarrow E_2$.

La stratégie optimale dépend alors du nombre de générations dans le premier environnement. Ce modèle simple propose un cadre conceptuel pour comprendre comment une probabilité hérissable d'un type de cellules passives peut être adapté par l'évolution en réponse à différents

schémas de changements environnementaux et suggère également que les stratégies actuelles contiennent une « mémoire » d’environnements passés récents (effet du nombre de générations dans l’environnement précédent).

Conclusions Ces résultats suggèrent une propriété surprenante pour les relations $G \rightarrow P$. Les différentes souches utilisées dans les expériences ont été construites à partir de populations de levures *Saccharomyces* issues d’environnements différents. Dans ces habitats d’origine les conditions d’application du stress hyper-osmotique, critique pour la survie, peuvent se manifester de manières très différentes en termes d’intensité, de durée, de fréquence, etc. ... Par exemple, des milieux arides, mais soumis à fortes précipitations brèves et soudaines, sont à l’opposé des conditions dans des milieux usuellement humides. Pour les levures issues d’habitats aussi différents, nos résultats suggèrent que la relation entre leur génotype et leur phénotype ($G \rightarrow P$) peut s’adapter avec flexibilité et continûment. Suivant la manière dont s’applique le stress dans des habitats différents, les levures peuvent adapter progressivement leur réponse optimale selon des critères de durées, fréquence et amplitude des événements. Le scénario mis en évidence rejoint les questions des stratégies microbienne de minimisation de risque (bet-hedging) observées dans d’autres systèmes [BALÁZSI et collab., 2011; OLOFSSON et collab., 2009; SIMONS et collab., 2011]. Nous avons mis en évidence comment de telles stratégies peuvent évoluer et s’adapter continûment à des situations environnementales particulières.

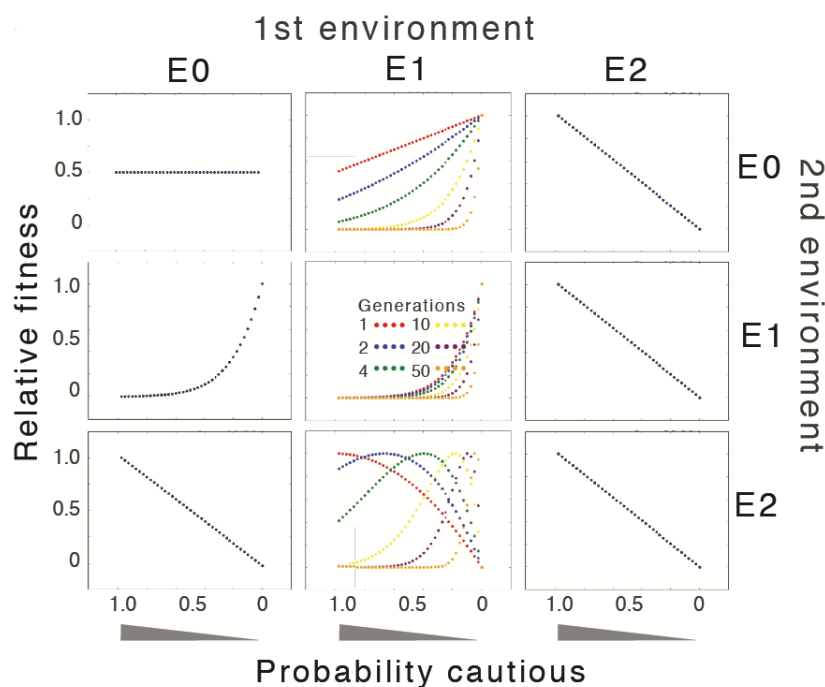


FIGURE 5.44 – **Modèle de prise de risque (« bet-hedging ») continu de la réponse au stress hyper-osmotique.** Comportement du modèle simple de variabilité de réponse au stress avec une proportion héritable de cellules passives et réactives. La probabilité p de cellules passives définit une stratégie d’adaptation à des changements entre les trois environnements possibles. La valeur adaptative relative est mesurée pour toutes les stratégies après 10 générations dans chaque environnement. Les 9 transitions possibles entre les 3 environnements de stress $E_i, i = (0, 1, 2)$ ont été considérées. E0 : toutes les cellules se divisent également. E1 : stress modéré, les cellules réactives se divisent, les passives survivent sans division. E2 : stress intense, les cellules actives meurent, les passives survivent sans division. Des stratégies intermédiaires ($0 < p < 1$) sont optimales seulement quand le changement d’environnement est d’un milieu modéré à plus sévère ($E1 \rightarrow E2$). Quand E1 est le premier environnement, la stratégie optimale p dépend du nombre de générations.

5.7 Autres travaux de biologie ds systèmes : Réponse au stress osmotique et contrôle temps réel de l'expression

Sans les développer en raison d'un mémoire déjà bien long, je mentionnerais très brièvement seulement une autre série de travaux de biologie des systèmes auxquels j'ai participé de 2008 à 2012 au laboratoire Matières et Systèmes Complexes en collaboration avec Pascal Hersen et Grégory Batt. Ces recherches exploitant comme ci-dessus le système de réponse au stress hyperosmotique de la levure visaient d'une part à caractériser d'un point de vue biophysique les mécanismes et les limitations physiques de la voie de signalisation HOG, et d'autre part à étudier comment développer sur des cellules un système de contrôle en temps réel permettant de maîtriser le niveau d'expression temporel d'un gène de cellules uniques et de populations. Un courant de recherche très excitant auquel cependant je ne participe plus ayant développé l'axe de recherche sur les cultures neuronales qui fera l'objet des chapitres suivants.

Publications dans la thématique :

- « Severe osmotic compression triggers a slowdown of intracellular signaling, which can be explained by molecular crowding. », A. Miermont, F. Waharte, S. Hu, MN. McClean, S. Bottani, S. Léon, P. Hersen. Proc Natl Acad Sci U S A. (2013)
- « Long-term model predictive control of gene expression at the population and single-cell levels », J. Uhlendorf, A. Miermont, T. Delaveau, G. Charvin, F. Fages, S. Bottani, G. Batt, P. Hersen, Proc Natl Acad Sci U S A. 2012 Aug 28;109(35) :14271-6. (2012)
- « Signaling cascade dynamics after a hyper-osmotic shock in the yeast *Saccharomyces cerevisiae*. », A. Miermont, S. Bottani, P. Hersen, European Biophysics Journal with Biophysics Letters 40, 213 (2011).
- « Towards real-time control of gene expression : controlling the hog signaling cascade. », J. Uhlendorf, S. Bottani, F. Fages, P. Hersen, G. Batt, Pac Symp Biocomput. 338-49 (2011).

5.8 Conclusion du chapitre

Dans ce chapitre j'ai rassemblé mes principales contributions dans le champ de la biologie des systèmes qui recouvrent une assez grande variété de thèmes. Je me suis intéressé ici aux questions d'organisation et de fonctionnement de systèmes génétiques impliquant de nombreux éléments en interaction. À la différence des objets physiques dont de nombreux collègues étudient les formes, propriétés et principes de genèse, il s'agit ici de structures virtuelles. Des objets d'étude non matériels mais qui reflètent néanmoins une réalité : l'ensemble des réactions biochimiques et génétiques possibles. Les réseaux des interactions potentielles ont eux aussi formes, propriétés et principes de genèse importants à déchiffrer pour comprendre les aspects moléculaires du vivant. La morphogenèse des systèmes biologiques moléculaire est encore bien en reste par rapport aux progrès de la compréhension du développement physique du vivant. Peu de principes clairs sont identifiés. La recherche concerne encore grandement la caractérisation des systèmes et la formulation d'hypothèses.

Un champ important de la recherche théorique, à peine mentionné dans ma discussion et qui complète les quelques pistes où je me suis engagé, est celui de l'évolution artificielle. Ces recherches où des populations d'organismes virtuels aux génomes artificiels évoluent sur ordinateurs permettent d'explorer les structures de régulation possibles conformes à des modèles de dynamiques et des modèles d'évolution [FORTUNA et collab., 2017]. La réalisation d'évolution *in silico* de modèles est un outil puissant d'exploration des idées théoriques sur la constitution des réseaux d'interactions [FRANÇOIS et HAKIM, 2004; PAYNE et WAGNER, 2015b]. Ces approches abordent parfaitement les questions de l'auto-organisation des systèmes de régulation. Je n'ai pas osé m'y aventurer toutefois, recherchant plus de liens avec la biologie et des données.

Les travaux avec Reiner Veitia m'ont permis d'explorer des hypothèses sur le fonctionnement du génome et son évolution en rapport avec les effets des nombres de copies des gènes. Les réseaux sont moins directement présents dans les sujets abordés si ce n'est pas les contraintes qu'ils imposent aux gènes individuels imbriqués, piégés dans un système de relations. L'originalité des idées dans ces contributions réside dans la mise en évidence d'effets globaux du système sur les gènes en particulier. Les propriétés des gènes et leur évolution dépendent de tout leur environnement et s'impose automatiquement à eux. Des conditions physiques, tel que le volume du noyau et de la cellule, sa fraction accessible, la dimension totale du génome, affectent chaque gène et les dotent tous de propriétés exploitées par le vivant. Des conditions globales physiques et chimiques façonnent les possibilités de la génétique.

Mon discours sur la biologie a démarré par le récit d'études de morphogenèse physique par des collègues. Propriétés de forces et d'écoulements de matière imposant aux cellules d'un organisme en développement et le façonnent. Il sera très pertinent dans l'avenir de rapprocher les différentes études de ces différents façonnages aux niveaux moléculaires et cellulaires car ces niveaux naturellement interagissent étroitement. Le vivant combine physique, chimie et génétique à toutes les échelles. Et il faudra bien arriver à combiner les approches scientifiques aussi pour y voir plus clair.

5.9 Références

- «Database of Interacting Protein», URL <http://dip.doe-mbi.ucla.edu/>. 146
- ACKERS, G. K., A. D. JOHNSON et M. A. SHEA. 1982, «Quantitative model for gene regulation by lambda phage repressor.», *Proc Natl Acad Sci U S A*, vol. 79, n° 4, p. 1129–1133. 160, 169
- ALBERCH, P. 1991, «From genes to phenotype : dynamical systems and evolvability», *Genetica*, vol. 84, n° 1, doi :10.1007/BF00123979, p. 5–11, ISSN 00166707. 196
- ALBERT, R. et A. L. BARABASI. 2002, «Statistical mechanics of complex networks», *Rev. Mod. Phys.*, vol. 74, doi :10.1103/RevModPhys.74.47, p. 47–97. 141
- ALON, U. 2003, «Biological networks : the tinkerer as an engineer.», *Science*, vol. 301, n° 5641, doi : 10.1126/science.1089072, p. 1866–1867. 144
- ALON, U. 2007, «An Introduction to Systems Biology : Design Principles of Biological Circuits», doi :citeulike-article-id:1314150. URL <http://www.loc.gov/catdir/enhancements/fy0654/2005056902-d.html>. 145
- ARITA, M. 2004, «The metabolic world of *Escherichia coli* is not small.», *Proc Natl Acad Sci U S A*, vol. 101, n° 6, doi :10.1073/pnas.0306458101, p. 1543–1547. 144
- ASHER, G., N. REUVEN et Y. SHAUL. 2006, «20S proteasomes and protein degradation “by default”», *BioEssays*, vol. 28, n° 8, doi :10.1002/bies.20447, p. 844–849, ISSN 0265-9247. 177
- BALÁZSI, G., A. VAN OUDENAARDEN et J. J. COLLINS. 2011, «Cellular decision making and biological noise : from microbes to mammals.», *Cell*, vol. 144, n° 6, doi :10.1016/j.cell.2011.01.030, p. 910–25, ISSN 1097-4172. 203
- BINTU, L., N. E. BUCHLER, H. G. GARCIA et collaborateurs. 2005a, «Transcriptional regulation by the numbers : applications.», *Curr Opin Genet Dev*, vol. 15, n° 2, doi :10.1016/j.gde.2005.02.006, p. 125–135. 168, 169
- BINTU, L., N. E. BUCHLER, H. G. GARCIA et collaborateurs. 2005b, «Transcriptional regulation by the numbers : models.», *Curr Opin Genet Dev*, vol. 15, n° 2, doi :10.1016/j.gde.2005.02.007, p. 116–124. 168, 169

- BIRCHLER, J. A. 1979, «A study of enzyme activities in a dosage series of the long arm of chromosome one in maize.», *Genetics*, vol. 92, n° 4, p. 1211–1229, ISSN 0016-6731. [176](#)
- BIRCHLER, J. A., U. BHADRA, M. P. BHADRA et D. L. AUGER. 2001, «Dosage-Dependent Gene Regulation in Multicellular Eukaryotes : Implications for Dosage Compensation, Aneuploid Syndromes, and Quantitative Traits», *Developmental Biology*, vol. 234, n° 2, doi :10.1006/dbio.2001.0262, p. 275–288, ISSN 00121606. [175](#)
- BIRCHLER, J. A. et K. J. NEWTON. 1981, «Modulation of protein levels in chromosomal dosage series of maize : the biochemical basis of aneuploid syndromes.», *Genetics*, vol. 99, n° 2, p. 247–266, ISSN 0016-6731. [176](#)
- BOTTANI, S. et B. GRAMMATICOS. 2007a, «Analysis of a minimal model for p53 oscillations.», *J Theor Biol*, vol. 249, n° 2, doi :10.1016/j.jtbi.2007.04.026, p. 235–245. [154](#)
- BOTTANI, S. et B. GRAMMATICOS. 2007b, «Analysis of a minimal model for p53 oscillations.», *Journal of theoretical biology*, vol. 249, n° 2, doi :10.1016/j.jtbi.2007.04.026, p. 235–245, ISSN 0022-5193. [154](#), [155](#), [156](#)
- BOTTANI, S. et B. GRAMMATICOS. 2008, «A simple model of genetic oscillations through regulated degradation», *Chaos, Solitons & Fractals*, vol. 38, n° 5, doi :10.1016/j.chaos.2008.03.010, p. 1468–1482, ISSN 09600779. [157](#)
- BOTTANI, S. et B. GRAMMATICOS. 2012, «Stochastic and deterministic simulations of a delayed genetic oscillation model : Investigating the validity of reductions», *Chaos, Solitons and Fractals*, vol. 45, n° 5, doi :10.1016/j.chaos.2011.12.012, ISSN 09600779. [158](#), [160](#), [161](#), [162](#), [165](#)
- BOTTANI, S. et R. A. VEITIA. 2016, «Hill function-based models of transcriptional switches : Impact of specific, nonspecific, functional and nonfunctional binding», *Biological Reviews*, doi : 10.1111/brv.12262, ISSN 1469185X. [165](#), [166](#), [168](#), [170](#), [171](#)
- BOTTANI, S., N. ZABET, J. WENDEL et R. VEITIA. 2018, «Gene Expression Dominance in Allopolyploids : Hypotheses and Models», *Trends in Plant Science*, doi :10.1016/j.tplants.2018.01.002, ISSN 13601385. [187](#), [188](#)
- BUCHLER, N. E., U. GERLAND et T. HWA. 2003, «On schemes of combinatorial transcription logic.», *Proc Natl Acad Sci U S A*, vol. 100, n° 9, doi :10.1073/pnas.0930314100, p. 5136–5141. [169](#)
- CAREY, N. 2015, *Junk DNA : A Journey Through the Dark Matter of the Genome.*, Columbia University Press, ISBN 9780231539418. [182](#)
- CARRÉ, C., A. MAS et G. KROUK. 2017, «Reverse engineering highlights potential principles of large gene regulatory network design and learning», *npj Systems Biology and Applications*, vol. 3, n° 1, doi :10.1038/s41540-017-0019-y, p. 17, ISSN 2056-7189. [141](#)
- CATALÁN, P., C. F. ARIAS, J. A. CUESTA et S. MANRUBIA. 2017, «Adaptive multiscapes : an up-to-date metaphor to visualize molecular adaptation.», *Biology direct*, vol. 12, n° 1, doi :10.1186/s13062-017-0178-1, p. 7, ISSN 1745-6150. [196](#)
- CAVALIER-SMITH, T. 1982, «Skeletal DNA and the evolution of genome size.», *Annual review of biophysics and bioengineering*, vol. 11, n° 54, doi :10.1146/annurev.bb.11.060182.001421, p. 273–302, ISSN 0084-6589. [182](#)
- CAVALIER-SMITH, T. et M. J. BEATON. 1999, «The skeletal function of non-genic nuclear DNA : new evidence from ancient cell chimaeras», *Genetica*, vol. 106, n° 1/2, doi :10.1023/A:1003701925110, p. 3–13, ISSN 0016-6707. [182](#)

- CHÈNE, P. 2003, «Inhibiting the p53-MDM2 interaction : An important target for cancer therapy», *Nature Reviews Cancer*, vol. 3, n° 2, doi :10.1038/nrc991, p. 102–109, ISSN 1474175X. [154](#)
- CHENG, F., C. SUN, J. WU et collaborateurs. 2016, «Epigenetic regulation of subgenome dominance following whole genome triplication in *Brassica rapa*», *New Phytologist*, vol. 211, n° 1, doi :10.1111/nph.13884, p. 288–299, ISSN 14698137. [188](#)
- CHOI, J., K. W. YANG, T. Y. LEE et S. LEE. 2008, «New time-scale criteria for model simplification of bio-reaction systems.», *BMC Bioinformatics*, vol. 9, n° 1, doi :10.1186/1471-2105-9-338, p. 338. [164](#)
- CILIBERTO, A., B. NOVAK et J. J. TYSON. 2005, «Steady states and oscillations in the p53/{M}dm2 network.», *Cell Cycle*, vol. 4, n° 3, p. 488–493. [154](#)
- CONANT, G. C., J. A. BIRCHLER et J. C. PIRES. 2014, «Dosage, duplication, and diploidization : clarifying the interplay of multiple models for duplicate gene evolution over time», *Current Opinion in Plant Biology*, vol. 19, doi :10.1016/J.PBI.2014.05.008, p. 91–98, ISSN 1369-5266. [175](#), [194](#)
- COURCELLE, J., A. KHODURSKY, B. PETER, P. O. BROWN et P. C. HANAWALT. 2001, «Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*.», *Genetics*, vol. 158, n° 1, p. 41–64, ISSN 0016-6731. [139](#)
- D'ALCHÉ-BUC, F., P. P.-J. LAHAYE, B. PERRIN et collaborateurs. 2005, «A Dynamic Model of Gene Regulatory Networks Based on Inertia Principle», *Bioinformatics Using*, vol. 117, doi :10.1007/10950913_5, p. 93. [140](#), [141](#)
- DEHAL, P. et J. L. BOORE. 2005, «Two rounds of whole genome duplication in the ancestral vertebrate», *PLoS Biology*, vol. 3, n° 10, doi :10.1371/journal.pbio.0030314, p. e314, ISSN 15449173. [182](#)
- DEMIRKIRAN, G., G. KALAYCI DEMIR et C. GÜZELİŞ. 2017, «Revealing determinants of two-phase dynamics of P53 network under gamma irradiation based on a reduced 2D relaxation oscillator model», *IET Systems Biology*, vol. 12, n° 1, doi :10.1049/iet-syb.2017.0041, p. 26–38, ISSN 1751-8849. [153](#)
- DUNLAP, J. C. 1999, «Molecular bases for circadian clocks.», *Cell*, vol. 96, n° 2, p. 271–290. [150](#)
- ELIAŠ, J., L. DIMITRIO, J. CLAIRAMBAULT et R. NATALINI. 2014, «The p53 protein and its molecular network : Modelling a missing link between DNA damage and cell fate», *Biochimica et Biophysica Acta - Proteins and Proteomics*, vol. 1844, n° 1 PART B, doi :10.1016/j.bbapap.2013.09.019, p. 232–247, ISSN 15709639. [153](#)
- ELOWITZ, M. B. et S. LEIBLER. 2000, «A synthetic oscillatory network of transcriptional regulators.», *Nature*, vol. 403, n° 6767, doi :10.1038/35002125, p. 335–338. [151](#)
- ENCODE PROJECT CONSORTIUM. 2012, «An integrated encyclopedia of DNA elements in the human genome.», *Nature*, vol. 489, n° 7414, doi :10.1038/nature11247, p. 57–74, ISSN 1476-4687. [182](#)
- FORTUNA, M. A., L. ZAMAN, C. OFRIA et A. WAGNER. 2017, «The genotype-phenotype map of an evolving digital organism», *PLoS Computational Biology*, vol. 13, n° 2, doi :10.1371/journal.pcbi.1005414, p. 1–20, ISSN 15537358. [196](#), [204](#)
- FRANÇOIS, P. et V. HAKIM. 2004, «Design of genetic networks with specified functions by evolution in silico.», *Proc Natl Acad Sci USA*, vol. 101, n° 2, doi :10.1073/pnas.0304532101, p. 580–585. [204](#)
- FUNG, E., W. W. WONG, J. K. SUEN et collaborateurs. 2005, «A synthetic gene-metabolic oscillator.», *Nature*, vol. 435, n° 7038, doi :10.1038/nature03508, p. 118–122. [151](#)

- GARDNER, T. S., D. DI BERNARDO, D. LORENZ et J. J. COLLINS. 2003, «Inferring genetic networks and identifying compound mode of action via expression profiling.», *Science (New York, N.Y.)*, vol. 301, n° 5629, doi :10.1126/science.1081900, p. 102–5, ISSN 1095-9203. [139](#)
- GAVRILETS, S. 1997, «Evolution and speciation on holey adaptive landscapes», *Trends in Ecology & Evolution*, vol. 12, n° 8, doi :10.1016/S0169-5347(97)01098-7, p. 307–312, ISSN 01695347. [197](#)
- GERSTEIN, A. C., H. J. E. CHUN, A. GRANT et S. P. OTTO. 2006, «Genomic convergence toward diploidy in *Saccharomyces cerevisiae*», *PLoS Genetics*, vol. 2, n° 9, doi :10.1371/journal.pgen.0020145, p. 1396–1401, ISSN 15537390. [183](#)
- GIBSON, M. A. et J. BRUCK. 2000, «Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels», *The Journal of Physical Chemistry A*, vol. 104, n° 9, doi : 10.1021/jp993732q, p. 1876–1889. [159](#)
- GILLESPIE, D. T. 1977, «Exact Stochastic simulation of coupled chemical reactions», *The Journal of Physical Chemistry*, vol. 81, p. 2340–2361. [159](#)
- GINOZA, R. et A. MUGLER. 2010, «Network motifs come in sets : Correlations in the randomization process», *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 82, n° 1, doi : 10.1103/PhysRevE.82.011921, p. 011 921, ISSN 15393755. [149](#)
- GNANAPRADEEPAN, K., S. BASU, T. BARNOUD et collaborateurs. 2018, «The p53 tumor suppressor in the control of metabolism and ferroptosis», *Frontiers in Endocrinology*, vol. 9, n° APR, doi : 10.3389/fendo.2018.00124, p. 124, ISSN 16642392. [152](#)
- GOLDBETER, A. 1991, «A minimal cascade model for the mitotic oscillator involving cyclin and cdc2 kinase», *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, n° 20, p. 9107–9111, ISSN 0027-8424. [185](#)
- GOLDBETER, A. 2002, «Computational approaches to cellular rhythms.», *Nature*, vol. 420, n° 6912, doi :10.1038/nature01259, p. 238–245. [150](#)
- GÓMEZ-URIBE, C. A., G. C. VERGHESE et A. R. TZAFRIRI. 2008, «Enhanced identification and exploitation of time scales for model reduction in stochastic chemical kinetics.», *The Journal of chemical physics*, vol. 129, n° 24, doi :10.1063/1.3050350, ISSN 1089-7690. [165](#)
- GONZE, D., J. C. LELOUP et A. GOLDBETER. 2000, «Theoretical models for circadian rhythms in {N}eurospora and {D}rosophila.», *C R Acad Sci III*, vol. 323, n° 1, p. 57–67. [150](#)
- GREENBLATT, J. 1997, «RNA polymerase II holoenzyme and transcriptional regulation.», *Curr Opin Cell Biol*, vol. 9, n° 3, p. 310–319. [166](#)
- GREGOR, T., D. W. TANK, E. F. WIESCHAUS et W. BIALEK. 2007, «Probing the limits to positional information.», *Cell*, vol. 130, n° 1, doi :10.1016/j.cell.2007.05.025, p. 153–64, ISSN 0092-8674. [168](#)
- GREGORY, T. 2001, «Cambridge Philosophical Society. Biological Reviews of the Cambridge Philosophical Society. Cambridge University Press; 1935. <https://www.cambridge.org/core/journals/biological-reviews/article/the-trophic-ecology-of-freshwater-gammarus-spp-crustacea-a>», *Biol Rev Camb Philos Soc* 7, vol. 76, p. 65–101, ISSN 1464-7931. [182](#)
- GUELZIM, N., S. BOTTANI, P. BOURGINE et K. FRANÇOIS. 2002, «Topological and causal structure of the yeast transcriptional regulatory network.», *Nat Genet*, vol. 31, n° 1, doi :10.1038/ng873, p. 60–63. [141](#), [142](#), [143](#), [144](#), [146](#)

- HAFNER, A., J. STEWART-ORNSTEIN, J. E. PURVIS et collaborateurs. 2017, «P53 pulses lead to distinct patterns of gene expression albeit similar DNA-binding dynamics», *Nature Structural and Molecular Biology*, vol. 24, n° 10, doi :10.1038/nsmb.3452, p. 840–847, ISSN 15459985. [153](#)
- HAUPT, Y., R. MAYA, A. KAZAZ et M. OREN. 1997, «Mdm2 promotes the rapid degradation of p53.», *Nature*, vol. 387, n° 6630, p. 296–299. [153](#)
- HIRATA, H., S. YOSHIURA, T. OHTSUKA et collaborateurs. 2002, «Oscillatory expression of the b{HLH} factor Hes1 regulated by a negative feedback loop.», *Science*, vol. 298, n° 5594, doi : 10.1126/science.1074560, p. 840–843. [150](#)
- HIRATE, Y., S. BOTTANI, W. L. PANG et S. RUTHERFORD. 2016, «Tunable bet hedging in yeast responses to osmotic stress», cahier de recherche, doi :10.1101/039982. URL <https://www.biorxiv.org/content/early/2016/04/05/039982><http://biorxiv.org/lookup/doi/10.1101/039982>. [198](#)
- HOFFMANN, A., A. LEVCHENKO, M. L. SCOTT et D. BALTIMORE. 2002, «The κ B-NF- κ B signaling module : temporal control and selective gene activation.», *Science*, vol. 298, n° 5596, doi :10.1126/science.1071914, p. 1241–1245. [150](#)
- HOLLISTER, J. D. et B. S. GAUT. 2009, «Epigenetic silencing of transposable elements : A trade-off between reduced transposition and deleterious effects on neighboring gene expression», *Genome Research*, vol. 19, n° 8, doi :10.1101/gr.091678.109, p. 1419–1428, ISSN 10889051. [188](#)
- HUGHES, T. R. et C. G. DE BOER. 2013, «Mapping yeast transcriptional networks.», *Genetics*, vol. 195, n° 1, doi :10.1534/genetics.113.153262, p. 9–36, ISSN 1943-2631. [143](#)
- HUNG, D. Y. et L. SHAPIRO. 2002, «A signal transduction protein cues proteolytic events critical to *Caulobacter* cell cycle progression.», *Proc Natl Acad Sci U S A*, vol. 99, n° 20, doi :10.1073/pnas.202495099, p. 13 160–13 165. [157](#)
- INGRAM, P. J., M. P. STUMPF et J. STARK. 2006, «Network motifs : structure does not determine function.», *BMC Genomics*, vol. 7, n° 1, doi :10.1186/1471-2164-7-108, p. 108, ISSN 14712164. [149](#)
- JENAL, U. et R. H. ARONIS. 2003, «Regulation by proteolysis in bacterial cells.», *Curr Opin Microbiol*, vol. 6, n° 2, p. 163–172. [157](#)
- JEONG, H., S. P. MASON, A. L. BARABASI et Z. N. OLTVAI. 2001, «Lethality and centrality in protein networks.», *Nature*, vol. 411, n° 6833, doi :10.1038/35075138, p. 41–42. [141](#)
- JOTHI, R., S. BALAJI, A. WUSTER et collaborateurs. 2009, «Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture.», *Molecular systems biology*, vol. 5, doi :10.1038/msb.2009.52, p. 294, ISSN 1744-4292. [143](#)
- KAZEMIAN, M., H. PHAM, S. A. WOLFE, M. H. BRODSKY et S. SINHA. 2013, «Widespread evidence of cooperative DNA binding by transcription factors in *Drosophila* development», *Nucleic Acids Research*, vol. 41, n° 17, doi :10.1093/nar/gkt598, p. 8237–8252, ISSN 1362-4962. [166](#)
- KNABE, J. F., C. L. NEHANIV et M. J. SCHILSTRA. 2008, «Do motifs reflect evolved function?—No convergent evolution of genetic regulatory network subgraph topologies», *Biosystems*, vol. 94, n° 1-2, doi :10.1016/j.biosystems.2008.05.012, p. 68–74, ISSN 03032647. [149](#)
- KONAGURTHU, A. S. et A. M. LESK. 2008, «On the origin of distribution patterns of motifs in biological networks», *BMC Systems Biology*, vol. 2, n° 1, doi :10.1186/1752-0509-2-73, p. 73, ISSN 17520509. [149](#)

- KURAS, L., A. ROUILLON, T. LEE et collaborateurs. 2002, «Dual regulation of the met4 transcription factor by ubiquitin-dependent degradation and inhibition of promoter recruitment.», *Mol Cell*, vol. 10, n° 1, p. 69–80. [157](#)
- LAHAV, G., N. ROSENFELD, A. SIGAL et collaborateurs. 2004, «Dynamics of the p53-Mdm2 feedback loop in individual cells.», *Nat Genet*, vol. 36, n° 2, doi :10.1038/ng1293, p. 147–150. [150](#), [152](#), [153](#), [154](#)
- LANDER, E. S., L. M. LINTON, B. BIRREN et collaborateurs. 2001, «Initial sequencing and analysis of the human genome», *Nature*, vol. 409, n° 6822, doi :10.1038/35057062, p. 860–921, ISSN 0028-0836. [182](#)
- LAUB, M. T., S. L. CHEN, L. SHAPIRO et H. H. MCADAMS. 2002, «Genes directly controlled by CtrA, a master regulator of the Caulobacter cell cycle.», *Proc Natl Acad Sci U S A*, vol. 99, n° 7, doi :10.1073/pnas.062065699, p. 4632–4637. [157](#)
- LEHMAN, A. M., K. B. ELLWOOD, B. E. MIDDLETON et M. CAREY. 1998, «Compensatory energetic relationships between upstream activators and the RNA polymerase II general transcription machinery.», *J Biol Chem*, vol. 273, n° 2, p. 932–939. [166](#)
- LEROY, B., J. L. FOURNIER, C. ISHIOKA et collaborateurs. 2013, «The TP53 website : An integrative resource centre for the TP53 mutation database and TP53 mutant analysis», *Nucleic Acids Research*, vol. 41, n° D1, doi :10.1093/nar/gks1033, p. D962–9, ISSN 03051048. [152](#)
- LEROY, B., J. L. FOURNIER, C. ISHIOKA et collaborateurs. 2017, «TP53 web site», URL <http://p53.fr/>. [152](#)
- LEV, R. MAYA, L. A. SEGEL et collaborateurs. 2000, «Generation of oscillations by the p53-Mdm2 feedback loop : a theoretical and experimental study.», *Proc Natl Acad Sci U S A*, vol. 97, n° 21, doi :10.1073/pnas.210171597, p. 11 250–11 255. [150](#), [151](#), [153](#)
- LEVINE, A. J. 2011, «Introduction : The changing directions of p53 research», *Genes and Cancer*, vol. 2, n° 4, doi :10.1177/1947601911413463, p. 382–384, ISSN 19476019. [152](#)
- LEWIS, J. 2003, «Autoinhibition with transcriptional delay : a simple mechanism for the zebrafish somitogenesis oscillator.», *Curr Biol*, vol. 13, n° 16, p. 1398–1408. [150](#)
- LEWONTIN, R. C. 1974, *The genetic basis of evolutionary change*, Columbia University Press, ISBN 9780231033923, 346 p.. [196](#)
- LI, F., T. LONG, Y. LU, Q. OUYANG et C. TANG. 2004, «The yeast cell-cycle network is robustly designed.», *Proc Natl Acad Sci U S A*, vol. 101, n° 14, doi :10.1073/pnas.0305937101, p. 4781–4786. [150](#)
- LIBEREK, K. et C. GEORGOPOULOS. 1993, «Autoregulation of the Escherichia coli heat shock response by the DnaK and DnaJ heat shock proteins.», *Proc Natl Acad Sci U S A*, vol. 90, n° 23, p. 11 019–11 023. [157](#)
- LOTEM, E. Y., S. SATTATH, N. KASHTAN et collaborateurs. 2004, «Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction.», *Proc Natl Acad Sci U S A*, vol. 101, n° 16, doi :10.1073/pnas.0306752101, p. 5934–5939. [144](#)
- LYNCH, M. 2016, «Mutation and human exceptionalism : Our future genetic load», *Genetics*, vol. 202, n° 3, doi :10.1534/genetics.115.180471, p. 869–875, ISSN 19432631. [194](#)
- LYONS, E. et J. SCHNABLE. 2015, «Plant paleopolyploidy», URL https://www.genomeevolution.org/wiki/index.php/Plant{}_paleopolyploidy. [182](#)

- MA, L., J. WAGNER, J. J. RICE et collaborateurs. 2005, «A plausible model for the digital response of p53 to {DNA} damage.», *Proc Natl Acad Sci U S A*, vol. 102, n° 40, doi :10.1073/pnas.0501352102, p. 14 266–14 271. [154](#)
- MANGAN, S. et U. ALON. 2003, «Structure and function of the feed-forward loop network motif.», *Proc Natl Acad Sci U S A*, vol. 100, n° 21, doi :10.1073/pnas.2133841100, p. 11 980–11 985. [145](#)
- MAZURIE, A., S. BOTTANI et M. VERGASSOLA. 2005, «An evolutionary and functional assessment of regulatory network motifs.», *Genome Biol*, vol. 6, n° 4, doi :10.1186/gb-2005-6-4-r35. [147](#), [149](#)
- MCADAMS, H. H. et A. ARKIN. 1997, «Stochastic mechanisms in gene expression.», *Proc Natl Acad Sci U S A*, vol. 94, n° 3, p. 814–819. [159](#)
- MCMILLEN, D., N. KOPELL, J. HASTY et J. J. COLLINS. 2002, «Synchronizing genetic relaxation oscillators by intercell signaling.», *Proc Natl Acad Sci U S A*, vol. 99, n° 2, doi :10.1073/pnas.022642299, p. 679–684. [151](#)
- MESHI, O., T. SHLOMI et E. RUPPIN. 2007, «Evolutionary conservation and over-representation of functionally enriched network patterns in the yeast regulatory network», *BMC Systems Biology*, vol. 1, n° 1, doi :10.1186/1752-0509-1-1, p. 1, ISSN 17520509. [149](#)
- MEYERS, L. A. et D. A. LEVIN. 2006, «On the abundance of polyploids in flowering plants», *Evolution*, vol. 60, n° 6, doi :10.1111/j.0014-3820.2006.tb01198.x, p. 1198–1206, ISSN 0014-3820. [182](#)
- MILO, R., S. S. ORR, S. ITZKOVITZ et collaborateurs. 2002, «Network motifs : simple building blocks of complex networks.», *Science*, vol. 298, n° 5594, doi :10.1126/science.298.5594.824, p. 824–827. [144](#), [145](#)
- MIRSKY, A. E. 1951, «THE DESOXYRIBONUCLEIC ACID CONTENT OF ANIMAL CELLS AND ITS EVOLUTIONARY SIGNIFICANCE», *The Journal of General Physiology*, vol. 34, n° 4, doi :10.1085/jgp.34.4.451, p. 451–462, ISSN 0022-1295. [182](#)
- MÖNKE, G., E. CRISTIANO, A. FINZEL et collaborateurs. 2017, «Excitability in the p53 network mediates robust signaling with tunable activation thresholds in single cells», *Scientific Reports*, vol. 7, doi :10.1038/srep46571, p. 46 571, ISSN 20452322. [153](#)
- MULLER, H. J. 1964, «The relation of recombination to mutational advance», *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 1, n° 1, doi :10.1016/0027-5107(64)90047-8, p. 2–9, ISSN 00275107. [194](#)
- NELSON, D. E., IHEKWABA, M. ELLIOTT et collaborateurs. 2004, «Oscillations in NF-kappaB signaling control the dynamics of gene expression.», *Science*, vol. 306, n° 5696, doi :10.1126/science.1099962, p. 704–708. [150](#), [151](#)
- O'BRIEN, E. L., E. VAN ITALLIE et M. R. BENNETT. 2012, «Modeling synthetic gene oscillators», *Mathematical Biosciences*, vol. 236, n° 1, doi :10.1016/j.mbs.2012.01.001, p. 1–15, ISSN 00255564. [151](#)
- OHNO, S. 1970, *Evolution by Gene Duplication*, Springer Berlin Heidelberg, ISBN 978-3-642-86661-6, 170 p.. [175](#)
- OLOFSSON, H., J. RIPA et N. JONZÉN. 2009, «Bet-hedging as an evolutionary game : the trade-off between egg size and number.», *Proceedings. Biological sciences / The Royal Society*, vol. 276, n° 1669, doi :10.1098/rspb.2009.0500, p. 2963–9, ISSN 0962-8452. [203](#)
- ORR, H. A. 2009, «Fitness and its role in evolutionary genetics.», *Nature reviews. Genetics*, vol. 10, n° 8, doi :10.1038/nrg2603, p. 531–9, ISSN 1471-0064. [195](#)

- OTTO, S. P. 2007, «The Evolutionary Consequences of Polyploidy», *Cell*, vol. 131, n° 3, doi :10.1016/j.cell.2007.10.022, p. 452–462, ISSN 00928674. [182](#)
- PAHLE, J. 2009, «Biochemical simulations : stochastic, approximate stochastic and hybrid approaches», *Brief Bioinform*, vol. 10, n° 1, doi :10.1093/bib/bbn050, p. 53–64, ISSN 1477-4054. [159](#)
- PAYNE, J. L. et A. WAGNER. 2015a, «Function does not follow form in gene regulatory circuits», *Scientific Reports*, vol. 5, n° 1, doi :10.1038/srep13015, p. 13 015, ISSN 2045-2322. [149](#)
- PAYNE, J. L. et A. WAGNER. 2015b, «Function does not follow form in gene regulatory circuits.», *Scientific reports*, vol. 5, doi :10.1038/srep13015, p. 13 015, ISSN 2045-2322. [204](#)
- PIEGU, B., R. GUYOT, N. PICAULT et collaborateurs. 2006, «Doubling genome size without polyploidization : dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice.», *Genome research*, vol. 16, n° 10, doi :10.1101/gr.5290206, p. 1262–9, ISSN 1088-9051. [182](#)
- PIGLIUCCI, M. 2010, «Genotype-phenotype mapping and the end of the 'genes as blueprint' metaphor.», *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 365, n° 1540, doi :10.1098/rstb.2009.0241, p. 557–66, ISSN 1471-2970. [196](#)
- PIGLIUCCI, M. 2013, «Landscapes, Surfaces, and Morphospaces : What Are They Good For?», dans *The Adaptive Landscape in Evolutionary Biology*, édité par E. Svensson et R. Calsbeek, chap. 3, Oxford Scholarship, ISBN 9780199595372, p. 26–40, doi :10.1093/acprof:oso/9780199595372.003.0003. [196](#)
- POTVIN-TROTTIER, L., N. D. LORD, G. VINNICOMBE et J. PAULSSON. 2016, «Synchronous long-term oscillations in a synthetic gene circuit», *Nature*, vol. 538, n° 7626, doi :10.1038/nature19841, p. 514–517, ISSN 14764687. [151](#)
- POURQUIÉ, O. 2003, «The segmentation clock : converting embryonic time into spatial pattern.», *Science*, vol. 301, n° 5631, doi :10.1126/science.1085887, p. 328–330. [150](#)
- PRUTEANU, M. et R. H. ARONIS. 2002, «The cellular level of the recognition factor {R}ss{B} is rate-limiting for sigma{S} proteolysis : implications for {R}ss{B} regulation and signal transduction in sigma{S} turnover in {E}scherichia coli.», *Mol Microbiol*, vol. 45, n° 6, p. 1701–1713. [157](#)
- PRZULJ, N., D. G. CORNEIL et I. JURISICA. 2004, «Modeling interactome : scale-free or geometric?», *Bioinformatics*, doi :10.1093/bioinformatics/bth436. [144](#)
- PTASHNE, M. et A. GANN. 2002, *Genes and Signals*, New York : Cold Spring Harbor Laboratory Press. [166](#)
- PURCELL, O., N. J. SAVERY, C. S. GRIERSON et M. DI BERNARDO. 2010, «A comparative analysis of synthetic genetic oscillators», *Journal of The Royal Society Interface*, vol. 7, n° 52, doi :10.1098/rsif.2010.0183, p. 1503–1524, ISSN 1742-5689. [151](#)
- RANDRUP, Y. A., S. J. FLEISHMAN, N. B. TAL et L. STONE. 2004, «Comment on "{N}etwork motifs : simple building blocks of complex networks" and "{S}uperfamilies of evolved and designed networks".», *Science*, vol. 305, n° 5687, doi :10.1126/science.1099334. [149](#)
- REINBERG, D. et R. G. ROEDER. 1987, «Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of initiation factors IIB and IIE.», *J Biol Chem*, vol. 262, n° 7, p. 3310–3321. [166](#)
- RENNY-BYFIELD, S., L. GONG, J. P. GALLAGHER et J. F. WENDEL. 2015, «Persistence of subgenomes in paleopolyploid cotton after 60 my of evolution», *Molecular Biology and Evolution*, vol. 32, n° 4, doi :10.1093/molbev/msv001, p. 1063–1071, ISSN 15371719. [188](#)

- RIESEBERG, L. H. 1997, «Hybrid origins of plant species», *Annual Review of Ecology and Systematics*, vol. 28, n° 1, doi :10.1146/annurev.ecolsys.28.1.359, p. 359–389, ISSN 0066-4162. [188](#)
- RIESEBERG, L. H. et J. H. WILLIS. 2007, «Plant speciation», *Science*, vol. 317, n° 5840, doi :10.1126/science.1137729, p. 910–914, ISSN 00368075. [182](#)
- ROEDER, R. G. 1996, «Nuclear RNA polymerases : role of general initiation factors and cofactors in eukaryotic transcription.», *Methods Enzymol*, vol. 273, p. 165–171. [166](#)
- RONEN, M., R. ROSENBERG, B. I. SHRAIMAN et U. ALON. 2002, «Assigning numbers to the arrows : parameterizing a gene regulation network by using accurate expression kinetics.», *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, n° 16, doi :10.1073/pnas.152046799, p. 10 555–60, ISSN 0027-8424. [139](#)
- ROUILLON, A., R. BARBEY, E. E. PATTON, M. TYERS et D. THOMAS. 2000, «Feedback-regulated degradation of the transcriptional activator {M}et4 is triggered by the {SCF}{(M)et30} complex.», *EMBO J*, vol. 19, n° 2, doi :10.1093/emboj/19.2.282, p. 282–294. [157](#)
- SAIZ, L. et J. M. G. VILAR. 2006, «Stochastic dynamics of macromolecular-assembly networks», *Molecular Systems Biology*, vol. 2, doi :10.1038/msb4100061, p. 2006.0024, ISSN 1744-4292. [160](#)
- SHAI, R. MILO, S. MANGAN et U. ALON. 2002, «Network motifs in the transcriptional regulation network of {E}scherichia coli.», *Nat Genet*, vol. 31, n° 1, doi :10.1038/ng881, p. 64–68. [144](#), [145](#), [146](#)
- SHEA, M. A. et G. K. ACKERS. 1985, «The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation.», *J Mol Biol*, vol. 181, n° 2, p. 211–230. [169](#)
- SIMONS, A. M., S. C. STEARNS, S. L. CHOWN et E. COLLAB. 2011, «Modes of response to environmental change and the elusive empirical evidence for bet hedging.», *Proceedings. Biological sciences / The Royal Society*, vol. 278, n° 1712, doi :10.1098/rspb.2011.0176, p. 1601–9, ISSN 1471-2954. [203](#)
- SOL, A. D., H. FUJIIHASHI et P. O’MEARA. 2005, «Topology of small-world networks of protein-protein complex structures.», *Bioinformatics*, vol. 21, n° 8, doi :10.1093/bioinformatics/bti167, p. 1311–1315. [141](#)
- SOLÉ, R. V. et S. VALVERDE. 2006, «Are network motifs the spandrels of cellular complexity?», *Trends in ecology & evolution*, vol. 21, n° 8, doi :10.1016/j.tree.2006.05.013, p. 419–22, ISSN 0169-5347. [149](#)
- STEIGE, K. A. et T. SLOTTE. 2016, «Genomic legacies of the progenitors and the evolutionary consequences of allopolyploidy», *Current Opinion in Plant Biology*, vol. 30, doi :10.1016/j.pbi.2016.02.006, p. 88–93, ISSN 13695266. [188](#)
- STEWART-ORNSTEIN, J., H. W. J. CHENG et G. LAHAV. 2017, «Conservation and Divergence of p53 Oscillation Dynamics across Species», *Cell Systems*, vol. 5, n° 4, doi :10.1016/j.cels.2017.09.012, p. 410–417.e4, ISSN 24054720. [153](#)
- STEWART-ORNSTEIN, J. et G. LAHAV. 2017, «P53 dynamics in response to DNA damage vary across cell lines and are shaped by efficiency of DNA repair and activity of the kinase ATM», *Science Signaling*, vol. 10, n° 476, doi :10.1126/scisignal.aah6671, p. eaah6671, ISSN 19379145. [153](#)
- STRAUS, D., W. WALTER et C. A. GROSS. 1990, «Dna{K}, {D}na{J}, and {G}rp{E} heat shock proteins negatively regulate heat shock gene expression by controlling the synthesis and stability of sigma 32.», *Genes Dev*, vol. 4, n° 12A, p. 2202–2209. [157](#)

- STRICKER, J., S. COOKSON, M. R. BENNETT et collaborateurs. 2008, «A fast, robust and tunable synthetic gene oscillator», *Nature*, vol. 456, n° 7221, doi :10.1038/nature07389, p. 516–519, ISSN 0028-0836. [151](#)
- SUTTON, M. D., B. T. SMITH, V. G. GODOY et G. C. WALKER. 2000, «The SOS response : recent insights into umuDC-dependent mutagenesis and DNA damage tolerance.», *Annual review of genetics*, vol. 34, n° 1, doi :10.1146/annurev.genet.34.1.479, p. 479–497, ISSN 0066-4197. [139](#)
- TOMPA, P., J. PRILUSKY, I. SILMAN et J. L. SUSSMAN. 2008, «Structural disorder serves as a weak signal for intracellular protein degradation.», *Proteins*, vol. 71, n° 2, doi :10.1002/prot.21773, p. 903–9, ISSN 1097-0134. [177](#)
- TURNER, T. 2004, «Stochastic approaches for modelling in vivo reactions», *Computational Biology and Chemistry*, vol. 28, n° 3, doi :10.1016/j.compbiolchem.2004.05.001, p. 165–178, ISSN 14769271. [159](#)
- TYSON, J. J., K. CHEN et B. NOVAK. 2001, «Network dynamics and cell physiology.», *Nat Rev Mol Cell Biol*, vol. 2, n° 12, doi :10.1038/35103078, p. 908–916. [150](#)
- TYSON, J. J. et B. NOVAK. 2001, «Regulation of the eukaryotic cell cycle : molecular antagonism, hysteresis, and irreversible transitions.», *J Theor Biol*, vol. 210, n° 2, doi :10.1006/jtbi.2001.2293, p. 249–263. [150](#)
- VEITIA, R., S. BOTTANI et J. BIRCHLER. 2008, «Cellular reactions to genomic imbalances», *Submitted to Trends in Genetics*. [175](#), [176](#), [177](#), [178](#), [180](#)
- VEITIA, R., F. VEYRUNES, S. BOTTANI et J. BIRCHLER. 2015, «X chromosome inactivation and active X upregulation in therian mammals : Facts, questions, and hypotheses», *Journal of Molecular Cell Biology*, vol. 7, n° 1, doi :10.1093/jmcb/mjv001, ISSN 17594685. [191](#)
- VEITIA, R. A. 2002a, «Exploring the etiology of haploinsufficiency.», *Bioessays*, vol. 24, n° 2, doi : 10.1002/bies.10023, p. 175–184. [176](#)
- VEITIA, R. A. 2002b, «Exploring the etiology of haploinsufficiency.», *BioEssays : news and reviews in molecular, cellular and developmental biology*, vol. 24, n° 2, doi :10.1002/bies.10023, p. 175–184, ISSN 0265-9247. [187](#)
- VEITIA, R. A. 2003, «Nonlinear effects in macromolecular assembly and dosage sensitivity.», *J Theor Biol*, vol. 220, n° 1, p. 19–25. [176](#), [188](#)
- VEITIA, R. A. 2005, «Gene dosage balance : Deletions, duplications and dominance», *Trends in Genetics*, vol. 21, n° 1, doi :10.1016/j.tig.2004.11.002, p. 33–35, ISSN 01689525. [173](#), [175](#)
- VEITIA, R. A. et S. BOTTANI. 2009, «Whole Genome Duplications and a ‘Function’ for Junk DNA? Facts and Hypotheses», *PLoS ONE*, vol. 4, n° 12, doi :10.1371/journal.pone.0008201, p. e8201, ISSN 1932-6203. [182](#), [184](#), [185](#)
- VEITIA, R. A., S. BOTTANI et J. A. BIRCHLER. 2007, «Cellular reactions to genomic imbalance», *Submitted to Trends in Genetics*. [194](#)
- VEITIA, R. A., S. BOTTANI et J. A. BIRCHLER. 2013, «Gene dosage effects : Nonlinearities, genetic interactions, and dosage compensation», *Trends in Genetics*, vol. 29, n° 7, doi :10.1016/j.tig.2013.04.004, p. 385–393, ISSN 01689525. [176](#), [178](#), [179](#), [181](#)
- VEITIA, R. A., D. R. GOVINDARAJU, S. BOTTANI et J. A. BIRCHLER. 2017, «Aging : Somatic Mutations, Epigenetic Drift and Gene Dosage Imbalance», *Trends in Cell Biology*, vol. 27, n° 4, doi :10.1016/j.tcb.2016.11.006, p. 299–310, ISSN 18793088. [193](#)

- WAGNER, A. et D. A. FELL. 2001, «The small world inside large metabolic networks.», *Proc R Soc Lond B Biol Sci*, vol. 268, n° 1478, doi :10.1098/rspb.2001.1711, p. 1803–1810. [141](#)
- WAHL, G. M., J. M. STOMMEL, K. KRUMMEL et M. WADE. 2005, «25 Years of P53 Research», chap. Chapter 4 :, Springer. [152](#)
- WALKER, G. C. 1995, «SOS-regulated proteins in translesion DNA synthesis and mutagenesis.», *Trends in biochemical sciences*, vol. 20, n° 10, p. 416–20, ISSN 0968-0004. [139](#)
- WENDEL, J. F., D. LISCH, G. HU et A. S. MASON. 2018, «The long and short of doubling down : polyploidy, epigenetics, and the temporal dynamics of genome fractionation», *Current Opinion in Genetics and Development*, vol. 49, doi :10.1016/j.gde.2018.01.004, p. 1–7, ISSN 18790380. [188](#)
- WRIGHT, S. 1932, «The roles of mutation, inbreeding, crossbreeding and selection in evolution.», URL <http://www.blackwellpublishing.com/ridley/classictexts/wright.pdf>. [195](#), [196](#), [197](#)
- ZAMAL, F. A. et D. RUTHS. 2012, «On the contributions of topological features to transcriptional regulatory network robustness», *BMC Bioinformatics*, vol. 13, n° 1, doi :10.1186/1471-2105-13-318, p. 318, ISSN 1471-2105. [143](#)
- ZATORSKY, N. G., N. ROSENFELD, S. ITZKOVITZ et collaborateurs. 2006, «Oscillations and variability in the p53 system.», *Mol Syst Biol*, vol. 2, doi :10.1038/msb4100068, p. 2006.0033. [150](#)

Troisième partie

Neurophysique

Chapitre 6

Approches physiques pour les cultures neuronales

« Les machines faites de la main de l'homme ne sont pas des cerveaux, mais les cerveaux sont une variété, très mal comprise, de machines computationnelles »

McCulloch, cité dans Pélissier et Tête 1995, p.189 et Alexandre Brunet chapitre 1.12

Sommaire

6.1 Présentation	220
6.1.1 Introduction du chapitre	220
6.2 Les cultures de neurones dissociées	222
6.2.1 Protocoles de cultures neuronales	222
6.2.2 Milieux de cultures	224
6.2.3 Interagir avec les cultures	228
6.2.4 Les réseaux de neurones des cultures de neurones	232
6.3 Applications des cultures de neurones	238
6.3.1 Applications biomédicales	239
6.3.2 Applications aux troubles neurologiques	239
6.3.3 Criblage moléculaire et biosenseurs	241
6.4 Vers des processeurs neuronaux	241
6.4.1 Codage neuronal et réseaux neuronaux	242
6.4.2 Robustesse du calcul neuronal	242
6.4.3 Approches de recherche	244
6.4.4 Réalisation de processeurs neuronaux	244
6.5 Activité spontanées des cultures	247
6.5.1 Synchronisation et bouffées de potentiels d'action	249
6.5.2 Avalanches et comportements invariants d'échelle	251
6.5.3 Dynamique neuronale et inhibition	254
6.6 Programmer les cultures neuronales	255
6.6.1 Codage neuronal et dynamique collective	255
6.6.2 Codage dynamique de l'information	255
6.6.3 Apprentissage dans les cultures	257
6.7 Modélisation des cultures neuronales	262
6.7.1 Modélisation de la structure des cultures	262
6.7.2 Modélisation de la croissance neuronale : état de l'art	263
6.7.3 Modélisation de l'activité des neurones	266
6.8 Conclusions et perspectives : vers la neurobiologie synthétique.	267
6.9 Références	268

6.1 Présentation

Les cultures cellulaires *in vitro* de cellules cérébrales permettent l'observation dans un milieu artificiel, hors de leur environnement original, de la croissance des neurones et du développement sur de longues périodes du réseau de leurs connexions, support des fonctions cognitives. Ces possibilités et les facilités des cultures cellulaires pour des manipulations et des traitements chimiques, en font depuis longtemps en neurosciences une plateforme très utilisée de recherches à visées pharmaco-médicales impossibles *in vivo* ou sur des prélèvements de tissus naturels.

L'application de techniques récentes de microfluidiques ouvre depuis une dizaine d'années la voie à de nouveaux usages des cultures neuronales en vue de la réalisation de dispositifs artificiels capables de calcul à base de cellules vivantes. Cette troisième partie « Neurophysique » de mon mémoire commence par le chapitre actuel, offrant une introduction assez développée aux cultures de neurones, leurs usages et les problématiques de nature cognitive, calculatoires, qui me motivent en premier lieu. Le second chapitre de cette partie, chap. 7, présentera mes travaux spécifiques dans ce second volet de mon champ d'activité. Pour prendre connaissance plus directement de mes recherches, il est possible de se reporter directement au chapitre suivant. On y trouvera une introduction abrégée qui résume les points principaux de la discussion qui va maintenant suivre sur les cultures neuronales.

6.1.1 Introduction du chapitre

Les neurosciences ont pour objectif fondamental la compréhension des fonctions cognitives, c'est-à-dire des capacités de traitement de l'information, d'apprentissage, de mémorisation et du contrôle de l'activité musculaire des animaux. Un corpus considérable de connaissances anatomiques du cerveau est disponible de nos jours avec la connaissance des structures du cerveau sur toutes les échelles, depuis les cellules individuelles à celle de l'organe tout entier. Différents projets de recherche, discutés plus loin, sont en effet en cours pour reconstituer l'organisation spatiale de cerveaux entiers avec une résolution spatiale au niveau des neurones eux-mêmes, comprenant aussi la caractérisation biologique détaillée des types de neurones et cellules gliales dans les différentes zones. En sus de l'anatomie, de nombreuses techniques d'imagerie dite *fonctionnelle* permettent d'observer l'activité électrique cérébrale portée par les neurones avec des résolutions spatiales et temporelle de plus en plus élevées.

Les neurosciences sont clairement un des domaines scientifiques les plus en vogue de cette première partie de siècle, largement soutenus par les enjeux médicaux mais également certainement par la fascination publique actuelle pour les progrès de l'intelligence artificielle de l'industrie informatique. Cependant, malgré les énormes avancées des neurosciences dans tous ses sous domaines, la question fondamentale de la relation entre matière (la biologie) et esprit (la psychologie) reste largement irrésolue. L'étude de cette interrogation centrale sur notre essence d'être vivants fait appel aux contributions de la plupart des disciplines scientifiques, croisant leurs points de vue et approches. Il en est forcément ainsi car la complexité du système nerveux fait intervenir un mélange interdépendant d'échelles et de phénomènes : molécules, cellules, organisme, psychologie et physiologie. Ce questionnement interpelle donc les scientifiques d'horizons très différents, mathématiciens, informaticiens, physiciens ainsi que naturellement biologistes, psychologues et médecins.

De très nombreuses études du cerveau poursuivent la description détaillée des différentes structures et propriétés cellulaires et moléculaires, motivées en grande partie par leurs enjeux médicaux. D'autres démarches vont plutôt s'intéresser aux principes fondamentaux de constitution de l'organe cérébral et de ses éléments qui lui confèrent ses capacités remarquables de traitement d'information. Il s'agit de comprendre la faculté d'interaction active des êtres vivants avec l'environnement en vue de favoriser leur reproduction. Une capacité qui implique à la fois la perception des caractéristiques du milieu ambiant et la détermination d'une ligne d'action optimale

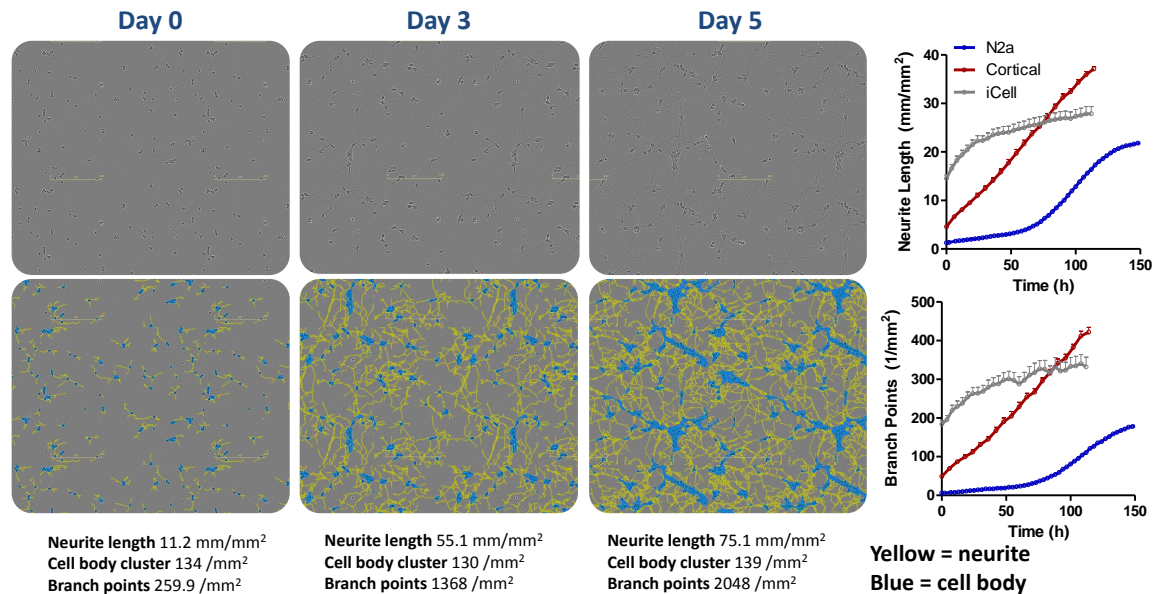


FIGURE 6.1 – **Exemple de réseau de neurone en culture.** Gauche : Photo en microscopie par contraste de phase (dessus) et reconstruction informatique avec le logiciel NeuroTrack (en dessous) des neurites pour des cultures primaires de neurones corticaux à différents jours après l’incubation des cellules. Droite : évolution au cours du temps des longueurs totales par unité de surface des neurites et du nombre de branchements par unité de surface pour des cultures corticales de rat, de neurones de cellules souches humaines (iCell), et neurones issues de lignées immortalisées (N2a). Ces dispositifs de cultures et systèmes de caractérisation sont commercialisés par la société IncuCyte comme plateformes de recherche sur les troubles neurologiques, la toxicité, la régénération et la différenciation neuronales [ALCANTARA et collab.; ESSEN BIOSCIENCE, 2018]. Source : [ALCANTARA et collab.]

suivant les conditions. En d’autres termes encore, la propriété de réception et collecte de signaux en entrée, ainsi que la propriété du système à réaliser des opérations pour décider en fonction des entrées perçues d’une réaction qui affectera un retour de l’organisme sur l’environnement. Il s’agit de comprendre ici les principes logiques suivant lesquels un système neuronal effectue des calculs. Une faculté qui outre le calcul lui-même nécessite la perception des entrées et la manifestation en sortie d’un résultat.

Chez les animaux ces opérations sont portées par l’activité collective des réseaux de neurones interconnectés dans des structures anatomiques bien identifiées, résultant de l’excitabilité des neurones individuels et de la communication synaptique entre eux¹. La question centrale des neurosciences est donc de comprendre comment les propriétés calculatoires du cerveau dépendent de la structure des réseaux neuronaux.

La question fondamentale des neurosciences est ainsi d’identifier, de décrire les structure et les activités qui se déroulent dans le cerveau pour ensuite donner sens aux relations entre la composition et l’organisation des circuits neuronaux et leurs propriétés fonctionnelles. Les recherches typiques portent sur l’observation de cerveaux évolués, des structures cérébrales naturelles, fruit de millions d’années de sélection naturelle. Elles portent sur l’observation d’objets très sophistiqués et délicats avec des possibilités d’interventions expérimentales sur les réseaux étudiés relativement limitées.

L’étude de cultures neuronales est une approche complémentaire aux recherches sur le cer-

1. Fondamentalement, suivant la définition donnée d’un système cognitif on peut aussi parler de systèmes cognitifs cellulaires impliquant récepteurs membranaires, voies de transduction de signaux intracellulaires et interrupteurs génétiques sélectionnant une réponse métabolique. Des idées développées notamment par les philosophes et biologistes Humberto Maturana et Francisco Varela [MATURANA et VARELA, 1994] et qui constituent aussi un lien conceptuel entre cette partie sur les neurosciences et la partie précédente sur la biologie des systèmes.

veau, permettant d'élargir le champ d'exploration et d'explorer les propriétés issues de structures de réseaux de neurones très différentes de celles naturelles. Les cultures de neurones, dont des exemples sont présentés dans la figure 6.1, sont des outils établis de recherche sur les troubles neurologique et les effets de drogues et toxines sur l'activité et le développement neuronal. Quoique plus rarement, ces dispositifs trouvent aussi des applications dans le champs des recherches cognitives qui m'intéressent principalement.

Alors que les structures naturelles ont pu se développer par la sélection naturelle avec des propriétés très particulières, peut-être spécialisées et optimisées, l'étude de réseaux artificiels dont la structure ne résulte pas de l'histoire évolutive peut aider à mettre en évidence les caractéristiques indispensables aux processus calculatoires cognitifs. L'espoir est qu'en manipulant les structures, on puisse trouver les conditions qui leur permettent d'assurer des fonctions cohérentes. C'est-à-dire identifier l'organisation et la nature des réseaux conduisant à des propriétés calculatoires.

Les cultures neuronales sont ainsi des tissus vivants, actifs et plastiques qui offrent dont des conditions fascinantes de pouvoir suivre le développement propre d'un réseau neuronal, son auto-organisation, tout en observant sa dynamique d'activité électrique. Dans la suite de ce chapitre je présenterai plus en détail ces objets d'étude, les questions qui s'y rapportent et leurs perspectives.

6.2 Les cultures de neurones dissociées

La capacité de maintenir vivantes et de faire se développer des cellules de neurones hors de leur tissu naturel est ancienne [MILLET et GILLETTE, 2012; THOMPSON, 1977]. L'histoire de cette technique est racontée par Millet et Gillet [MILLET et GILLETTE, 2012] qui expliquent comme dès son introduction en 1907 par Harrison [HARRISON et collab., 1907], l'isolation de neurones *in vitro* et l'observation de leur développement au microscope a permis une des avancées les plus fondamentales en neuroscience : la confirmation de manière définitive de la nature cellulaire du système nerveux [HARRISON, 1910]. Les structures en réseau très complexes que l'on observait au microscope sont bien constituées d'une multitude de cellules individuelles différentes interconnectées, et non pas, comme il était possible de le supposer un objet unique continu tel un filet. Contrairement à d'autres cellules d'animaux, les neurones sont ramifiés sur de très grandes longueurs par rapport à leur corps cellulaire. Les neurones adultes ne peuvent donc généralement pas être extraits et isolés à partir de biopsies sans destruction. En conséquence leur culture est délicate et demande des techniques différentes d'autres types cellulaires.

6.2.1 Protocoles de cultures neuronales

Les protocoles pour des cultures de neurones sont de nos jours assez bien définis même si chaque groupe expérimental a ses recettes, dont des détails peuvent affecter les comportements. La mise en culture implique l'incubation dans un milieu nutritif de cellules précurseurs de neurones, d'origine soit embryonnaire, soit provenant de cellules souches.

Les protocoles utilisés par mes collaborateurs à l'Institut Weizmann et à l'Institut Curie sont basés sur des cellules de cerveaux d'embryons de rats ou souris prélevés par dissection. Des cellules dites *primaires* car ne résultant pas d'une division cellulaire post prélèvement. Les cultures sont préparées typiquement à partir de neurones des régions du cerveau spécifiques comme l'hippocampe ou le cortex. D'autres auteurs font usage de cultures de neurones d'invertébrés, comme la sauterelle dont la grande taille des cellules facilite les observations et les cultures [DE SANTOS-SIERRA et collab., 2014; SHEFI et collab., 2002b], mais dont les propriétés sont différentes de celles des rongeurs. La discussion ci-dessous concerne plutôt les cultures à bases de neurones primaires de souris et de rats telles que pratiquées par les équipes avec lesquelles je collabore.

Au stade embryonnaire quand la dissection est réalisée (embryons de 18 ou 19 jours chez la souris) les neurones sont encore à un stade précoce avec une forme ramifiée peu développée.

Une fois prélevés ces neurones précoces sont incubées typiquement sur du verre et baignés dans une solution nutritive. Les neurones sont dits dissociés car les cellules sont indépendantes; les ramifications et connexions synaptiques s'établiront avec la croissance des neurites. Ces cultures sont dites aussi *primaires* car les neurones présents sont ceux du prélèvement, sans qu'il y ait prolifération cellulaire. À l'âge pratique pour la dissection il n'y a en effet plus de division cellulaire des neurones. Le réseau neuronal se développe avec les neurones initialement incubés qui connaissent uniquement une mortalité.

Une fois incubés, le développement des neurones commence sans retard avec la croissance progressive des neurites dont l'une évoluera en axone. Un élément essentiel pour l'usage des cultures est la préservation bien attestée des caractéristiques physiologiques et morphologiques des neurones qui correspondent à celles des tissus originaux [KRIEGSTEIN et DICHTER, 1983]. Les connexions entre neurones vont aussi commencer à s'établir à peine quelques heures après l'incubation avec la formation des premières synapses. Si les caractéristiques des cellules individuelles semblent correspondre à celles des tissus d'où elles sont extraites, la structure émergente du réseau en culture est cependant différente, sans rapport avec celle qui se serait développé dans le tissu original. En culture les connexions semblent s'établir de manière aléatoire. Le réseau et les synapses croissent sans ordre, suivant des principes différents encore mal caractérisés. Sans un plan d'organisation ou un système de guidage qui existerait dans les cerveaux. Cependant, ce réseau *anormal* n'empêche pas l'apparition 3 à 4 jours après l'incubation d'une activité électrique spontanée au sein des neurones qui commence à se propager sur le réseau au fur et à mesure de sa croissance.

Pour compléter la description des protocoles de cultures il est nécessaire de préciser qu'au même temps du prélèvement des neurones primaires, on extrait également dans la même opération des cellules gliales avec lesquelles ils sont mélangés. Ces cellules, ramifiées elles aussi occupent en effet pour près de 50% du volume cérébral. Elles jouent un rôle essentiel à la maintenance de la bonne activité du cerveau et sont constitués de oligodendrocytes (près de 76%), microglies (7%) et astrocytes (17%). Les oligodendrocytes qui interviennent dans la myélination des axones et les microglies qui constituent le mécanisme de résistance immunitaire du cerveau sont peu présentes dans les cultures. Les astrocytes participent à l'équilibrage des concentrations ioniques, notamment de calcium, du potassium dans le milieu extra cellulaire et au recyclage des neurotransmetteurs synaptiques. Ils sont présents dans le milieu extra cellulaire et peuvent aussi continuer à proliférer. Malgré la présence attestée et le rôle significatif des astrocytes, leur caractérisation, fraction et interactions avec les neurones des cultures sont très mal caractérisés et mériteraient plus d'études; la littérature n'est pas très riche sur cette question [KIM et MAGRANÉ, 2011] et les discussions avec les collaborateurs expérimentalistes confuses à ce sujet².

Hors la morphologie, les cultures répliquent à plusieurs échelles les comportements cérébraux naturels. Les neurones en cultures préservent de nombreuses propriétés morphologiques, pharmacologiques et électriques [DICHTER, 1978; KRIEGSTEIN et DICHTER, 1983]. La part de neurones excitateurs et inhibiteurs y est équivalente au contexte naturel, ainsi que les propriétés observées des récepteurs, canaux ioniques et de libération de neurotransmetteurs. De composition identique et de propriétés individuelles des éléments égales à celles *in vivo*, les cultures offrent ainsi un modèle fonctionnel des zones cérébrales dont la caractéristique principale semble juste être son absence de structure naturelle. Les neurones en cultures constituent donc des modèles cellulaires valables pour des applications de recherche variées [ROBERT et collab., 2012]. Cependant même si les caractéristiques générales des neurones en cultures semblent similaires à celles de neurones dans des tissus, leur développement est soumis à un déroulement de contrôles génétiques et environnementaux différents dans les cultures que dans les cerveaux. Ainsi, l'activation différente en fonction du temps et de l'environnement d'interrupteurs génétiques de choix de développement conduit certainement à des différences dans la croissance des réseaux, leur structure sur le long

2. Des protocoles existent cependant pour cultiver plus spécifiquement de cellules gliales [ALBUQUERQUE et collab., 2009; SKAPER et collab., 2012] pour des études qui leurs sont consacrées.

terme et de leur dynamique.

Alors que la croissance des neurones peut se poursuivre longtemps après l'incubation, on considère généralement que les cultures de neurones issues de rongeurs ont atteint une maturité au bout d'environ deux à trois semaines après incubation, quand l'essentiel des connexions semble s'être établie [CHIAPPALONE et collab., 2006; VAN PELT et collab., 2004b]. La plupart des expériences publiées ont lieu après cette date et se poursuivent sur une période typique de trois à quatre semaines. Même si les cultures peuvent rester actives très longtemps au-delà, celles-ci sont facilement sujettes à des contaminations et des périodes d'observations plus longues restent difficiles.

Je mentionnerai enfin que sur un long terme les cultures connaissent un vieillissement sous forme de déplacements et regroupement des corps cellulaires et des neurites. Ces restructurations complexes dans des vieilles cultures sont dues à l'évolution des interactions entre les différents éléments et notamment avec le substrat et ont fait l'objet d'encore peu d'études. Cette évolution serait intéressante à étudier, car sur le long terme elle pourrait permettre de mettre en évidence des caractéristiques fines des neurones particulièrement dans des contextes pharmaco-médicaux.

Je n'ai pas discuté ici de protocoles de culture impliquant des cellules souches. Ceux-ci mettent en œuvre des cellules reprogrammées pour se développer en neurones précurseurs et permettent de réaliser des cultures avec des neurones de type humains. Ce type de cultures n'est usuellement pas utilisés par mes collaborateurs en raison d'une part du coût élevé des cellules nécessaires, ainsi que d'une plus faible activité généralement observée. Les cultures de neurones primordiaux de rongeurs, de cellules souches humaines ou à partir de neurones d'invertébrés [AMERICAN FRIENDS OF TEL AVIV UNIVERSITY, 2011] offrent encore de nombreux avantages pour construire des modèles neuronaux de recherches appliquées ou fondamentales. Elles sont un outil répandu pour de nombreuses applications qui seront discutées plus loin. Auparavant pour appréhender ces méthodes et leurs applications il est encore nécessaire d'introduire les divers dispositifs techniques pour accueillir les cultures et interagir avec elles.

6.2.2 Milieux de cultures

Les dispositifs disponibles actuellement pour réaliser les cultures bénéficient depuis une dizaine d'années des technologies de *microfabrication* de la *microfluidique* qui sont encore en train de révolutionner les méthodes expérimentales en biologie cellulaire. La possibilité de structurer le milieu de culture avec des compartiments et des formes particulières a en effet ouvert de nombreuses applications aux cultures cellulaires de tous types et en particulier de neurones. Elles permettent par exemple, de tenter de reproduire *in vitro* des architectures cérébrales mimant celles observées *in vivo*, permettant une plus grande facilité et champ d'étude de l'activité neuronale dans des structures modulaires.

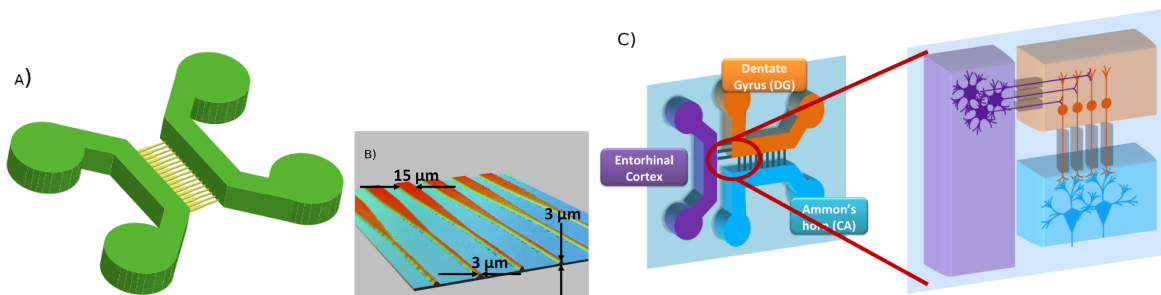


FIGURE 6.2 – Exemples structures microfluidique pour neurones : A) Schéma conçu sur ordinateur de différentes chambres interconnectées par des microcanaux. B) Moule positif de microcanaux asymétriques. C) Système développé au laboratoire de Jean-Louis Viovy à l'Institut Pierre-Gilles De Gennes pour reproduire sur une puce trois couches cérébrales successives comme modèle d'étude de la propagation de la maladie d'Alzheimer. Source : [VIGNES, 2013]

En réalité les premières méthodes pour créer des substrats avec différents compartiments de culture ont été introduites dès les années 70. Il s'agissait déjà de forcer la croissance des neurones dans des environnements imposant des structures pré-définies pour contraindre l'architecture globale des réseaux tout en accédant aux neurones des différentes régions (voir la discussion historique dans [MILLET et GILLETTE, 2012]).

Microfluidique et micro-impression

Les techniques issues de la microfluidique découplent et facilitent les possibilités de structurer les milieux de culture cellulaires. Le point de départ de ces techniques est l'usage de méthodes de photolithographie permettant de réaliser par gravure chimique sur des substrats photosensibles des structures creuses à partir de plans conçus sur ordinateur suivant l'imaginaire des chercheurs, figure 6.3. Cette étape qui s'apparente à un développement photographique classique permet la production d'une puce avec des structures sub-millimétriques gravées avec précision mais est relativement laborieuse et onéreuse. La microfluidique s'est surtout développée depuis qu'elle a été complétée par différentes approches pour faciliter la multiplication des dispositifs produits. Le substrat gravé par photolithographie peut en effet être avantageusement utilisé de nombreuses fois comme patron pour la réplique des structures, soit en tant que moule, soit en tant que tampon d'impression.

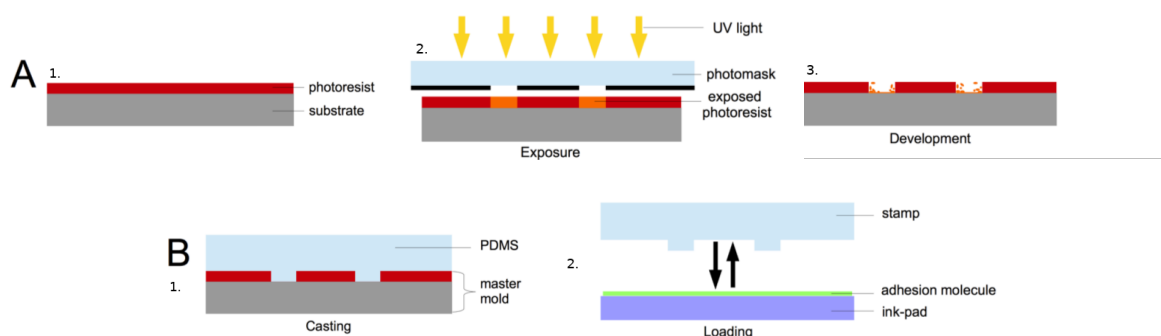


FIGURE 6.3 – **Méthodes de microfabrication. Principe de gravure photolithographique.** A) Préparation d'un moule principal. Un substrat solide est recouvert d'une couche de résine photosensible qui est illuminée au travers d'un masque portant le dessin des structures désirées. En fonction de la résine photosensible les zones illuminées ou non deviennent solubles, permettant de les dissoudre et exposer le substrat susjacent tout en laissant les autres zones protégées par la résine. B) La structure en résine issue de l'étape précédente sert de masque principale pour la reproduction, en négatif, de dispositifs souples par le moulage avec un polymère souple, le polydiméthylsiloxane (PDMS). Les dispositifs souples obtenus peuvent être directement utilisés en exploitant les structures sous forme de reliefs. Alternativement ils peuvent être utilisés comme tampons pour transférer des molécules adhésives ou répulsives sur un substrat. Source : [VIGNES, 2013]

Dans la technique de moulage le substrat, solide, issu de la photolithographie, sert de moule pour la réplique en négatif des structures en lui coulant dessus un gel de polymères initialement liquide et le laissant polymériser. Une fois démoulés, ces dispositifs souples portent en relief les structures à l'échelle microscopique du modèle initial. Les chambres et canaux creusés dans le substrat permettent la circulation de fluides matérialisant des réacteurs chimiques et des bioréacteurs à des échelles sub-millimétriques³. Par leur miniaturisation et intégration dans des systèmes de contrôle informatisés, ces dispositifs offrent la possibilité de manipuler précisément les conditions biochimiques de développement spatial et temporel de cellules qui y sont mises en culture.

Alternativement le substrat issu de la photolithographie peut-être utilisé comme tampon d'impression, pour déposer sur un substrat plan un dépôt de molécules répulsives ou attractives pour les cellules suivant les motifs du patron modèle [COREY et FELDMAN, 2003; MILLET et collab.,

3. On parle aussi de laboratoire sur une puce; le nom justement de la principale revue du domaine « Lab on a chip » lancée en 2001.

2010; TOMBA et collab., 2014] (en anglais techniques de « patterning », impression de patrons). Ces dispositifs imprimés permettent la fixation de cellules à des localisations très précises, mais contrairement aux structures en relief des moulures, ne matérialisent des structures confinant des fluides et une architecture de bioréacteurs. Les différentes techniques de microfabrication, moulage de canaux et impression de patrons adhésifs ou répulsifs peuvent aussi être combinées pour un contrôle encore plus précis [YAMADA et collab., 2016].

Appliquée aux cultures neuronales, la microfluidique ouvre tout particulièrement de nombreuses perspectives d'applications liées à la morphologie complexe étendue sur de grandes zones spatiales de ce type de cellules. De nombreuses revues [GROSS et collab., 2007, 2005; SOE et collab., 2012; WANG et collab., 2009] présentent le développement important de ces techniques en neuroscience et l'ouvrage Emilia Biffi « Microfluidic and compartmentalized platforms for neurobiological research » de 2015 [BIFFI, 2015] donne une revue encore d'actualité de la richesse des techniques disponibles maintenant pour des applications de cultures neuronales. Les dispositifs microfluidiques ou imprimés permettent ces études avec des systèmes constitués de plusieurs compartiments interconnectés de différentes manières et donc d'organiser les cultures dans des zones différenciées fonctionnellement. Au sein des zones accessibles les neurones se développent spontanément pour construire des réseaux désordonnés mais contraints par la géométrie et l'interaction des cônes de croissance avec l'environnement. Certains systèmes microfluidiques ou imprimés visent même un contrôle au niveau de la cellule unique et du plan de câblage précis du réseau des connexions par la localisation précise des corps cellulaires et des directions de croissance des neurites [PARK et collab., 2014; ROTH et collab., 2012a,b; STENGER et collab., 1998; TAYLOR et collab., 2005; TOMBA et collab., 2014; WHEELER et BREWER, 2010].

Cependant les mécanismes de croissances de neurones dans ces différents systèmes restent encore très mal connus et les développements de solutions technologiques opérationnelles restent exploratoires. Les dispositifs microfluidiques mettent en évidence le rôle de l'environnement physique pour le développement des neurones [BUGNICOURT et collab., 2014; ROTH et collab., 2012a] dont le mécanisme moteurs doivent agir aussi *in vivo*. Des modèles théoriques manquent pour comprendre les processus de croissances des réseaux dans ces différents contextes. Je reviendrai sur cette question dans le chapitre 7, section 7.4 présentant mes travaux visant à contribuer à cette problématique. Pour des designs complexes il s'avère en effet difficile de prédire comment la croissance des axones va se comporter. De nombreuses recherches expérimentales s'efforcent ainsi de concevoir des designs de dispositifs permettant un guidage fiable des axones et la production de réseaux bien reproductibles. C'est notamment le cas des travaux au sein de l'équipe de Jean-Louis Viovy et de Catherine Villard à l'Institut Curie et l'Institut Pierre Gilles de Gènes pour la microfluidique et de retombées de la thèse en co-direction avec moi de Renaud Renault [RENAULT, 2015; RENAULT et collab., 2016, 2015]. Une de mes ambitions est d'aider dans ce sens, avec une meilleure compréhension théorique des mécanismes de croissance et le développement d'outils de modélisation et simulations de ces systèmes.

Vers des cultures 3D

En dépit de leurs avantages importants d'accessibilité et d'observabilités les cultures planes sont géométriquement très différentes des systèmes neuronaux *in vivo* organisés dans l'espace à trois dimensions.

Dans une culture plane le corps cellulaire et les cônes de croissance sont anormalement aplatis et la croissance des neurites bloquée dans une fine couche. Quoique des croisements et des enchevêtrements sont possibles la morphologie de réseaux plans est limitée. En conséquence les possibilités d'interactions avec d'autres neurones, les cellules gliales et l'environnement sont réduites par rapport à un réseau pouvant se développer à trois dimensions [CULLEN et collab., 2011] et les comportements collectifs des réseaux peuvent aussi être affectés par la différence de dimensions [ULLOA SEVERINO et collab., 2016].

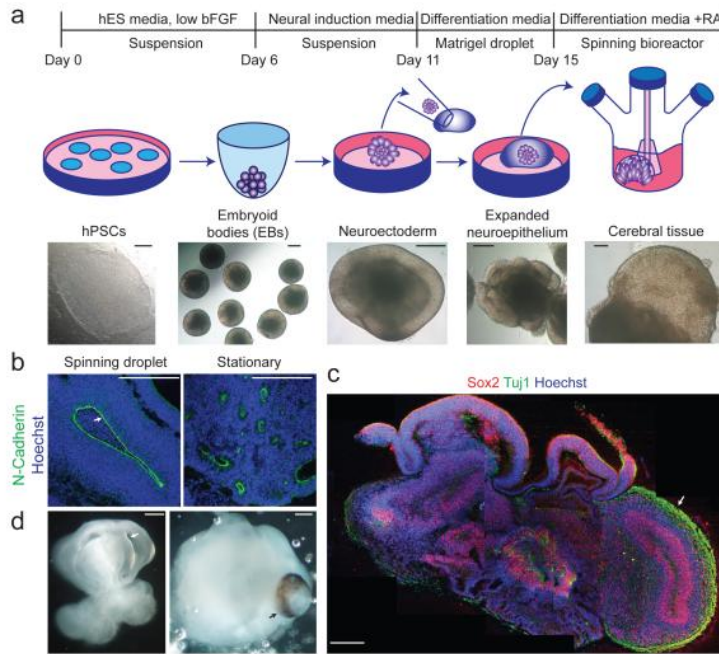


FIGURE 6.4 – **Description d'un système de culture d'organoïdes cérébraux.** a : Schéma du système de culture, partant initialement de cellules souches humaines pluripotentes (hPSCs) ; la suite d'images représente les étapes successives. b-d : images relevant plusieurs traits morphologiques des tissus qui se sont développés notamment des plissements caractéristiques de cerveaux. Échelle : 200 μ m. Source : [LANCASTER et collab., 2013]

Dans sa thèse publiée FREGA [2016b] discute page 37 comment pour un certain nombre d'auteurs les réseaux à deux dimensions sont trop lointains des situations naturelles pour démontrer des caractéristiques suffisantes pour être réellement utiles pour l'étude de systèmes cérébraux. L'introduction de modèles *in vitro* 3D permettrait alors de se rapprocher de conditions de croissance physiologique, de mieux reproduire les architectures cérébrales naturelles et de conduire à des comportements complexes.

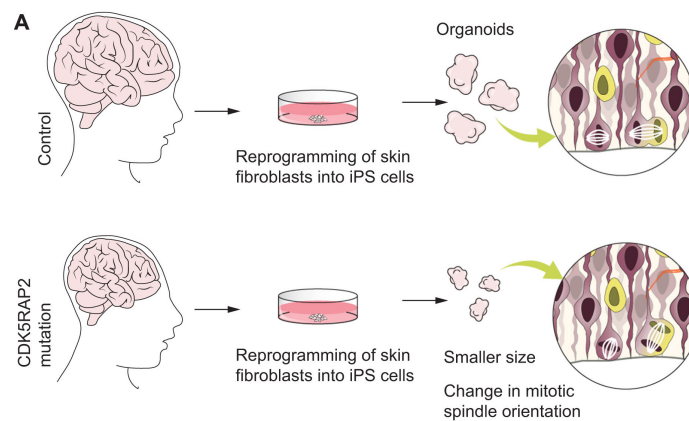


FIGURE 6.5 – **Organoïdes cérébraux comme outil d'étude de désordres neurologiques** Utilisation d'organoïdes cérébraux par [LANCASTER et collab., 2013] pour étudier le développement du cerveau d'un patient souffrant de microcéphalie portant une mutation du gène CDK5RAP2 qui lui est associé. Les organoïdes produits à partir du patient étaient plus petits que le contrôle et montraient des changements d'orientation des faisceaux de division mitotiques qui pourraient être reliés à une abondance réduite de cellules progénitrices et au développement d'un cerveau de taille réduite. Source : [KELAVA et LANCASTER, 2016]

Pour ces raisons, des recherches sont en cours à plusieurs endroits sur des méthodes de production de cultures de neurones à 3 dimensions. Une orientation qui a notamment connu une certaine médiatisation à partir de 2013 [LANCASTER et collab., 2013] avec la première publication dans Nature d'une première culture 3D de neurones issus de cellules souches humaines. Ces objets qui ont été dénommés *organoïdes cérébraux*, « cerebral organoids », présentent un développement de plusieurs régions neuronales discrètes mais liées reproduisant certains traits de différenciation

du développement du cerveau humain, figure 6.4. Cette première étude consistait essentiellement dans le développement d'un protocole de culture de cellules souches vers un tissu de cellules neuronales, sans aucune sophistication d'architecture visée. Elle met en évidence des capacités d'auto-organisation de cellules souches pour la formation de tissus et organes qu'il serait intéressant d'étudier d'un point de vue physique dans la ligne des travaux de morphogenèse discutés au chapitre 4. Depuis cette première publication en 2013 surtout on assiste à une vague de travaux sur les cultures neuronales à trois dimensions particulièrement comme systèmes d'étude de maladies du cerveau [KELAVA et LANCASTER, 2016; LEE et collab., 2017; VAN PEL et collab., 2018; WANG, 2018; YAN et collab., 2018], voir figure 6.5.

Différentes approches sont aussi recherchées pour structurer les cultures à 3 dimensions, via par exemple l'impression 3D de cellules [ONG et collab., 2018], ou la mise en culture dans des milieux dotées d'échafaudages dans l'espace [TORRE et CHEN, 2018]. Bien que ces objets implémentent la bonne dimensionnalité dans l'espace les organoïdes cérébraux présentent de nombreuses difficultés, en particulier l'accès aux nutriments et aux traitements pour les cellules internes limitent à l'heure actuelle les possibilités de structuration et d'usage. Les applications envisagées des organoïdes cérébraux comme plateformes de comparaison entre situations pathologiques et saines, tel qu'illustré sur la figure 6.5 sont très analogues sur le principe aux études comparatives avec des cultures planes, mais avec des complications méthodologiques supplémentaires significatives. Pour ces raisons les cultures de neurones à deux dimensions restent des dispositifs d'étude privilégiés offrent des possibilités d'interaction avec les cellules uniques.

6.2.3 Interagir avec les cultures

Comme déjà mentionné un des intérêts des cultures neuronales réside dans les possibilités d'excitation, de modulation et d'observation de leurs activité. Différentes techniques de stimulation et d'enregistrement de l'activité électrique venant d'autres contextes des neurosciences peuvent y être appliquées, possédant chacune différents avantages et inconvénients. Il s'agit d'une part de méthodes électrophysiologiques qui mesurent directement l'activité électrique, et d'autre part des méthodes indirectes optiques basées sur la chimie et la fluorescence.

Méthodes électrophysiologiques

Les neurones en culture sont accessibles à des études électrophysiologiques classiques à l'aide de microélectrodes précisément placées de part et d'autre de la membrane d'une cellule. Pour un plus grand nombre de mesures des systèmes parallèles comme les matrices multi-électrodes (multi-electrode array - MEA) sont possibles, voir la figure 6.6. Consistant en un arrangement de multiples microélectrodes ces dispositifs permettent d'étendre à près d'une centaine de neurones des stimulations et enregistrements électriques directs [GROSS, 2011]. Les MEA sont particulièrement intéressants pour les études fonctionnelles car ils permettent des enregistrements à haute résolution temporelle de potentiels d'action individuels [WHEELER et collab., 2011].

La mesure électrophysiologique de la variation temporelle de la différence de potentiel membranaire caractéristique d'un potentiel d'action demande une mesure au travers de la membrane neuronale. Ce type de mesure électrophysiologique nécessite donc deux électrodes l'une à l'extérieur de la cellule et l'autre à l'intérieur, demandant donc de traverser la membrane. Ces mêmes électrodes peuvent également stimuler les neurones et déclencher une dépolarisation en injectant un courant. La parallélisation de telles mesures est donc délicate et ces méthodes traditionnelles dites « patch clamp » ne permettent alors d'observer qu'une ou quelques cellules.

Alternativement à la mesure directe du potentiel trans-membranaire on peut sonder les variations de potentiel externes au neurone à proximité immédiate de la membrane, dues aux changements de concentration ioniques extracellulaires résultant des courants trans-membranaires entrants et sortants lors des potentiels d'actions. Pour cette mesure une électrode doit être localisé

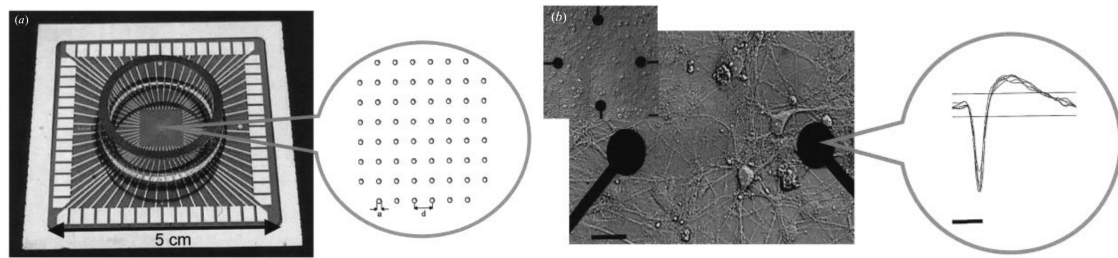


FIGURE 6.6 – **Culture de neurones avec matrice d'électrodes.** Matrice d'électrodes sur une culture de neurones. Photo en haut à gauche : 4 électrodes de diamètre $30\ \mu\text{m}$ sont visibles comme tâches circulaires noires ainsi que leur connectique; les somas des neurones sont visibles répartis dans le plan. Photo principale : agrandissement autour de deux électrodes; les neurones et leurs neurites à proximité sont visibles, la barre donne l'échelle de $30\ \mu\text{m}$. Les électrodes enregistrent le potentiel des neurones dans leur voisinage immédiat. Un exemple d'enregistrements alignés de potentiels d'action est indiqué dans le cercle; l'échelle temporelle est de 1 ms. Source : [MAROM et SHAHAF, 2002]

tout proche de la membrane et la seconde servir de référence en étant placée au loin. En changeant la tension l'électrode extracellulaire peut induire des courants ioniques trans-membranaires de canaux sensibles au voltage et permettre ainsi la stimulation des neurones.

Ce type de mesures est non invasif et permet la parallélisation en multipliant les électrodes pour sonder simultanément les champs de potentiel à l'extérieur des neurones qui leur sont les plus proches. Il est à la base des matrices multi-électrodes, développées à partir des années 1970 [GROSS et collab., 1977; THOMAS et collab., 1972] avec des systèmes de l'ordre d'une trentaine de points de mesure. Jusqu'au début des années 2000 le développement de cette technique a cependant été assez restreint en raison notamment des difficultés de fabrication des électrodes avec les propriétés nécessaires et de leur isolation dans un substrat permettant l'expérimentation [LIU et collab., 2012]. Les matrices d'électrodes offrent une très grande résolution temporelle et étant non invasives permettent des enregistrements sur de longues périodes.

Les techniques actuelles issues des progrès de microfabrication et de la microélectronique comportent typiquement de 60 à 256 électrodes de 10 à $30\ \mu\text{m}$ de diamètre espacées de 100 à $500\ \mu\text{m}$ [FREGA, 2016a; PINE, 2006] pour lesquelles existe une offre commerciale. Cependant malgré de progrès significatifs pour augmenter leur résolution, le nombre réduit d'électrodes et la petite couverture en surface limitent le nombre de neurones d'une population pouvant être suivi. Elles restent en outre relativement coûteuses et de mise en œuvre délicate notamment si elles doivent être couplées avec des chambres de culture multi-compartiments (des réalisations dans ce sens sont rapportées dans [FREGA, 2016b] p. 35-36).

Méthodes optiques

Alternativement à la mesure électrique directe des méthodes optiques permettent d'interagir avec les neurones en culture pour exciter et enregistrer indirectement l'activité neuronale proposant des performances complémentaires aux matrices d'électrodes.

Un nombre croissant de ressources et de méthodes sont disponibles pour observer les propriétés électrophysiologiques de neurones de manière optique. Plusieurs types de marqueurs fluorescents qui délivrent des signaux de lumière en présence d'activité neuronale ont ainsi été développés. Il s'agit de rapporteurs fluorescents sensibles soit au voltage (« voltage sensitive dyes ») ou bien sensibles à la présence intracellulaire calcium [BAKER et collab., 2005]. De tels marqueurs permettent des mesures d'activité sur un grand nombre de neurones à la fois par imagerie de fluorescence. On pourra se rapporter au chapitre « Visualising Neuronal Function » de la seconde édition de 2015 manuel de référence « Guide to Research Techniques in Neuroscience » de Matt Carter et Jennifer Shieh [CARTER et collab.] pour une introduction complète à ces techniques qui ont connu un fort développement, je rapporterai ici pour complétude uniquement des principes

de base.

Les rapporteurs sensibles au voltage sont des protéines membranaires fluorescentes dont l'absorption ou l'émission subit un décalage en fonction du potentiel transmembranaire, permettant de mesurer l'état électrique des neurones. Elles permettent ainsi de mesurer non seulement les potentiels d'action mais également la variation du potentiel sous le seuil, ce qui n'est pas possible par des mesures extracellulaires comme avec les matrices d'électrodes [BAKER et collab., 2005]. Il existe une grande variété de marqueurs sensibles au voltage avec des propriétés différentes en termes de signal et de toxicité [CARTER et SHIEH, 2015b]. Jusqu'à récemment ces rapporteurs demandaient cependant des préparations pas toujours reproductives et régulières des tissus à observer pour faire adsorber les rapporteurs dans les cellules, et surtout manquant de sélectivité des types cellulaires cibles. Des rapporteurs sensibles au voltage codés génétiquement et permettant plus de ciblage et de régulation existent seulement depuis peu [SEPEHRI RAD et collab., 2017].

L'imagerie calcique est une autre approche pour révéler par fluorescence visible au microscope l'activité neuronale grâce au flux entrant de calcium dans un potentiel d'action [SMETTERS et collab., 1999]. Les ions calcium jouent en effet de nombreux rôles cellulaires notamment comme messagers intracellulaires affectant la régulation génétique et dans les neurones pour contrôler la libération de neurotransmetteurs à partir des vésicules synaptiques. La concentration de calcium intracellulaire est ainsi une variable très dynamique grâce à la présence de pompes ioniques sélectives effectuant un transport actif de ces molécules en réponse à différents signaux. Dans les neurones la sensibilité au potentiel trans-membranaire de certains de ces canaux conduit à des flux calciques lors des potentiels d'action. La découverte de molécules devenant fluorescentes en présence de calcium a alors permis le développement de colorants indicateurs de calcium (calcium indicators) et de la technique dite d'« imagerie calcique » très utiles pour de nombreuses applications quand des activités cellulaires sont liées à des mouvements de calcium au travers la membrane. En neurosciences elle rend possible l'utilisation de tous les progrès de la microscopie par fluorescence pour étudier non seulement la structure de tissus cellulaire mais également une activité qui s'y déroule. Des microscopes avec des optiques à large champ et des caméras sensibles permettent alors de suivre dans le temps l'activité se déroulant dans un champ d'observation jusqu'à plusieurs centimètres carrés bien plus large que celui possible avec des matrices d'électrodes avec des résolutions spatiales élevées pouvant en outre être augmentée par l'utilisation de système co-focaux.

De très nombreuses approches sont possibles, avec différents types de colorants pour traiter les cultures à examiner, ou bien via des indicateurs encodés génétiquement. Il s'agit alors de protéines fluorescentes typiquement dérivées de la GFP et fusionnées à des domaines d'attachement de calcium. Avec ce système on peut alors créer des lignées d'animaux transgéniques ce qui permet des constructions dans lesquelles la protéine indicatrice est exprimée sélectivement uniquement dans certains types cellulaires.

Stimulation de l'activité neuronale en culture

Une variété de méthodes de stimulation électriques directes ont été développées pour les neurones et peuvent s'appliquer également sur des cultures. Elles sont notamment passées en revue dans le chapitre 8 du manuel de Carter et Sieh [CARTER et SHIEH, 2015a]. L'application de champs magnétiques peut également induire une stimulation de la culture [ECKMANN et collab., 2007].

Pour les cultures, tout comme pour de nombreuses applications *in vivo* et sur des tissus naturels, ce sont les méthodes optiques développées surtout depuis une quinzaine d'années qui suscitent le plus grand intérêt actuellement par la grande flexibilité qu'elles permettent.

L'excitation ou l'inhibition d'activité neurones est en effet possible par un biais optique, via l'éclairage avec une lumière appropriée, grâce à des modifications génétiques pour exprimer dans les cellules observées des gènes codant pour des canaux ioniques activés par un flux lumineux.

Cette approche dite d'« optogénétique » a considérablement renforcé l'outillage d'expérimentation en neurosciences sur le lien entre structures érébrales et fonction. Elle a été permise par la découverte et l'amélioration de différentes protéines de canaux ioniques dont l'ouverture ou la fermeture peut être induite par un éclairage à une bonne longueur d'onde. Ainsi une stimulation optique de neurones porteurs de ces canaux permet l'induction rapide d'une activité électrique.

Suivant le choix de canal optogénétique, le flux ionique induit dépolarisera ou « hyper » polarisera la membrane permettant d'induire ou inhiber un potentiel d'action. Les canaux optiques disponibles réagissent très rapidement à la lumière, de l'ordre de la milliseconde, et les cellules sensibles peuvent être sélectivement manipulées et contrôlant la focalisation du laser d'excitation dans des zones précises des tissus. L'activité induite peut ensuite être observée de différentes manières via des électrodes ou par fluorescence avec des indicateurs calciques.

L'optogénétique est ainsi une technique particulièrement flexible et adaptables pour des expériences diverses aussi bien *in vivo* qu'*in vitro* pour interagir en temps réel avec les réseaux neurones. La plus grande difficulté pour sa mise en œuvre consiste dans l'implantation dans les cellules ciblées des constructions génétiques permettant l'expression des canaux ioniques photosensibles. La création de lignées transgéniques généralement de rongeurs portant les gènes pour les protéines optogénétiques associés à des promoteurs spécifiques à des types de tissus est une procédure lourde, mais permet de disposer de modèles animaux prêts et exprimant les récepteurs seulement dans certains types cellulaires. De telles lignées sont cependant maintenant commercialement disponibles. Alternativement mais avec des résultats plus aléatoires, on peut introduire les codes des protéines optogénétiques dans des cellules cibles en transfectant les tissus avec des vecteurs viraux porteurs des constructions. En inoculant dans celles-ci des promoteurs spécifiques aux neurones (ou de tout autre type cellulaire ciblé), seules les cellules transfectées avec le régulateur correspondant actif vont exprimer les protéines des canaux ioniques photosensibles et devenir activables.

Comme nous l'avons montré dans le cadre de la thèse de Renaud Renault [RENAULT, 2015; RENAULT et collab., 2015], et sera discuté au chapitre suivant, section 7.2.7, il a été également possible de combiner les systèmes pour à la fois stimuler et enregistrer optiquement l'activité neuronale dans des cultures en sélectionnant des molécules pour l'excitation et pour le signal avec des spectres sans chevauchements.

Stimulation chimique

Les potentiels d'actions dépendent de l'équilibre des échanges membranaires d'ions au travers de canaux ioniques. Ces structures moléculaires sont sensibles à quantité de réactifs chimiques qui en alternent les propriétés et peuvent conduire à l'induction de dépolérisation du potentiel. L'activité neuronale peut-être également stimulée chimiquement par l'ajout de drogues offrant une approche complémentaire aux méthodes électriques ou optogénétiques.

De nombreux protocoles sont établis, notamment avec l'application transitoire d'une concentration de K^+ dans les culture. Cet ion intervenant dans le potentiel d'action, son changement de concentration extracellulaire induit facilement une dépolérisation de la membrane [FITZJOHN et collab., 2001]. Il est assez commun aussi d'agir par l'ajout de neuromodulateurs ou de bloqueurs synaptiques pour augmenter l'activité persistante de déclenchements des neurones en augmentant l'excitabilité des neurones (sur l'effet de ces drogues dans les cultures voir les références dans [PASQUALE et collab., 2008]) et leur sensibilité au bruit. Nous avons étudié de telles modulation dans le cadre des prédictions de notre étude sur les bouffées de potentiels d'actions, discuté au chapitre 7 à la section 7.3.

6.2.4 Les réseaux de neurones des cultures de neurones

Dans la section 6.2 plus haut j'ai discuté des dispositifs physiques utilisés pour réaliser les cultures ainsi que des questions liées aux contraintes géométriques. Je traiterai ici des propriétés des l'ensemble connexions qui s'établissent dans la culture et qui forment le réseau.

Croissance auto-organisée et fonction

Les caractéristiques de réseaux neuronaux en culture sont différentes de celles de neurones dans leurs environnements naturels. Les neurones se développent et les réseaux s'établissent dans un cadre biologiquement anormal. En premier lieu, les réseaux cultivés sont constitués par la croissance de cellules neuronales « dissociées » (disembodied). Tous les protocoles de préparation impliquent la suspension des cellules neuronales précurseurs dans des solutions. C'est le cas à partir de cellules souches ainsi que par les méthodes plus communes et classique de prélèvement de cellules cérébrales embryonnaires. Toutes les relations entre cellules existant dans cette phase critique du développement des organismes sont rompues dans le processus d'extraction. Les suspensions des neurones précurseurs primaires, issues directement du tissu, ensemble avec les cellules gliales prélevées simultanément détruisent toute l'organisation naturelle déterminante pour la croissance de l'embryon en mélangeant tous les différents types de neurones et glies des zones d'excision. Le milieu de cultureensemencé par ces cellules est désordonné, sans cohérence de position des différents types de neurones et de glies des tissus d'origine, sans aucune structure de l'environnement biochimique des cellules nécessaire au guidage normal du développement et sans l'influence d'échanges sensoriels et moteurs avec l'environnement du cerveau en développement.

La structure du réseau en culture, et donc également sa fonction, n'auront aucun rapport avec ceux du tissu original. Ces réseaux vont pousser à partir d'un milieu initial désordonné, les neurites vont croître, se différencier, faire des branchements et établir des connexions synaptiques à partir d'un mélange initial aléatoire de cellules cérébrales dont généralement seule la densité cellulaire est connue. En outre, à la différence des systèmes neuronaux naturels; la plupart des cultures neuronales sont sur des surfaces dans des couches cellulaires minces de topologies quasiment bidimensionnelles qui contraignent les possibilités de connexions. Pour les cultures de neurones primaires, directement issues d'un prélèvement sur des cerveaux de fœtus, la composition exacte en types de neurones et glies est mal définie. À partir du désordre initial de la culture, les neurones vont se développer et donner naissance à un réseau initialement non organisé et qui va croître selon une structure résultant des mécanismes de croissance et d'interaction des neurones entre eux et l'environnement géométrique et chimique qui leur est imposé. Les réseaux neuronaux vont se développer suivant des mécanismes et des conditions purement internes.

La structure des réseaux qui en résulte est encore de nos jours mal connue. Hors des réseaux jeunes (comme dans le cas de la figure 6.1) ou à faible densité, l'observation directe au microscope ne voit qu'un enchevêtrement dense de neurites et corps cellulaires sans possibilité de reconstituer la connectivité. Les réseaux ont des caractéristiques désordonnées, héritage de l'ensemencement au hasard. Mais ils ne sont pas complètement aléatoires. La biologie et l'environnement physique des chambres de cultures imposent des contraintes qui structurent le système qui s'y développe. L'observation d'activité dans des réseaux cultivés indique en effet des comportements différents en fonction de la composition cellulaire d'origine (zones cérébrales d'origine des cellules, espèce, cellules souches) et des conditions expérimentales [BETTENCOURT et collab., 2007; GROSS, 2011]. Ces relations s'interprètent en supposant l'action de mécanismes de croissance auto-organisés qui conduisent à l'émergence spontanée de structures de réseaux complexes déterminées par leur type de désordre.

Si les propriétés résultantes de l'évolution propre, auto-organisée du réseau neuronal en culture peuvent intéresser d'un point de vue théorique, il n'en demeure pas moins que du point de vue

des neurosciences elles restent problématiques. Dans ces systèmes le réseau s'est établi sans interactions fonctionnelles avec d'autres structures cérébrales et l'environnement. Comme je le discuterai dans la section suivante les cultures sont dominées par une activité collective sous forme de bouffées synchronisées de potentiels d'actions (network bursts) constituées de rafales de déclenchements successifs durant un court intervalle de temps de tous les neurones. Bien que ce type comportement montre une grande richesse de variantes [WAGENAAR et collab., 2006a], cette dynamique semblerait être liée au développement du réseau en absence de stimulations sensorielle [WAGENAAR, 2005]

En effet malgré cette diversité des bouffées de réseau, la constance de cette forme de dynamique dans les cultures semble contraster avec le répertoire varié d'oscillations spontanées à basse et à haute fréquence observés dans des études *in vivo* [TIMOFEEV, 2000]. Comme indiqué par Wagenaar [WAGENAAR, 2005], *in vivo* on observe également des bouffées de réseau, mais seulement durant une certaine durée limitée du développement en jouant un rôle dans l'établissement des connexions synaptiques [BANERJEE et ELLENDER, 2009] (et autres références dans [WAGENAAR, 2005]). La persistance de bouffées de réseaux sur de longues durées dans les cultures neuronales serait un signe du développement inachevé du réseau neuronal probablement en raison du manque de stimulation sensorielle. Ces idées correspondent aux observations de différentes expériences de privation sensorielle *in vivo* (occlusion de la paupière, inactivation médicamenteuse de la rétine) qui ont montré des conséquences majeures sur les propriétés de neurones individuels et par conséquent celles des réseaux et de l'activité de réseaux [BANERJEE et ELLENDER, 2009; LINDEN et collab., 2009]. Il serait alors intéressant d'étudier la formation de réseaux en culture avec une stimulation continue durant le développement et de rechercher si celle-ci influence la topologie du réseau et de la dynamique spontanée. Un sujet faisant justement l'objet du sujet de la thèse de Mallory Dazza en co-direction avec Catherine Villard de l'Institut Pierre Gille de Gènes sélectionné pour un contrat doctoral au concours de l'école doctorale 564 Physique en Île de France ce mois de juin.

Connectivité des réseaux en culture

L'unité de base d'une culture neuronale typique, y compris dans des structures complexes, est un compartiment de culture dans lequel le réseau va d'abord s'établir avec d'éventuelles ramifications projetant via des micro-canaux vers d'autres zones. La densité des cultures empêche en règle générale la caractérisation directe par microscopie des réseaux qui s'établissent dans un compartiment homogène. Des propriétés générales semblent se dégager, en tout cas pour les neurones de mammifères.

Au sein d'un compartiment les neurones primaires sontensemencés aléatoirement et les neurites croissent librement en interaction avec les cellules voisines et les bords du milieu. Les cultures typiques rassemblent entre 10^3 et 10^5 neurones avec des densités entre 500 et 5000 neurones par mm^2 . En dehors de cultures à très faible densité possible avec des neurones d'invertébrés [DE SANTOS-SIERRA et collab., 2014; SHEFI et collab., 2002a] il n'est pas possible de déchiffrer directement la connectivité structurale physique de réseaux neuronaux à deux dimensions à partir d'images de microscopie. Une culture de quelques 10^5 neurones possède plusieurs millions de connexions de neurites densément enchevêtrées. La morphologie des neurones en cultures globalement désorganisée est très mal connue en comparaison d'architectures cérébrales structurées. En raison du désordre, les approches électrophysiologiques combinées à la microscopie et aux reconstructions à trois dimensions [WEARNE et collab., 2005] qui ont permis de grandes avancées pour la cartographie des structures cérébrales, ne peuvent s'appliquer que dans quelques situations à basse densité [MARI et collab., 2015; DE SANTOS-SIERRA et collab., 2015]. Faisant la synthèse de plusieurs travaux expérimentaux de stimulation et d'enregistrement d'activité à l'aide de paires d'électrodes sur cultures matures dans des conditions usuelles, [MAROM et SHAHAF, 2002] concluent qu'il est raisonnable d'estimer qu'un neurone se connecte directement à 10-30%

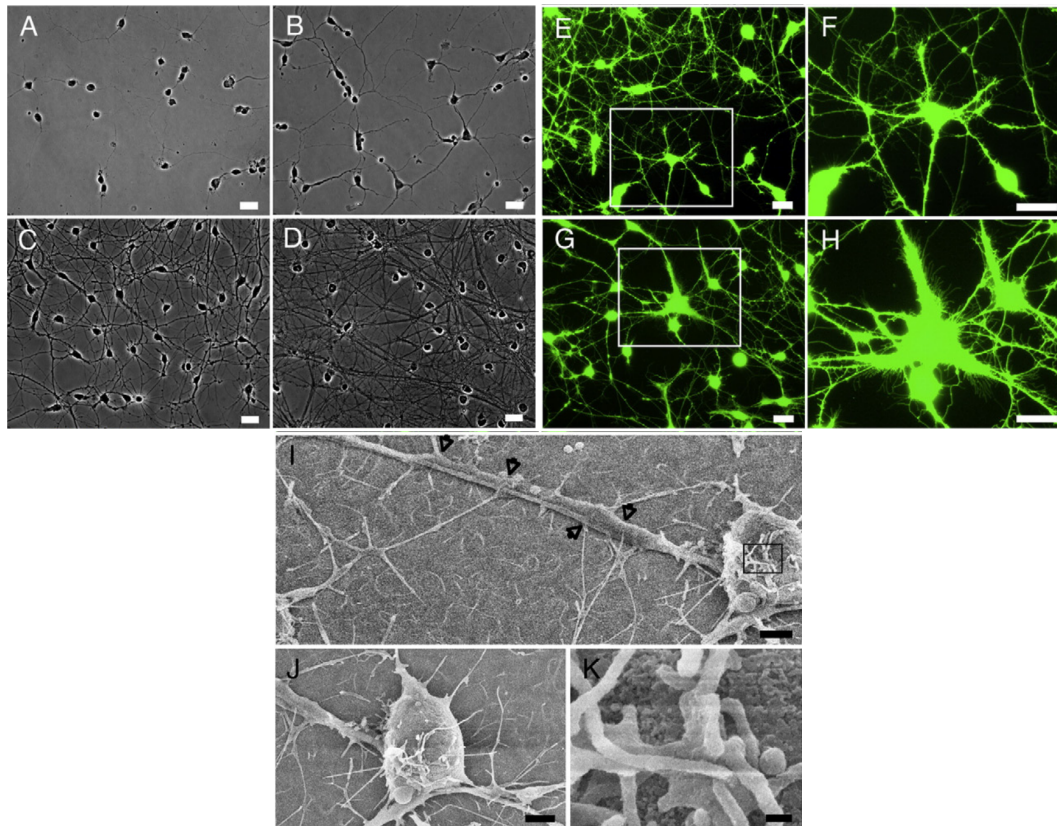


FIGURE 6.7 – Réseau de neurones en culture en développement et liaisons synaptiques potentielles Photos en microscopie par contraste de phase d'un réseau neuronal en culture en développement montrant la croissance des neurites et la formation du réseau à 2 DIV (« Days *in vitro* », jours après incubation), 4 DIV (B), 8 DIV (C), et 14 DIV (D) à une faible densité de 250–400 neurones/mm² (échelle 20µm). (E-H) : image en fluorescence de cultures de neurones à 8 DIV mettant en évidence, grâce à ces faibles densités d'incubation, la genèse des neurites et le développement du réseau (échelle 20µm). (I-J) : images en microscopie électronique de cultures neuronales à basse densité à 7 DIV, mettant en évidence les neurites et les points de contacts où des synapses peuvent potentiellement se former (flèches). (I) : Vue d'ensemble d'une grande projection neuronale (échelle 2µm). (J) : Vue du dessus montrant la croissance de neurites au-dessus d'un soma (échelle 2µm). (K) : Zoom sur un soma montrant la croissance de neurites sur la surface avec de nombreux points de contacts représentant des points de synapses potentielles. (I) : Zoom sur la région encadrée de (I), (échelle 200 nm). Source : figure 1 de [CULLEN et collab., 2010]

des autres neurones par des synapses fonctionnelles [AĆIMOVIĆ et collab., 2011; NAKANISHI et KUKITA, 1998]. Sans déchiffrer la structure du réseau neuronal CULLEN et collab. [2010] ont pu caractériser la densité de synapses en fonction de la densité cellulaire, un paramètre relié au réseau important. Pour des cultures matures (21 DIV) et des densités d'incubation de 10 to 5000 neurones/mm² ils ont mis en évidence que le développement des neurones se produisant dans une couche d'épaisseur de 18µm avec une distribution à trois dimensions complexe des neurites même dans ces cultures planes. Leurs résultats indiquent une relation inverse entre la densité des synapses et le nombre de neurones dans des cultures matures (le rapport synapse/neurone ~ 400 pour des densités d'incubation < 500 neurones/mm², contre 150 synapses/neurone à des densités 500–4500 neurones/mm²). La figure 6.7 montre des images de réseau neuronal en développement dans une culture à différentes dates après l'incubation et des points de contacts où pourront se former les synapses. Les concentrations relativement basses pour ces observations permettent une bonne visualisation de l'établissement progressif du réseau.

La caractérisation des réseaux en culture est encore largement indirecte combinant des données expérimentales d'activité avec des modèles théoriques et des simulations numériques pour construire des modèles de la structure *in vivo*.

Approches via la physique statistique et la théorie des graphes

Ceci a notamment été une des contributions importantes des modèles via le concept de percolation par Elisha Moses et Tsvy Tlusty qui ont permis de proposer un modèle de connectivité de cultures en réseau compatible avec un ensemble d'expériences d'activité. Leur théorie ajustée sur des observations expérimentales suggère que la connectivité dans des cultures de neurones d'hippocampe et de cortex de rat est dominée par le voisinage dans une géométrie plane avec une distribution statistique du nombre de liens entrants essentiellement Gaussienne. Cette conclusion se base sur des mesures du nombre de neurones qui se déclenchent suite à une stimulation électrique d'ensemble d'intensité croissante du réseau. L'expérience montre que lorsque la fraction des neurones initialement excités par le stimulus extérieur dépasse une certaine valeur critique, alors l'activité du réseau connaît un saut soudain avec une propagation très rapide recrutant l'intégralité de la population. En revanche, en dessous de ce seuil de stimulation l'activation ne se propage essentiellement pas [COHEN et collab., 2010]. Ce phénomène peut être capturé par un modèle théorique simple d'activité neuronale qui permet de décrire le comportement en termes d'un processus de percolation particulier dit de « bootstrap » (amorçage) [CHALUPA et collab., 1979] sur un graphe aléatoire dirigé.

Ce modèle statistique présente des transitions de phase qui dépendent du seuil d'excitabilité des neurones et des caractéristiques topologiques du réseau des connexions. Il permet donc de relier des observables macroscopiquement, les transitions de phase, à des paramètres topologiques et neuronaux. Connue aussi sous le nom de « Quorum Percolation » ce modèle pour l'activation collective d'une culture neuronale permet ainsi d'interpréter des observations expérimentales de réponses macroscopiques d'une culture avec un modèle théorique dépendant de paramètres topologiques [SORIANO et collab., 2008]. L'ajustement de ceux-ci pour reproduire les courbes expérimentales a permis plusieurs conclusions sur la connectivité subjacente inaccessible par toute autre technique. En premier lieu il apparaît que les cultures utilisées, une chambre d'une surface de l'ordre du cm^2 avec une distribution homogène de quelques centaines de neurones par mm^2 , peuvent être représentées par des graphes aléatoires avec une distribution Gaussienne du nombre de connexions entrantes sur chaque neurone [SORIANO et collab., 2007]. Cette représentation par un graphe aléatoire sans besoin de notion d'espace et de voisinage suffit pour rendre compte de la réponse collective de la population et a permis de proposer des évaluations des nombres moyens de connexions entrantes de l'ordre de 60 à 120 pour de neurones d'hippocampe et 75 à 150 pour de neurones corticaux, augmentant linéairement avec la densité et des variances autour de la connectivité moyenne entre 25% et 50% [COHEN et collab., 2010]. Les composantes inhibitrices ont également pu être déterminées avec 30% de liens inhibiteurs pour l'hippocampe et 20% pour le cortex [SORIANO et collab., 2008]. Bien que la connectivité moyenne est probablement sous-estimée (voir la discussion sur le modèle de percolation avec une décroissance des signaux du chapitre suivant, section 7.2.3) les valeurs trouvées sur les réseaux sont inférieures au moins d'un ordre de grandeur avec la connectivité moyenne dans le cerveau de plusieurs milliers de connexions.

Des raffinements de ces expériences et analyses ont par la suite montré quelques divergences à la distribution Gaussienne avec une coupure sur les nombres maximal et minimal des liens entrants [COHEN et collab., 2010] mais aussi une sous population de 10% avec une queue de distribution de connectivité diminuant en loi de puissance [ECKMANN et collab., 2010]. Cette sous population fortement connectée regrouperait des neurones *guides*, « leaders » qui se retrouveraient les premiers à déclencher dans l'initiation des bouffées d'activité globale spontanées observées dans les cultures.

Cependant ces descriptions du réseau en culture sont en termes de graphes aléatoires sans aucune notion d'espace. L'organisation spatiale des neurones et des connexions dans la culture doit certainement jouer un rôle dans la dynamique du réseau. Les expériences discutées-ci dessus sont clairement avec des cultures organisées sur un plan. Le succès d'une représentation dénouée de

métrique pour décrire le comportement statistique d'ensemble d'un réseau spatial à deux dimensions est surprenant. Elle s'apparente à celle d'un graphe de connexions plus proches voisins dans un espace à très grand nombre de dimensions. Tlusty et Eckman [TLUSTY et ECKMANN, 2009] ont cependant résolu cette apparente contradiction en montrant qu'effectivement un graphe aléatoire est une approximation acceptable pour des réseaux denses de faible surface par rapport à l'extension des axones et une connectivité moyenne suffisamment élevée (voir aussi [ECKMANN et collab., 2010]).

Si certaines caractéristiques d'un réseau fortement connecté d'extension limitée peuvent être similaires à un graphe aléatoire, toutes les propriétés ne font pas abstraction de l'organisation spatiale. En particulier dans un réseau métrique où la notion de voisinage a un sens, il peut exister un noyau localisé d'initiation d'un front de propagation d'activité. Dans un graphe aléatoire cette notion n'existe pas. Si l'on s'intéresse à la dynamique de propagation de l'activité dans la culture il est donc nécessaire de compléter la représentation du réseau par un modèle métrique.

En absence d'observations directes plusieurs groupes ont proposé des modélisations de l'organisation spatiale de réseaux en culture. Il s'agit de procédures de simulation de la croissance d'un réseau neuronal dans l'espace plus ou moins inspirés par la biophysique qui permettent de générer des réseaux qu'on espère représentatifs des cultures réelles. Ces modèles doivent reproduire les quelques propriétés topologiques déjà caractérisées, mais n'amènent pas directement des informations nouvelles confirmées. ECKMANN et collab. [2010] et ZBINDEN [2011] ont ainsi proposé un modèle spatial compatible avec leurs résultats de distribution statistiques de connectivité qui les conduits à plusieurs conjectures sur les distributions de tailles des dendrites (Gaussienne avec une fraction de neurones avec une extension lointaine). Ce sujet encore très ouvert fait l'objet d'un de mes projets actuels de recherche.

Comme l'organisation spatiale du réseau influe en premier lieu sur la propagation de l'activité, ses propriétés doivent se refléter dans la dynamique spatio-temporelle relevées dans des études dites fonctionnelles qui observent l'activité dans le temps et l'espace.

Approches via les réseaux fonctionnels

Alternativement à l'identification de structures de connectivité physique trop denses pour être directement caractérisées, plusieurs auteurs ont abordé l'organisation des neurones en culture sous l'angle de réseaux fonctionnels reliant les activités enregistrées à un ensemble de positions fixes, typiquement les électrodes d'une matrice MEA, pour obtenir une estimation du réseau neuronal subjacent [CHIAPPALONE et collab., 2006; DOWNES et collab., 2012; NAPOLI et collab., 2014; ORENSTEIN et KEREN, 2017; SCHROETER et collab., 2015]. Cette approche reprend pour les cultures les méthodes d'analyse statistique de dépendance entre nœuds d'activité développés pour l'étude encéphalographiques cérébrales pour interpréter les enregistrements spatio-temporels d'activité électrique de type électro-encéphalographiques à de nombreux points du cerveau [FELDT et collab., 2011; SPORNS et collab., 2004].

Le lien entre réseau fonctionnel relevé par l'activité de quelques points de mesure et morphologie réelle du réseau physique des connexions entre les neurones n'est cependant pas direct et ne me semble avoir été caractérisé que partiellement en culture [MAHESWARANATHAN et collab., 2012]. Dans le cerveau, des études ont montré l'absence de relation simple entre connectivité fonctionnelle et structurelle (anatomique). Des corrélations fortes d'activité entre zones distinctes sont en effet possibles sans connexion physique directe [HONEY et collab., 2009]

L'interprétation des résultats fonctionnels ne doit s'interpréter qu'avec prudence sur le réseau de connectivité physique, mais les conclusions renseignent déjà sur des tendances probables [JOVANOVIĆ et ROTTER, 2016; OCKER et collab., 2017a,b; POLI et collab., 2016]. En outre la possibilité de suivre facilement l'activité des cultures sur plusieurs semaines, en répétant par exemples des mesures MEA à des dates successives, permet d'acquérir des représentations du réseau fonction-

nel à des âges différents et de donner des informations précieuses sur la maturation du réseau réel, même par ce biais indirect.

Comme le rapportent notamment les résultats de [DOWNES et collab. \[2012\]](#) le réseau fonctionnel d'une culture corticale se structure selon un modèle de réseau aléatoire avec une propriété petit monde, un faible parcours moyen et un haut clustering. Des propriétés qui réduiraient les durées de propagation de signaux dans la population favorisant la synchronisation, ainsi que l'apparition de nœuds fortement connectés se comportant comme des centres d'influence.

Cependant je resterai prudent sur les conclusions des analyses par réseau fonctionnel sur les cultures. En premier lieu pour interpréter les résultats il serait intéressant d'explorer le lien entre les propriétés d'un réseau fonctionnel relevé par l'activité de quelques points de mesure et la morphologie réelle du réseau physique des connexions entre les neurones. D'autres part les expériences d'activité en culture sont très sensibles aux conditions et protocoles expérimentaux exacts (par exemple la sensibilité au milieu de culture illustré par les expériences de [PENN et collab. \[2016\]](#)). Il est probable que cette variabilité conduise à des observables statistiques de réseau très différents dans des analyses de graphes fonctionnels sur un nombre de points de mesure très limités.

Résultats dans des cultures à faible densité

Les neurones d'invertébrés comme la sauterelle sont de grande taille et peuvent être prélevés adultes et dissociés de leur tissu original (glanglions). Ce faisant ils perdent leurs neurites, mais celles-ci, dendrites et axone se régénèrent dans le milieu de culture. En raison probablement de leur taille, ces neurones survivent aussi avec un ensemencement à faible densité. La croissance des réseaux peut s'observer distinctement au microscope optique et des méthodes d'analyse d'image et de segmentation peuvent aider dans ces conditions à reconstruire les réseaux [[DE SANTOS-SIERRA et collab., 2015](#)]. Ainsi [SHEFI et collab. \[2002a,c\]](#) puis [DE SANTOS-SIERRA et collab. \[2014\]](#) ont pu mettre en évidence que le réseau en croissance n'est pas purement aléatoire et fait apparaître une structure de petit monde avec un chemin moyen court et un clustering élevé. Une propriété caractéristique de nombreux réseaux biologiques et supposée favoriser le transfert efficace d'information. Dans le contexte de la culture cette propriété se relie à l'apparition d'une activité spontanée synchronisée dans les cultures dont je parlerais dans la section suivante. La connectivité observée par [DE SANTOS-SIERRA et collab. \[2014\]](#) correspond à un réseau avec un nombre régulier homogène de liens augmentant avec l'âge de la culture, sans que cette publication distingue le sens des connexions. Un degré homogène de connexions par neurone qui peut signifier une distribution Gaussienne, comme celles déduite par le groupe de E. Moses, avec une variance faible.

Dans les conditions utilisées, [SHEFI et collab. \[2002a,c\]](#) ont observé un processus de croissance dans lequel les cônes de croissance établissaient des connexions non seulement aux neurones voisins mais également à des neurites de son propre corps cellulaire sans évitement apparent de soi-même. Cette situation conduit à des réseaux présentant des boucles sur soi-même.

Modèles de réseaux en culture

En absence de caractérisations expérimentales complète des réseaux dans les cultures de nombreux modèles théoriques ont été introduits et utilisés afin d'explorer théoriquement des comportements collectifs qui pourraient être attendus. Ces travaux en grande partie numériques cherchent ainsi à caractériser les liens entre comportements possibles de réseaux, fonctions biologiques qui pourraient être associées, et structure sous-jacente du réseau.

Faute d'éléments suffisants ces études se sont assez généralement appuyées sur des réseaux purement aléatoires de théorie des graphes. De tels réseaux sont décrits statistiquement par un

nombre limité de paramètres comme, par exemple, la probabilité de connexion entre deux nœuds quelconques dans le modèle classique de Erdős-Rényi. Les modèles de connexions aléatoires avec une distribution du nombre de connexions sur chaque neurone sont construits autour d'une connectivité moyenne et d'une variance et permettent d'approcher une connectivité ordonnée. Alternativement le modèle de petit-monde inclut à ces modèles la possibilité de varier la proportion de connexions à longue portée. Enfin les réseaux invariants d'échelle possèdent des structures à toutes les échelles et sont caractérisés par une distribution en loi de puissance du degré de connexion des nœuds. Dans ces cas les réseaux font abstraction de toute localisation spatiale. Ces différents types de réseaux abstraits ont été utilisés dans de nombreux travaux numériques appliquant des modèles dynamiques d'activité neuronale et de communication synaptique recherchant de manière plutôt exploratoires les comportements attendus avec les différents types d'organisation. On peut lire une synthèse de tels travaux sur des modèles de réseaux issus de théorie des graphes dans [MÄKI-MARTTUNEN et collab., 2013]. En particulier on peut relever la généralité de comportements synchronisés à large échelle qui sont retrouvés dans des réseaux divers [BRUNEL et BRUNEL, 2000; BRUNEL et HAKIM, 1999; MASQUELIER et collab., 2013; VAN VREESWIJK et HANSEL, 2001].

Malgré leur utilité les réseaux aléatoires ne sont qu'une première approche qui s'avère utile et pertinente dans des configurations particulières [TLUSTY et ECKMANN, 2009]. Plusieurs expériences sur des cultures montrent en effet la complexité de la dynamique spatiale dans les cultures [LONARDONI et collab., 2017; ORLANDI et collab., 2013] avec la présence de zones d'initiation d'événements et la propagation de vagues d'activité. De telles observations ont naturellement conduit à l'introduction de modèles de réseaux aléatoires dans un espace métrique. Dans la forme la plus simple il s'agit d'une distribution aléatoire des corps cellulaires formant les nœuds du réseau sur une surface et de l'établissement de connexions entre les nœuds suivant une loi probabiliste dépendant de la distance. Cette règle implémente l'idée des connexions à distance via des axones interagissant avec un arbre dendritique plus centré sur le soma d'un neurone receveur [ACIMOVIC et collab., 2015; MAHESWARANATHAN et collab., 2012]. Alternativement d'autres modèles incorporent des représentations qui se veulent plus explicites des neurites. C'est le cas d'axones représentés par une marche aléatoire et se connectant à des neurones cibles en fonction de l'intersections de l'axone avec un disque représentatif de l'arbre dendritique de la cible [ORLANDI et collab., 2013], ou plus finement d'axones implémentant un modèle morphologiquement réaliste du cheminement d'un axone [GRITSUN et collab., 2012]. C'est dans l'évolution de ces approches vers une représentation plus réaliste du développement de neurones que se sont orientées une partie des activités de notre groupe, ainsi que sera développé dans la section 7.4 du prochain chapitre 7. En effet, comme l'ont aussi démontré théoriquement [HERNÁNDEZ-NAVARRO et collab., 2017] du groupe de Jordi Soriano à Barcelone, dans des cultures les longueurs finies des axones et l'hétérogénéité effective dans une surface de taille finie de la distributions des neurones dans un plan génèrent de fortes corrélations métriques qui déterminent la dynamique davantage que les distributions de connectivités seules considérées dans les approches de champ moyen. Ces résultats très convaincants motivent une approche spatiale plus réaliste des réseaux.

6.3 Applications des cultures de neurones

Les cultures neuronales dans des environnements artificiels structurés offrent par leurs possibilités d'observation et de contrôle un potentiel de recherche aussi bien pour l'étude de fonctions cognitives que pour la conception de modèles *in vitro* de pathologies neuronales difficiles d'accès *in vivo* ou sur des modèles animaux. Les cultures sont ainsi devenues un outil précieux pour comprendre le système nerveux, depuis les constituants moléculaires des fonctions neuronales jusqu'à au comportement de populations.

En premier lieu les cultures permettent la conception de plateformes d'expérimentation fondamentales en neurosciences et des modèles d'études de troubles neurologiques. Construites à

partir des cellules élémentaires du système nerveux central, ces cultures sont naturellement un outil expérimental unique sur les principes d'organisation et de fonctionnement soutenant le traitement de l'information cérébrale. En raison des propriétés particulières d'excitabilité les cultures neuronales suscitent également de l'intérêt hors d'un cadre proprement biologique pour des applications dans des dispositifs technologiques tels des bio-senseurs.

Les usages de cultures neuronales connaissent depuis une dizaine d'années un renouveau important par l'apport des techniques de microfabrication issues de la microfluidique [NETO et collab., 2016]. S'y ajoutent également des avancées dans les différentes des techniques de stimulation et de mesure des activités, par des matrices d'électrodes ou par l'optogénétique. Ces raisons conduisent le champ des cultures neuronales à être bien actif et touchant des domaines d'applications très différents. Un champ d'exploration interdisciplinaires aux croisements des neurosciences, des sciences de l'information, des mathématiques et de la physique.

6.3.1 Applications biomédicales

Les cultures de neurones dissociés sont utilisées dans de nombreuses études physiologiques, pathologiques ou pharmacologiques en raison de la plus grande accessibilité des cellules à des traitements chimiques, à l'imagerie et des stimulations et enregistrements électriques que des modèles *in vivo*. Elles sont ainsi utilisées dans études traitant de propriétés d'activité des réseaux [BEGGS et PLENZ, 2003; MAEDA et collab., 1995; SEGEV et BEN-JACOB, 2001; VAN PELT et collab., 2004b; WAGENAAR et collab., 2006a], du fonctionnement des synapses et la plasticité [GAL et MAROM, 2013] [SHAHAF et MAROM, 2001a] [MAROM et EYTAN, 2004], de criblage de molécules ciblant le système nerveux, de tests pharmacologiques [MATSUDA et collab., 2018; MOREFIELD et collab., 2000] ou encore de modèles *in vitro* de pathologie comme l'épilepsie [DELORENZO et collab., 2007].

Grâce à la possibilité de réaliser par la microfluidique des cultures dans des géométries sophistiquées combinant différentes zones de cultures reliées entre elles, d'autres questions fondamentales concernant les propriétés spatiales des neurones deviennent plus accessibles. Il en est ainsi de l'étude de la croissance [FRIMAT et collab., 2010; PEYRIN et collab., 2011; YAMADA et collab., 2016], de la formation des synapses [DELEGLISE et collab., 2013] ou encore du transport intracellulaire de molécules dans les neurites [CRONIER et collab., 2007].

Les co-cultures de différents types cellulaires dans des dispositifs microfluidiques permettent de tester les interactions entre cellules tout en ayant accès à des contrôles distincts de leurs conditions environnementales (différents fluides, traitements). Ces co-cultures peuvent être de neurones de même nature ou de types différents [DINH et collab., 2013, 2014], ou bien de neurones et de cellules gliales [CERQUEIRA et collab., 2018; DAMBACH et collab., 2014]. Des cellules de propriétés différentes de zones cérébrales distinctes peuvent être maintenues dans des zones de culture adjacentes tout en permettant des connections. Ceci permet la création de modèles *ex vivo* de communication entre zones cérébrales brain-on-a-chip [DAUTH et collab., 2017]. Ces études trouvent un contexte d'application pour une variété de maladies cérébrales.

6.3.2 Applications aux troubles neurologiques

Plusieurs troubles neurologiques sont liés à des désordres du réseau neuronal. Cette connectivité anormale est causée par le dysfonctionnement des mécanismes de croissance neuronales durant le développement ou dans le cas de maladies neurodégénératives, dont Alzheimer et Parkinson, par des altérations du cytosquelette ou en conséquence d'agrégation de protéines, on trouvera des nombreuses références dans [HOCQUEMILLER et collab., 2010].

Les cultures neuronales sont un outil précieux pour ces études car elles permettent l'observation de la morphologie de neurones et réseaux durant la croissance, ou de leur altération dans le cadre de modèles de pathologies. En particulier il est intéressant d'étudier ces observations avec

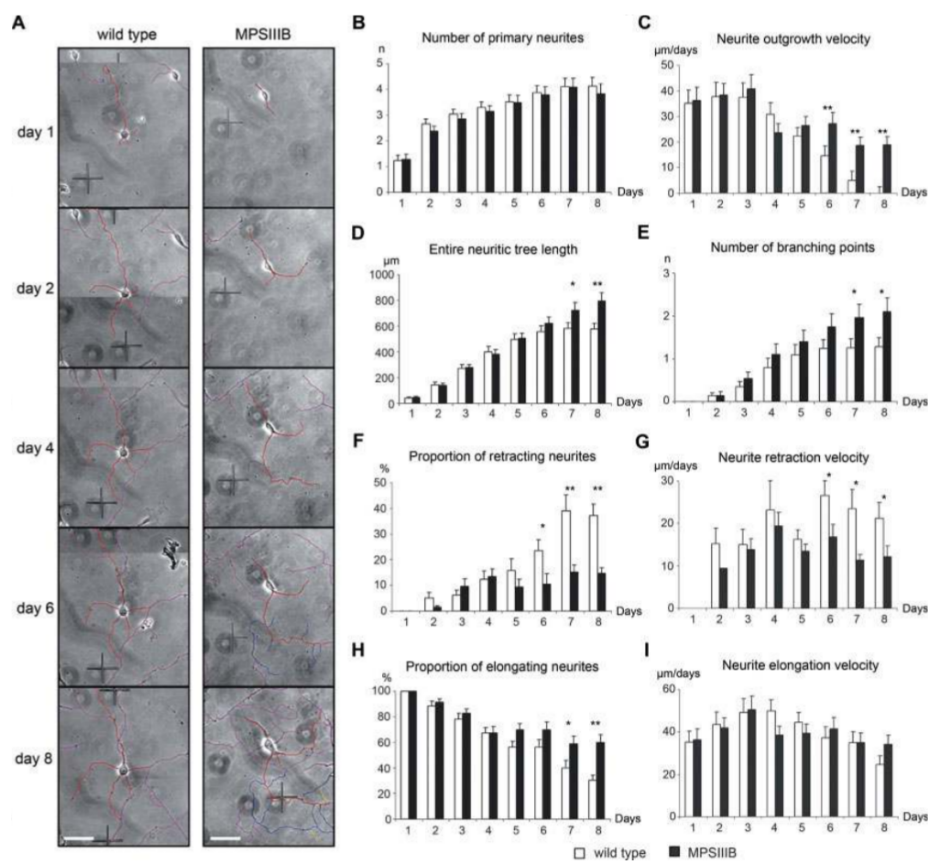


FIGURE 6.8 – Suivi au jour le jour de la dynamique de croissance de neurites. Suivi sur des cultures corticales de souris en microscopie par contraste de phase entre une culture normale (wild type) et une culture de cellules issues d'un modèle murin du syndrome San Filippo (MPSIII). (A) : exemple de développement de neurites saines (gauche) et pathologiques (droite). (B – I) : dynamique de plusieurs caractéristiques morphologiques des neurites. Source : [HOCQUEMILLER et collab., 2010]

d'autres sur des structures saines pour tester des hypothèses sur les mécanismes de la maladie.

Ceci est par exemple le cas pour des troubles du spectre de l'autisme (Autism spectrum disorder – ASD) qui impliquent les protéines dites Contactines. Des altérations de leurs gènes codant semblent en effet modifier la croissance le branchement et le guidage des neurites dans les cerveaux des patients ASD [STOECKLI, 2004]. L'étude de la morphologie des neurones en cultures en présence de différents facteurs chimiques permet de rechercher les mécanismes de développement neuronal, de maturation des synapses et de plasticités qui interviennent dans la maladie [MERCATI et collab., 2013].

Le syndrome Sanfilippo (mucopolysaccharidose de type III, MPSIII) est une autre pathologie avec des caractéristiques morphologiques importantes. L'observation en culture du développement de neurones modèle de la pathologie a montré des paramètres de croissance anormaux avec des élongations et branchements de neurites plus fréquents et moins de rétraction des neurites qui aboutissent à des arborescences dendritiques et axonales surdimensionnées par rapport à des contrôles normaux. Les tests sur les cultures permettent de rechercher les facteurs moléculaires et génétiques associés à ces altérations [HOCQUEMILLER et collab., 2010].

La maladie d'Alzheimer est un processus dégénératif progressif la maladie. La dégénérescence se propage dans le temps et dans le cerveau suivant un chemin régulier par le cortex et l'hippocampe. Les cultures dans des dispositifs microfluidiques sont prometteuses pour construire un modèle *ex vivo* les chemins cérébraux impliqués. Ainsi la collaboration des groupes de JM Peyrin et de JL Viovy a visé à construire un dispositif de culture de trois types neuronaux différents interconnectés du cerveau (cortex, gyrus denté et corne d'Ammon) dans le but d'étudier la progression

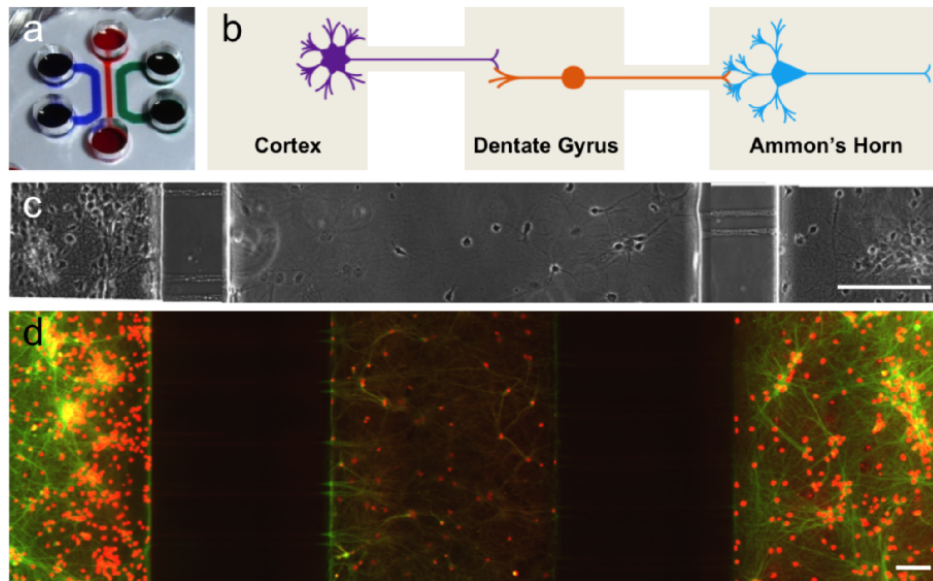


FIGURE 6.9 – Schéma de conception pour la reproduction dans un dispositif de culture microfluidique du chemin initial de propagation la maladie d'Alzheimer. A) Image du circuit intégral avec les trois chambres pour le cortex, le gyrus denté et la corne d'Ammon (de gauche à droite). B) Illustration du réseau. C) Superposition des images montrant les trois chambres ensemencées avec des neurones issus des 3 zones cérébrales respectives. D) immunomarquage de l'α-tubuline (en vert) et des noyaux cellulaires (relevés en fausse couleur en rouge). Échelle 100 μm. Source : M. Vignes thèse 2013.

temporelle et spatiale des lésions associées à la maladie d'Alzheimer.

6.3.3 Criblage moléculaire et biosenseurs

Outre les observations morphologiques les cultures sont utilisées dans un contexte pharmacologique plus large car elles se sont révélées avoir des propriétés similaires à celles des tissus cérébraux d'origine, tout en amenant des facilités de manipulation. Ceci conduit à l'idée d'utiliser ces cultures comme des plateformes rapides de criblage pharmacologique pour des tests de toxicité ou de la recherche de molécules actives [STENGER et collab., 2001] Il semble en effet assez facile de soumettre en quantité les cultures à des produits divers et de mesurer simplement électroniquement les changements d'activité et de détecter ainsi des effets possibles.

Ce principe peut être repris dans une perspective différentes pour utiliser les cultures comme senseurs chimiques [SELINGER et collab., 2004]. On peut en effet envisager de détecter des composés chimiques particuliers neuro-actifs en détectant les modifications d'activité de cultures auxquelles elles ont été exposées. Cette idée pourrait être affinée et étendue en insérant dans les neurones des récepteurs membranaires spécifiques parmi les milliers de récepteurs connus.

Un dernier champ d'applications de cultures que je traiterai à part en raison de mon intérêt particulier à leur sujet concerne l'exploration à l'aide de cultures de propriétés cognitives.

6.4 Vers des processeurs neuronaux

« La réflexion engendre la faculté de raisonner. »

Gustave Le Bon, Les opinions et les croyances (1911)

Dans les organismes vivants les circuits neuronaux ont la capacité d'effectuer des opérations de calcul complexes, correspondant au traitement d'information. Dans cet axe d'investigation on

veut rechercher les conditions nécessaires pour que des ensembles de neurones démontrent des capacités calculatoires. Par ces dernières on désigne les propriétés d'un système à produire des résultats différenciés en fonction des conditions particulières auxquelles il est soumis. L'entrée correspond à une information sur l'état extérieur, elle peut être changeante et correspondre à des situations environnementales différentes. La sortie est une transformation de cette entrée en une autre information, sous une autre forme, qui est l'interprétation par le système des conditions externes élaborée pour la suite des processus dans lesquels il est impliqué.

La capacité de calcul est ici synonyme à celle de traitement d'information, en d'autres termes, d'activité cognitive élémentaire : à des conditions données le système sait produire en réponse un comportement particulier utile pour la suite. En définitive la nature calculatoire d'un comportement sera déterminé par son usage dans un environnement plus large : une bille en chute libre après être lâchée n'a pas de capacité calculatoire (même si elle obéit aux équations de Newton), si la bille tombe sur un bouton poussoir, elle actionne un interrupteur et donc une fonction qui fait passer d'un état à un autre.

6.4.1 Codage neuronal et réseaux neuronaux

La capacité calculatoire d'un cerveau implique que celui-ci élabore une représentation des informations qu'il perçoit et traite. Une des questions majeures des neurosciences est bien d'établir la manière, ou les manières, avec laquelle le cerveau stocke et traite les informations en utilisant certaines propriétés des potentiels d'action [BIALEK et ZEE, 1990]. En effet une séquence de déclenchements de neurones (un « train » de potentiels d'actions) peut coder de l'information de différentes manières. Il est ainsi bien établi que la force de contraction musculaire est contrôlée par les neurones périphériques moteurs via la fréquence moyenne sur un intervalle temporel.

Dans d'autres cas comme les signaux des systèmes visuels et auditifs, ce sont les écarts temporels ou les corrélations entre déclenchements, jusqu'à des échelles aussi fines que la milliseconde, entre déclenchements individuels qui représentent l'information. Outre ces deux cas classiques d'autres mécanismes de codage ont encore été proposés [GERSTNER et collab., 1997; VOGELS et ABBOTT, 2005]. Il est clair maintenant qu'un code unique universel n'existe pas, l'évolution a fait émerger différentes stratégies pour des types d'information différentes. Le sujet est encore compliqué par la découverte de capacités calculatoires au niveau des arbres dendritiques eux-mêmes.

Le schéma de représentation des informations n'est qu'un élément de l'algorithme général complet du cerveau pour ses capacités cognitives. La réflexion théorique sur les caractéristiques du système de codage neuronal se poursuit de longue date et fait apparaître différentes composantes. De nombreuses références récentes [LÁBOS, 2000; ORZÓ et LÁBOS, 2000; VOGELS et ABBOTT, 2005] citent ainsi BULLOCK et PERKEL [1968] pour avoir identifié au moins 4 composantes fondamentales du système de codage cérébral : un schéma de représentation de l'information perçue basé sur les potentiels d'action, une capacité d'interprétation de l'information par rapport à des critères nécessaires en vue de guider une action (le cœur du traitement de l'information, la prise de décision), la transformation de cette interprétation dans un résultat exprimé aussi selon un certain code et enfin la capacité de le transmettre de manière fiable vers d'autres structures cérébrales (le cerveau est modulaire) ou bien vers le système nerveux périphérique.

6.4.2 Robustesse du calcul neuronal

Les mécanismes de codage neuronal doivent mettre en œuvre ces différentes fonctions de manière fiable et efficace. En effet le fonctionnement cérébral est soumis à la fois à des contraintes énergétiques et à une exigence de fiabilité. Plusieurs études expérimentales et techniques ont montré que l'activité électrique des neurones, la génération de potentiels d'action est un processus énergétiquement coûteux. On peut naturellement supposer que l'évolution du cerveau a du tendre à optimiser la capacité de traitement fiable de l'information par rapport au nombre de

neurones et la complexité du réseau [BARLOW, 1961]. Étant donné la variabilité des structures biologiques et les fluctuations environnementales la fidélité du codage et la tolérance à des erreurs est un facteur essentiel dans chaque fonction du codage neuronal.

Une question ancienne toujours pertinente est en effet la fiabilité des calculs complexes réalisés par les circuits neuronaux. L'organisation des neurones dans le cerveau permet en effet la réalisation d'opérations sophistiquées que l'on peut assimiler à des calculs complexes avec une grande efficacité en terme de vitesse et de faible erreur. On peut songer par exemple à toutes les opérations nécessaires au système nerveux visuel pour interpréter les données en provenance des yeux pour reconstituer une représentation à trois dimensions de l'espace et guider le mouvement des membres. Des opérations demandant des algorithmes sophistiqués et des moyens de calculs puissants pour des robots électroniques. Or le fonctionnement du cerveau est soumis à des fluctuations continues, notamment au niveau des synapses. Comment le cerveau compense ce fonctionnement thermodynamique? Il est connu depuis longtemps qu'un système redondant permet de compenser les fluctuations et erreurs de chaque composant individuel. La répétition des mêmes opérations de nombreuses fois en parallèles permet de pallier pour la défaillance d'une unité de calcul élémentaire. Cette redondance est-elle également le principe par lequel le cerveau assure la robustesse de ses opérations? On récupère dans ce cas un résultat fiable simplement par la moyenne d'une sortie opérée sur une population ou une durée d'intégration. Le cerveau obéit-il à un tel principe de redondance? C'est ce que suggèrent des résultats indiquant que l'effet moyennant du codage de l'information par le taux moyen de déclenchements des neurones est la manière la plus efficace de transmettre de manière fiable de l'information le long de chaînes de neurones [VAN ROSSUM et collab., 2002; SHADLEN et NEWSOME, 1994, 1998; VOGELS et ABBOTT, 2005].

Cependant cette vue à été contestée surtout à partir des années 1990 par des séries d'études [AERTSEN et collab., 1996; DLESMANN et collab., 1999] montrant que des populations de neurones peuvent parfaitement transmettre avec fidélité l'information codée dans les instants temporels des potentiels d'actions. Au contraire de la vue classique, les neurones de couches successives d'une ligne de transmission voient leur taux moyen d'activité tendre vers une limite indépendante de celle en entrée [LITVAK et collab., 2003]. L'information peut se propager dans un réseau cortical dans des successions de volées de potentiels d'action précisément synchronisés nommées « synfire chain » [DLESMANN et collab., 1999; GEWALTIG et collab., 2001]. En outre l'observation, surtout dans le contexte de stimulations sensorielles, de séries de potentiels d'action se répétant avec des précisions jusqu'à la milliseconde doit bien signifier une importance fonctionnelle de la temporalité précise [GEWALTIG et collab., 2001].

Ces différents points de vue sur l'information cérébrale sont au cœur de la question générale de la représentation de l'information par le cerveau et sont toujours d'actualité. Les énormes progrès d'imagerie fonctionnelle et d'enregistrement de l'activité cérébrale offrent des perspectives de progrès significatives, mais il faut faire face à la complexité des données, encore partielles, et des méthodes d'analyse. La modélisation et la simulation jouera certainement un rôle central en support des études expérimentales afin de tester des hypothèses. Étant donné les dimensions des systèmes étudiées les relations entre structure et activité ne pourront être réellement abordées qu'avec un support théorique et numérique. Cependant un autre défi pour comprendre la nature de l'information dans le cerveau et le lien avec l'organisation de celui-ci est la difficulté de manipuler ces relations dans des expérimentations. Les cultures neuronales, avec les contrôle possible aujourd'hui de la connectivité des réseaux ont ouvert des perspectives pour construire des modèles d'architectures de populations neuronales permettant d'étudier expérimentalement les possibilités biologiques de traitement d'information.

6.4.3 Approches de recherche

Ces contraintes et les manières d'y répondre font l'objet d'un très grand nombre de publications expérimentales et théoriques [GERSTNER et collab., 1997; GILSON et collab., 2009; KUMAR et collab., 2010; LÁBOS, 2000; ORZÓ et LÁBOS, 2000; ?]. Une manière d'aborder ce questionnement est de relier la dynamique d'un réseau neuronal ses propriétés de stockage, de traitement et de propagation d'information à la structure du réseau. Avec nos connaissances du fonctionnement synaptique et du déclenchement des potentiels d'action, en fonction de la connectivité de neurones, quelles sont les performances de différents schémas de codage neuronal? Réciproquement un autre regard consisterait à s'interroger sur l'émergence de structures de réseau tenant compte de règles de plasticité permettant l'évolution du réseau. Quelles sont alors les réseaux capables de soutenir des propriétés cognitives élémentaires de traitement d'information qui émergent [LÁBOS, 2000]?

Les cultures de neurones sont un moyen pour explorer expérimentalement les relations entre structure des réseaux et propriétés informationnelles. Plutôt que de déduire les relations *in vivo* entre codage neuronal et structure sur des réseaux sophistiqués déjà évolués par la nature, on peut chercher par la construction de réseaux et l'essai de schémas de stimulation à découvrir des relations rendant possible le traitement d'information par des populations de neurones. La réalisation de systèmes neuronaux artificiels capables de calculer, sera utile pour un éclairage sur les mécanismes naturels. Et ouvrira également peut-être la voie à des applications technologiques inédites. Les cultures neuronales permettent cette exploration. En contrôlant la géométrie des cultures on manipule la connectivité globale du réseau et l'accessibilité de neurones permet de leur appliquer de schémas de stimulation variés. On peut alors tester quelles structures de réseau et quelles conditions de stimulation conduisent à une activité pouvant être interprétée comme un traitement de signal, c'est-à-dire la réalisation d'un calcul.

6.4.4 Réalisation de processeurs neuronaux

Exploitant les techniques de micro-impression le groupe de Elisha Moses à l'Institut Weizmann a été pionnier dans la réalisation de dispositifs à base de cultures neuronales capables d'opérations de calcul. Ces expériences ont permis une réalisation artificielle avec des neurones des fonctions principales attendues dans un système de codage neuronal d'information discutées plus haut à savoir perception de signaux, leur transformation interprétant le contexte, la propagation du signal ou du résultat et sa restitution.

Ces dispositifs ont été construits en imprimant avec des molécules adhésives des schémas de structures sur des lamelles de verre rendues préalablement répulsives par un autre traitement. Les neurones adhèrent à la lamelle uniquement sur les zones imprimées. Seules les surfaces imprimées permettent l'adhésion cellulaire. La réalisation la plus simple a consisté dans la culture de neurones dans un canal, forçant une croissance quasiment uni-dimensionnelle permettant d'étudier la propagation simple d'un signal le long d'un chemin unique. Un système suffisamment simple pour être précisément stimulé et compris quantitativement.

Une sophistication de ces canaux réguliers de croissance a ensuite consisté à rompre la symétrie et produire des canaux asymétriques réalisés sous forme d'une succession de micro chambres de culture triangulaires. Ce mélange de structures linéaires et bi-dimensionnelles conduit à la sélection d'une direction privilégiée de croissance des axones (figure 6.10). Ceci matérialise avec des neurones l'analogie de diodes sélectionnant une direction pour la propagation des signaux, voir figure 6.11.

Ces systèmes ont ensuite été utilisées pour construire des configurations de cultures capables d'exécuter des fonctions logiques simples : une fonction seuil 6.12 et une fonction « AND » 6.13. Les opérations réalisées sont ici la conséquence de l'architecture des cultures qui détermine une fonction.

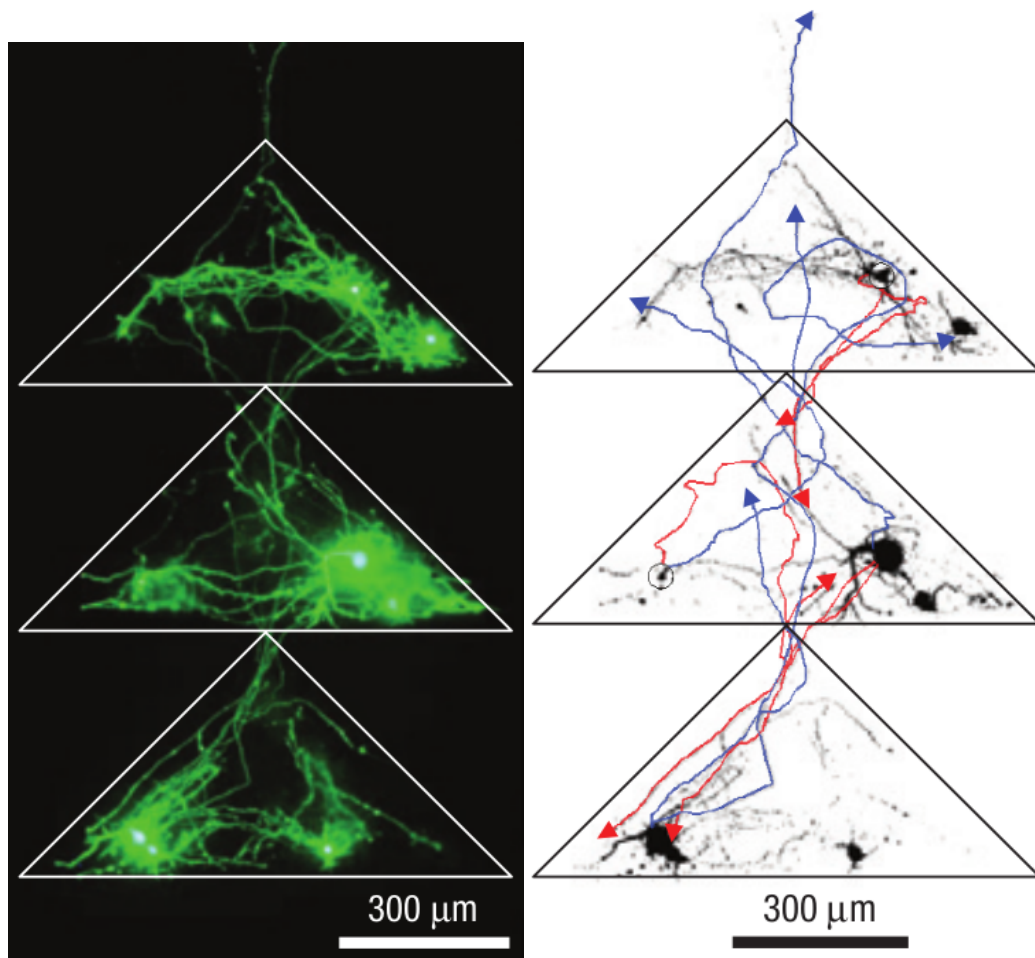


FIGURE 6.10 – **Dispositif à diode.** Gauche : Marquage par de la GFP fluorescence des axones (filaments fins long), somas des neurones (tâches brillantes) et arbres dendritiques (nuage de filaments autour des somas). Droite : reconstruction manuelle des axones traversant les triangles. Les projections vont préférentiellement du bas vers le haut. Source : [FEINERMAN et collab., 2008]

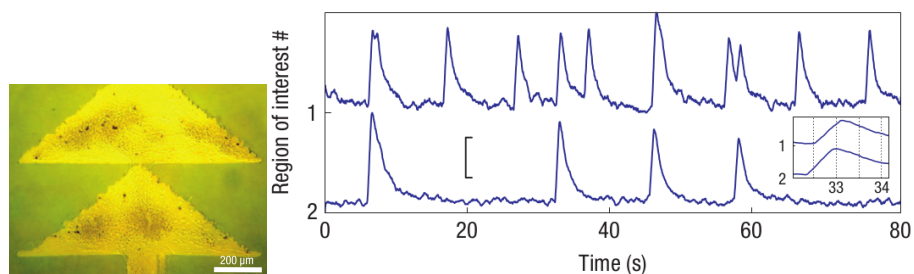


FIGURE 6.11 – **Implémentation d'une diode.** Un signal ne se propage que de la zone d'intérêt 2 à la zone d'intérêt 1, manifestant une croissance directionnelle préférentielle des axones de la zone triangulaire basse à celle supérieure. Source : [FEINERMAN et collab., 2008]

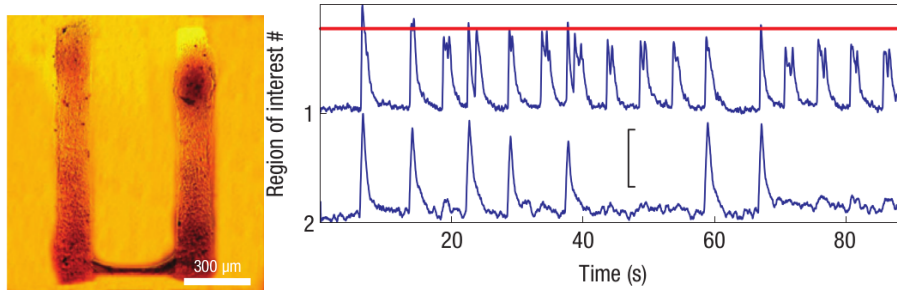


FIGURE 6.12 – **Implémentation de la fonction «Seuil»** A) fonction à seuil. Les graphes à droite indiquent la fluorescence au cours du temps des rapporteurs Ca^{2+} dans les neurones qui signalent leurs déclenchements de potentiels d'action; les pics indiquent l'intensité d'une bouffée de potentiels d'action dans les régions d'intérêt indiquées sur la photo du dispositif à gauche. Les deux colonnes sont reliées par un canal fin mettant en communication les populations neuronales des deux zones. La communication d'une zone à l'autre s'effectue uniquement quand le signal dépasse un seuil, indiqué par la ligne horizontale rouge. Source : [FEINERMAN et collab., 2008]

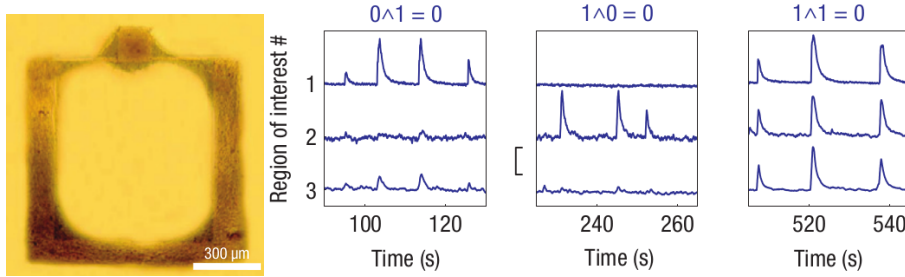


FIGURE 6.13 – **Implémentation de la fonction «AND»**. Les deux chambres droite et gauche communiquent vers la chambre supérieure par de fins canaux. Les signaux individuels des deux chambres latérales sont insuffisants seuls pour le déclenchement de la chambre supérieure (les deux graphes temporels à gauche). Celle-ci n'est activée que lorsque les deux chambres latérales sont actives simultanément (graphe temporel à droite). Source : [FEINERMAN et collab., 2008]

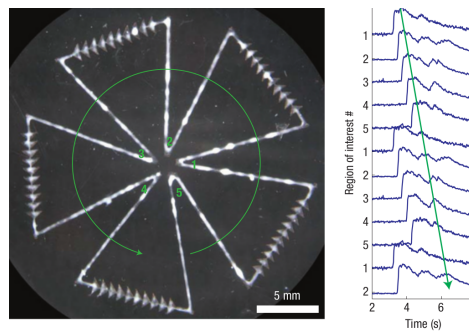


FIGURE 6.14 – **Implémentation d'un oscillateur**. Le dispositif consiste en un canal quasi-unidimensionnel dans lequel sont ensemencés les neurones. Ce canal forme un circuit fermé bouclant sur lui-même et comportant cinq lobes portant le long de leurs segments extérieurs une succession de diodes neuronales (figure 6.11). La chaîne de diodes sélectionne une direction privilégiée de projection des axones et un sens de la connectivité du réseau neuronal s'établissant dans le circuit indiqué par le cercle orienté vert. Les graphiques à droite indiquent les variations dans le temps des activités neuronales mesurées simultanément par fluorescence Ca^{2+} dans les 5 zones d'intérêt indiquées par les chiffres sur l'image. Le décalage des pics d'activité selon les zones successives indiqué par la ligne verte manifeste la circulation orientée de stimulation de potentiels d'actions autour du circuit. Source : [FEINERMAN et collab., 2008]

En combinant ces structures le groupe de Elisha Moses a construit un dispositif plus complexe générant une activité cyclique (figure 6.14) et ont également imaginé d'autres implémentations de fonctions complexes. Cependant les fonctions de ces dispositifs sont uniquement déterminées par la géométrie. Elles sont figées par les arrangements entre populations et les voies de communication. Ces systèmes peuvent être intéressants pour explorer des systèmes de codage et la manière de réaliser des opérations mais les possibilités restent figées. Ils s'apparentent à la réalisation de circuits électroniques avec des composants élémentaires statiques. Les possibilités des dispositifs neuronaux seraient grandement étendues si ces dispositifs pourraient être programmables. C'est-à-dire que pour une géométrie donnée il serait possible de changer le comportement. La manifestation plus élémentaire de ce changement, et l'unité manquante aux dispositifs précédant étant la construction d'une mémoire. Ce que fait formidablement le cerveau grâce à l'élasticité, comme obtenir dans une culture une mémoire d'entrée qui lui seraient envoyées? Dans la section suivante je discuterai donc de la plasticité dans les cultures indispensables pour la mémorisation et l'apprentissage.

Les systèmes proposés par le groupe de Elisha Moses sont fascinants mais restent préliminaires et très élémentaires. Leur conception est inspirée par la logique booléenne et une information codée sous la forme de la présence ou non de potentiels d'action. Cependant ces travaux ouvrent la voie à l'exploration expérimentale du calcul neuronal. On voudrait appliquer ces techniques pour explorer d'autres paradigmes de codage, de programmation et apprentissage avec les cultures de neurones. Aider par cette démarche à déterminer les conditions pour le calcul neuronal.

6.5 Activité spontanées des cultures

Les systèmes neuronaux aussi bien naturels que *in vitro* sont animés au cours de leur développement d'une activité électrique spontanée avec des corrélations dynamiques à grande échelle et la propagation de vagues macroscopiques d'activité [KIRKBY et collab., 2013]. Cette activité autonome, qui ne résulte donc pas de stimuli externes, a été observée sur de nombreux tissus cérébraux. Elle est caractéristique des circuits cérébraux durant les premières phases de leur développement et semble associée à leur formation [BLANKENSHIP et FELLER, 2009]. Elle pourrait alors être impliquée dans la différenciation entre types de neurones, la genèse de synapses et la plasticité synaptique [KHAZIPOV et LUHMANN, 2006]. Son évolution durant la progression de la morphogénèse du cerveau va de pair avec différentes étapes de la formation des circuits neuronaux. En particulier, au bout de quelques semaines ce régime d'activité fortement corrélé disparaît et donne la place à un nouveau régime d'activités neuronales asynchrones et décorréliées dans les réseaux [FELLER, 1999].

L'activité électrique spontanée de neurones représente près de 80% des besoins énergétiques du métabolisme des cerveau [MAZZONI et collab., 2007; RAICHLE, 2006]. Ce niveau élevé doit signifier que cette activité propre, hors du contexte d'un échange immédiat avec l'environnement, doit posséder un rôle important dans le fonctionnement cérébral. L'activité spontanée se retrouve notamment dans les réseaux neuronaux nommés « Central Pattern Generators » (CPG) responsables de la génération de rythmes physiologiques persistants toute la vie, tels la respiration, ou intermittent comme pour la locomotion.

En culture, on observe aussi une activité électrique spontanée. Ses propriétés globales et son évolution dans le temps sont remarquablement similaires à celles observées *in vivo* durant les premières phases du développement avec des séquences répétitives de déclenchements synchronisés à l'échelle de la population entière [WAGENAAR et collab., 2006a]. Un phénomène essentiellement désigné par le terme de « network burst » [VAN PELT et collab., 2004b], mais également par « network spikes » [EYTAN et MAROM, 2006] ou pour des auteurs dans un contexte de neurosciences plus appliquées aussi de « epilepticform events » [FURSHPAN et POTTER, 1989; LITT et ECHAUZ,

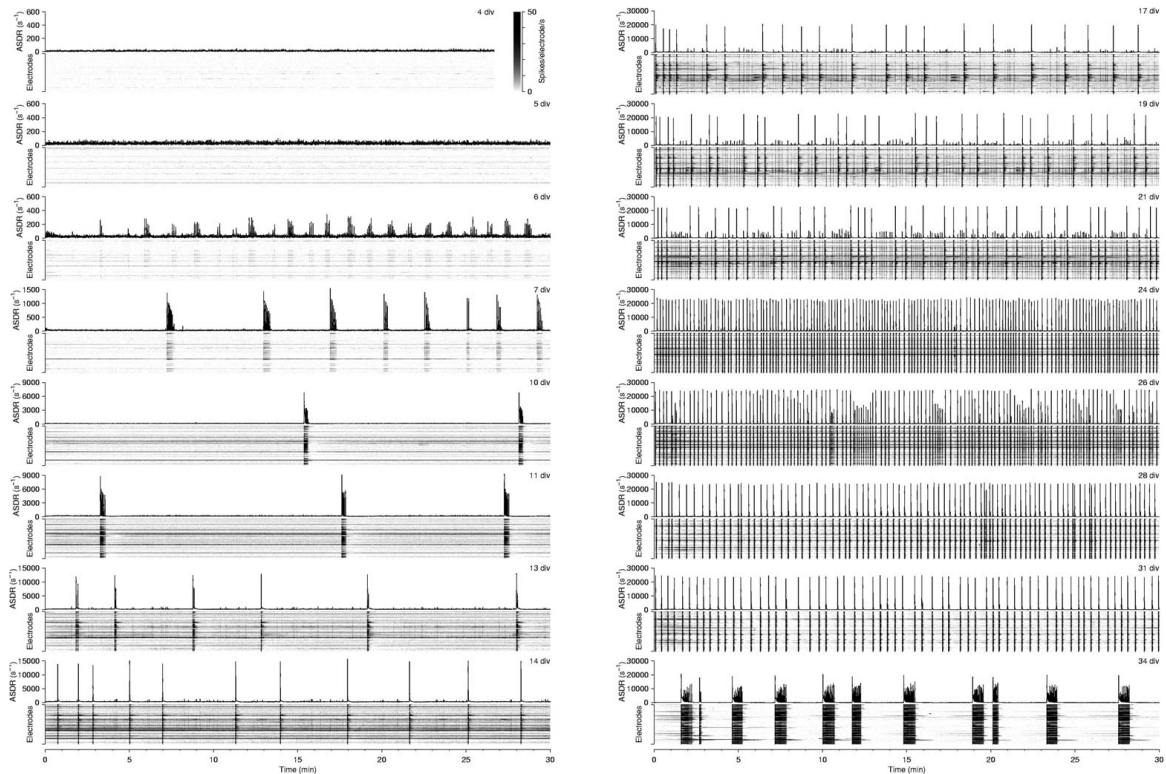


FIGURE 6.15 – **Développement de bouffées de réseau dans une culture dense.** Représentation de l'évolution de l'activité sur une période de 30 minutes sur 34 jours après incubation de la culture. Pour chaque journée, indiquée par le nombre DIV (day in vitro) en haut à droite de chaque série de graphes sont représentées le taux global d'activité de toutes les électrodes (ASDR : array-wide spike detection rate) et le graphe « raster » des 59 électrodes (chaque ligne représente une électrode, un point gris l'enregistrement de son niveau d'activité sur une seconde, selon l'échelle de gris proche de l'indication 4 div). Les bouffées de potentiel d'action sont visibles par les pics du taux global et par une ligne verticale indiquant l'activité simultanée de toutes les électrodes. Les bouffées apparaissent à partir de 6 div connaissent une évolution complexe le long du vieillissement de la culture. Source : [WAGENAAR et collab., 2006a]

2002; WAGENAAR, 2005] et *bouffées de réseau* en français.

Ces manifestations sont une forme de synchronisation collective de la population aussi bien dans les cerveaux que dans les cultures artificielles. Outre de caractéristiques typiques des bouffées de réseau – la présence de deux échelles de temps entre salves brèves de forte activité suivie de périodes de repos – la littérature fait apparaître une grande variété de modes d'activités qui ne font pas consensus sur la dynamique et la statistique spatio-temporelle des événements collectifs.

Les cultures de neurones, ou des régions cérébrale déterminées, sont des systèmes très complexes en raison de la structure de connexions et de la diversité des mécanismes et des facteurs en jeu qui déterminent la richesse des comportements physiologiques et pathologiques. Sont présent dans les populations des types de neurones différents mal caractérisés dans ces environnements. S'y trouvent également des interactions synaptiques différentes, excitatrices, inhibitrices avec plusieurs types de récepteurs; des propriétés des neurones individuels qui peuvent dépendre de conditions expérimentales précises affectant aussi les caractéristiques des interactions, ainsi que des schémas de connectivité non élucidés avec des variations topologiques locales possibles affectant la dynamique. Dans cette diversité, les comportements peuvent être dominés par des types cellulaires particuliers, des événements rares et les propriétés des queues de distributions.

Même au sein d'une même équipe avec des protocoles et conditions expérimentales identiques au mieux les comportements dynamiques des cultures peuvent présenter des comporte-

ments très variés. Les réseaux peuvent se développer de manière différente affectés par des facteurs non contrôlables dans les détails comme la densité d'incubation, la composition cellulaire, la proportion de glies, l'incorporation des marqueurs d'activité ou le couplage avec les électrodes de mesure pour les MEA [PASQUALE et collab., 2008].

6.5.1 Synchronisation et bouffées de potentiels d'action

En effet au bout de 5 à 7 jours après l'incubation les neurones acquièrent une activité électrique sporadique. Après encore quelques jours, un âge 9 à 12 jours *in vivo* cette activité prend de l'ampleur, commence à se structurer et aboutit à des vagues de déclenchements en cascade, finissant par entraîner la culture toute entière dans de brèves séquences de tirs simultanés [COHEN et collab., 2008]. Dans les cultures, les bouffées de réseau se poursuivent durant toute leur durée de leur vie [WAGENAAR et collab., 2006a]. Cette persistance est différente de ce que l'on observe *in vivo* où ce régime évolue vers une activité sporadique de potentiels d'action individuels témoignant d'une maturation du réseau. Cette comparaison a conduit à interpréter l'absence d'évolution du comportement des cultures comme un signe dans celles-ci d'une interruption du développement normal des réseaux *in vivo* [COHEN et collab., 2008; CORNER et collab., 2002; WAGENAAR, 2005].

Pendant les quelques centaines de milli-secondes que dure une bouffée d'activité synchronisée, presque tous les neurones tirent en bouffées avant de retourner au calme pendant une longue période, de plusieurs secondes et à la minute, avant une nouvelle bouffée [EYTAN et MAROM, 2006; MAROM et SHAHAF, 2002; NAPOLI et collab., 2014; VAN PELT et collab., 2004b; WAGENAAR et collab., 2006a]. Chaque bouffée implique pratiquement tous les neurones de la culture qui pendant quelques centaines de millisecondes émettent une rafale de déclenchements – la bouffée d'activité – avant le retour à une période inactive plus longue d'une durée de quelques secondes à la minute avant une nouvelle bouffée. Ces comportements sont très généraux et ont été retrouvés par de nombreux auteurs dans des cultures de différentes origines avec de neurones dissociés issus de cortex, d'hippocampe de rongeurs, ainsi que de ganglions de sauterelles [MAZZONI et collab., 2007] avec des propriétés et une dynamique très similaires.

Ainsi que je l'ai déjà indiqué, au-delà de leur comportement global distinctif, les bouffées d'activité spontanée collective présentent une grande diversité de caractéristiques particulières [VAN PELT et collab., 2004a,b; WAGENAAR, 2005] intervalles de temps entre les événements, nombre de déclenchements dans une bouffée, dynamique d'initiation, qui dépendent du système neuronal, des conditions de culture même si globalement leur forme macroscopique est similaire. En particulier les conditions de culture et l'âge semblent intervenir dans la cohésion du comportement collectif synchronisé. VAN PELT et collab. [2004a,b] ont montré un resserrement de l'activation des bouffées avec un déclenchement de toute la population soudain, sans séquence progressive de déclenchement. Une telle précision de synchronisation de phase a été également mise en évidence par PENN et collab. [2016] avec un milieu de culture chimiquement très contrôlé, sans sérum de cheval communément utilisé dans les protocoles de culture.

Cette complexité se retrouve aussi dans les séquences d'activations des bouffées qui ne sont pas aléatoires mais démontrent des formes d'organisation complexe comme l'existence de sous-ensembles de neurones se retrouvant régulièrement actifs en premier et agir en tant que précurseurs des vagues de déclenchements [ECKMANN et collab., 2008; EYTAN et MAROM, 2006; ORLANDI et collab., 2013]. La répétition de séquences particulières d'activité avec une récurrence précise sur de longues périodes a aussi été mise en évidence par ROLSTON et collab. [2007] dans des cultures corticales de rat dissociés. Ces séquences se retrouvent aussi bien dans les phases préliminaires de bouffées collectives de réseau que dans de l'activité sporadique entre des bouffées. Eytan et Marom ont également mise en évidence l'existence de neurones qui déclenchent de manière consistante dans les toutes premières phases des bouffées de réseaux alors que d'autres sont systématiquement dans les phases tardives [EYTAN et MAROM, 2006].

Malgré l'intérêt qu'à suscité depuis longtemps ce type de dynamique, ses propriétés et les mé-

canismes en œuvre restent encore très mal compris. En particulier les phénomènes responsables de l'initiation des bouffées restent confus et leur relation avec la morphologie du réseau sous-jacente. Plusieurs hypothèses ont été proposées à partir de différentes approches expérimentales et théoriques. L'initiation des bouffées pourrait ainsi dépendre de l'existence de types de neurones fortement connectés soumis à un bruit synaptique plus important facilitant leur déclenchement et la propagation d'activité à un grand nombre d'autres neurones [ECKMANN et collab., 2008; EYTAN et MAROM, 2006]. Une autre hypothèse implique un jeu complexe entre propriétés locales de connectivité du réseau et propagation du bruit pour créer des zones de nucléation à l'origine des cascades d'activité percolant sur toute la population [ORLANDI et collab., 2013].

Un certain nombre d'autres études [PENN et collab., 2016; RAMIREZ et collab., 2004; SURESH et collab., 2016] compliquent encore l'image en suggérant la présence d'une fraction significative de neurones dotés d'une activité propre périodique. Ceux-ci pourraient constituer une sous-population « pacemaker » responsable de la génération des rythmes collectifs. Les résultats de nos collaborateurs E. Moses et M. Segal [PENN et collab., 2016] dans des environnements de culture chimiquement très contrôlés suggèrent indiqueraient même qu'une grande majorité des neurones pyramidaux en culture issues d'hippocampe de rats seraient des oscillateurs intrinsèques. Le comportement d'ensemble de bouffées de réseau résulterait alors de la synchronisation collectives d'oscillateurs à relaxation. Une question dont les fondements théoriques sont abordés dans mes études discutées plus loin, dont un sujet particulier est l'explication de cette forme particulière de synchronisation collective avec de deux échelles de temps rapide inter-bouffées et lente entre les répétitions des bouffées. La synchronisation elle-même d'oscillateurs à relaxation semble être une propriété assez générique [BOTTANI, 1995; SOMERS et KOPELL, 1993].

Un autre exemple de l'effet de conditions de culture exactes se trouve dans [EYTAN et MAROM, 2006]. Les neurones corticaux de cultures dissociés de rats maintenues dans un milieu fait de sérum de cheval et nutriments en perfusion ont une activité spontanée avec un taux de déclenchements similaire à celui *in vivo*. Quand le renouvellement du milieu est interrompu, le réseau converge vers un mode d'action synchronisé pratiquement sans activité entre les bouffées.

Les poids relatifs de connexions excitatrices et inhibitrices dans les réseaux affectent également dans le degré de synchronisation observé dans l'activité des cultures. Ainsi le blocage de connexions synaptiques inhibitrices par des antagonistes des récepteurs GABA conduit les réseaux à des régimes caractérisée par davantage de grandes bouffées fortement synchronisées. Au contraire si l'excitation est réduite par des inhibiteurs des récepteurs synaptiques NMDA, les grandes bouffées sont réduites. [MAZZONI et collab., 2007]. Ces résultats indiquent comment les observations peuvent être déterminées par les conditions particulières entre les sous réseaux excitateurs obtenus dans les cultures. Même si le ratio entre neurones excitateurs et inhibiteurs est supposé être identiques, des variations sont possibles et les connexions sujettes à des variations subtiles expérimentales.

La disponibilité de méthodes d'observation avec une résolution temporelle élevée ont permis depuis une quinzaine d'années d'étudier plus finement la dynamique spatio-temporelle des réseaux neuronaux *in-vivo* et *in-vitro*. Il a fallu en effet attendre les travaux de Beggs et Plenz en 2003 pour identifier une dynamique spatio-temporelle d'avalanches de déclenchements, différente et de plus petites taille que les bouffées synchronisées collective qui seulement parfois en sont l'aboutissement. Ainsi des événements plus petits peuvent se produire, s'apparentant à des bouffées avortées qui n'auraient pas réussi à entraîner la population entière. Des avalanches de déclenchements, initiés spontanément par quelques nœuds du réseau par les causes méconnues discutées plus haut, et dont les déclenchements stimulent les nœuds auxquels ils sont connectés provoquant une chaîne de réactions en cascade. Ces avalanches peuvent constituer les phases initiales des bouffées de réseau, recrutant toujours plus de neurones dans une cascade de déclenchements avant l'entraînement explosif soudain de la population entière. Beggs et Plenz ont cependant montré la présence d'avalanches n'atteignant jamais le stade de recrutement global et pouvant avoir une variété de tailles et de durées. Ces avalanches neuronales constituent depuis

quinze ans un des sujets les plus débattus et controversés sur les propriétés cérébrales fondamentales pour le traitement d'information.

6.5.2 Avalanches et comportements invariants d'échelle

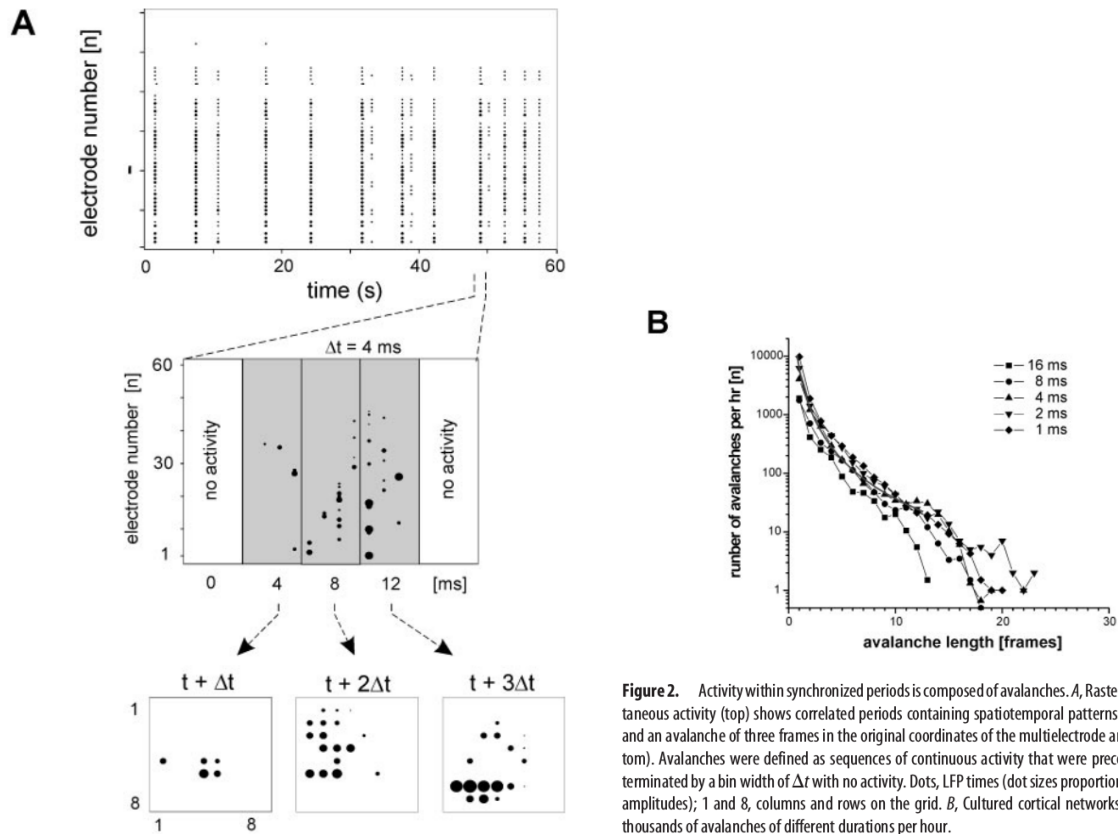


Figure 2. Activity within synchronized periods is composed of avalanches. *A*, Raster of spontaneous activity (top) shows correlated periods containing spatiotemporal patterns (middle) and an avalanche of three frames in the original coordinates of the multielectrode array (bottom). Avalanches were defined as sequences of continuous activity that were preceded and terminated by a bin width of Δt with no activity. Dots, LFP times (dot sizes proportional to LFP amplitudes); 1 and 8, columns and rows on the grid. *B*, Cultured cortical networks produce thousands of avalanches of different durations per hour.

FIGURE 6.16 – Activité sous forme d'avalanches au sein de bouffées synchronisées. *A* : La trame de l'activité spontanée (en haut) montre des séquences contenant des schémas corrélés spatio-temporels (milieu), un exemple en trois images successives suivant les enregistrements des microélectrodes (en bas). Les avalanches sont définies comme des séquences d'activité continue précédées et suivies d'intervalles sans activité. La surface des cercles est proportionnelle à l'intensité du potentiel local de l'électrode à cet endroit à l'instant de l'image. *B* : Distribution des tailles des avalanches de cultures corticales selon différentes durées des intervalles d'analyse. Source : figure 2 de Beggs et Plenz en 2003 [BEGGS et PLENZ, 2003]

L'existence et les propriétés d'avalanches neuronales a attiré une attention considérable depuis la publication de Beggs et Plenz en 2003 de résultats montrant que les tailles de ces événements n'auraient pas d'échelle caractéristique et seraient distribuées suivant des lois de puissances (voir figure 6.16). Un mode de comportement collectif qui se retrouve dans différents modèles théoriques d'avalanches de la théorie des systèmes critiques auto-organisés. Ces modèles, par exemple ceux dits « de feu de forêt », du tas de sable », des « tremblements de terre » produisent en effet des avalanches de tailles distribuées selon des lois de puissance qui suggèrent une organisation spontanée du système statistique dans un état de corrélations spatio-temporelles à toutes les échelles rappelant l'état dit « critique » lors d'une transition de phase de second ordre. Les avalanches dans les réseaux de neurones et leur distribution de taille en loi de puissance seraient ainsi les signes de comportements émergents collectifs corrélés à toutes les échelles dans le fonctionnement cérébral. Un mode hautement organisé peut-être nécessaire pour des fonctions cognitives.

Avec des mesures d'activité spontanée à l'aide de matrices d'électrodes sur des tranches cervicales de rat, Beggs et Plenz en 2003 [BEGGS et PLENZ, 2003] ont en effet montré la propagation spontanée d'avalanches d'activité avec des propriétés similaires à celles de systèmes physiques

critiques. À la fois la distribution des tailles des avalanches que leurs durées sont des lois de puissances dont les exposants correspondent aux valeurs champ-moyen de processus de branchement critiques ($-3/2$ pour les tailles, -2 pour les durées) [ZAPPERI et collab., 1995]. Ces exposants et le comportement critique paraissent par la suite avoir été confirmés sur des résultats plus étendus et une analyse statistique plus complète, vérifiant notamment l'invariance d'échelle entre systèmes de dimensions différentes [KLAUS et collab., 2011]. Des résultats avec des exposants identiques ont été aussi suggérés dans d'autres systèmes *in vivo* sur des rats anesthésiés [GIREESH et PLENZ, 2008] et sur de singes éveillés [PETERMANN et collab., 2009].

L'observation de telles lois de puissance et leur interprétation comme la manifestation d'un état critique de la dynamique cérébrale a stimulé de nombreuses idées sur leur signification. Cet état aurait notamment la conséquence majeure d'optimiser la transmission d'information dans le réseau tout en préservant la stabilité de celui-ci, c'est-à-dire sans emballement global. Les corrélations à toutes les échelles correspondant à un état critique signifieraient que les réseaux se trouveraient entre une phase où les activités seraient plutôt locales et s'estomperaient rapidement, et une phase d'amplification de l'activité à grande échelle. Dans la phase critique intermédiaire le traitement d'information serait optimisé car des activités de toutes échelles spatio-temporelles sont possibles [VAN KESSENICH et collab., 2016].

Ces idées ont naturellement stimulé la recherche de comportements analogues dans les cultures, dans lesquelles se posent en outre la question de l'auto-organisation de la structure de connectivité du réseau lui-même. Plusieurs groupes ont ainsi recherché des avalanches invariantes d'échelles dans la dynamique spontanée, ou induite, de cultures neuronales de différents types. Si comme le montre une littérature abondante citée dans [EYTAN et MAROM, 2006] les bouffées de réseau qui impliquent pratiquement tous les neurones sont le comportement macroscopique le plus notable, les mesures avec matrices d'électrode montrent également l'existence dans certaines situations de bouffées avortées de petite taille survenant entre grands événements synchronisés [EYTAN et MAROM, 2006]. Des manifestations peut-être significatives et pouvant manifester un état critique du réseau.

Dans les cultures neuronales l'existence d'avalanches et d'un état critique paraît en réalité plutôt controversée. À de nombreuses publications affirmant ce comportement auto-organisé dans les cultures s'opposent à la fois des résultats contradictoires d'expériences très précises et des interprétations différentes des données. Sans prétendre que ce chapitre soit une revue de ce sujet difficile en raison des nombreuses subtilités expérimentales, je rapporterai les faits qui m'ont paru essentiels sur cette question des avalanches et de la criticalité importante pour la compréhension des cultures et du fonctionnement neurobiologique. Une lecture qui m'invite à une grande prudence quant aux affirmations générales trop directes concernant la criticalité de cultures neuronales.

Étant donné la généralité des bouffées de réseaux à grande échelle, les avalanches de tailles différentes doivent être recherchées dans des événements moins visibles, tels des bouffées avortées, des événements sporadiques entre bouffées, ou les fronts de recrutements en cascade dans la montée d'une bouffée. Comme discuté plus haut, la dynamique précise de la synchronisation collective manifestant une bouffée est très dépendante des conditions expérimentales et de l'âge de la culture. Des situations particulières pourraient être nécessaires pour permettre une variabilité suffisante dans le temps de la dynamique spontanée autorisant des événements de tailles différentes. Ainsi des définitions différentes d'avalanches neuronales ont été introduites dans la littérature.

En effet, face à la difficulté d'identifier sans ambiguïté dans l'activité des cultures des avalanches neuronales claires au sens de modèles de la criticalité auto-organisé, des définitions différentes ont été introduites pour mesurer de tels événements. Comme il n'existe pas de manière standard de définir une bouffée (burst), parler de distributions de leurs tailles et de leurs durées n'est pas univoque. Enregistrés typiquement avec des matrices d'une soixantaine d'électrodes, pour l'essentiel les avalanches sur le réseau sont considérées comme des événements d'activité

enregistrée présentant un taux de déclenchements suffisant (supérieur à un seuil) sur une durée suffisante (au moins une certaine durée). Le choix des seuils possède une part d'arbitraire et correspond à différents critères pour isoler des séquences considérées comme indépendantes. Les tailles des avalanches seront alors soit le nombre total de déclenchements mesurés durant une telle séquence, soit le nombre, très limité, d'électrodes ayant répondu.

La première publication à ma connaissance discutant d'une recherche d'avalanches en culture et de lois de puissance est en 2006 par Eytan et Maron [EYTAN et MAROM, 2006] avec des expériences de matrices d'électrode sur des cultures corticales de rat mesurant de manière particulièrement précise la dynamique des bouffées de réseau (« network spikes » dans leur terminologie). Ils ont mis en évidence des régularités importantes dans les séquences d'initiation des bouffées avec la présence répétitive des mêmes neurones « early to fire » et un processus de recrutement « hiérarchique », signifiant probablement une propagation en cascade d'activité, durant la montée de la synchronisation d'une bouffée. Reprenant le type d'analyse que Beggs et Plenz avaient opérés sur des coupes cervicales [BEGGS et PLENZ, 2003] ils n'observent pas de distribution de lois de puissance équivalente. Leurs observations donnent, très différemment, une distribution bimodale manifestant que la culture se comporte avec une réponse collective d'ensemble, ou très petite réponse (« all-or-nothing nature of the network spikes »). Une conclusion similaire a été obtenue dans l'équipe de Steve Potter [MADHAVAN et collab., 2007] aussi sur des cultures de neurones de cortex de rat, obtenant une distribution de tailles de bouffées bimodale très différente d'une loi de puissance. Une autre étude de la même équipe [ROLSTON et collab., 2007] a complété l'investigation en regardant les distributions de tailles non pas de bouffées mais de séquences récurrentes d'activité de certaines électrode se répétant de manière stable. Ces distributions d'événements particuliers stables s'apparentent sur une décade à une loi de puissance avec un autre exposant que celui de Beggs et Plenz. Mais la faible étendue de l'ajustement ne permet pas de conclure sur la nature réelle de la distribution et la criticalité de la culture. Indépendamment de ces conclusions négatives par des groupes experts sur les cultures, d'autres auteurs ont cependant publié des résultats positifs.

MAZZONI et collab. [2007] est probablement la première suggestion d'avalanches suivant des lois de puissance en culture dans des expériences à partir de neurones dissociés d'hippocampe de rat. Ils affirment l'existence de lois de puissance avec des exposants identiques à ceux de [BEGGS et PLENZ, 2003]. Les courbes de distributions rapportées (leur figure 9) sont cependant difficilement lisibles et l'ajustement sur deux ou trois décades des données avec la loi de puissance incertain.

Aussi sur des cultures dissociées à partir de cortex de rat, PASQUALE et collab. [2008] rapportent des lois de puissance en comptabilisant comme avalanche toute séquence d'activité enregistrée sur leur matrice de soixante électrodes non interrompue par plus d'un faible intervalle de temps de silence de 2 ms. Y compris des phases de recrutement de bouffées de réseaux contribuent alors à plusieurs avalanches. La publication conclue à l'observation d'un état critique avec des exposants rapportés $-3/2$ pour les tailles et $-1/2$ pour les durées correspondent aux valeurs de Beggs et Plenz. L'ajustement des données à une loi de puissance porte cependant sur à peine plus d'une décade.

Depuis ces études datant d'une dizaine d'années la question de la criticalité des tissus neuroaux, de toute nature cerveau, coupe ou culture, a continué à susciter de vifs débats. Des tribunes enthousiastes revendiquent la nature fondamentalement auto-organisée et critique de la dynamique cérébrale [CHIALVO, 2010]. D'autres auteurs cependant ont pointé des limites fondamentales dans les analyses et données des arguments des travaux initiaux dans ce sens. Notamment l'usage exclusif du critère des lois de puissance pour la distribution des tailles et durées des événements pour conclure de la criticalité. Un critère en outre appliqué sur des jeux de données très limités en terme d'échantillons et d'intervalles très réduits de valeurs couvertes par les distributions. Une mode semble avoir eu lieu entre la fin des années 90 et le tournant du siècle voyant des lois de puissance dans une multitude de réseaux complexes naturels et techniques, dont les réseaux neuroaux. Mais comme présenté en détail par CLAUSET et collab. [2009] la détection et la caractérisation de lois de puissance est une opération délicate sujette à mauvaises conclusions

et fortes erreurs d'estimation. Les événements les plus grands, des queues de distributions sont les plus rares, et peu présents dans les données, donc incertaines dans les grandes tailles, rendant difficile l'identification des lois de puissance. Touboul et Destexhe ont aussi montré [TOUBOUL et DESTEXHE, 2010] que même si avérée la présence de loi de puissance dans les distributions d'événements n'est pas un critère suffisant pour conclure à la criticalité : l'observation expérimentale de lois de puissance peut aussi s'expliquer par des propriétés génériques de processus stochastiques avec un seuil. Comme ces auteurs le développent dans une autre publication [DEGHANI et collab., 2012] d'autres mécanismes stochastiques peuvent aussi générer des distributions en loi de puissance sans une situation critique statistique [BÉDARD et DESTEXHE, 2009; BÉDARD et collab., 2006; GISIGER, 2001; REED et HUGHES, 2002].

La présence trompeuse de lois de puissance est également une conclusion des expériences menées par le même groupe [DEGHANI et collab., 2012]. L'activité mesurée *in vivo* chez le chat, le singe et l'humain par des matrices d'électrode montre en effet des avalanches. Mais si superficiellement leurs distributions peuvent s'apparenter à des lois de puissance, elle sont mieux ajustées par d'autres distributions (bi-exponentielle). Quels que soient les conditions ces résultats ne montrent aucune preuve claire de loi de puissance et de criticalité.

6.5.3 Dynamique neuronale et inhibition

Une question sur les propriétés du codage neuronal possible à aborder avec les cultures concerne le rapport entre excitation et inhibition dans les réseaux. Ce rapport qui semble avoir un rôle essentiel dans le fonctionnement cérébral est encore très mal compris. Il correspond au bilan de l'ensemble des contributions de chaque type dans les courants synaptiques, que ce soit en volume dans une population ou bien intégré dans le temps. L'équilibre excitation sur inhibition semble essentiel pour permettre une bonne représentation des informations sensorielles et des opérations neuronales (voir [ZHOU et YU, 2018] et les citations qui s'y trouvent). Un exemple de cette importance est le déséquilibre de ce rapport lors de différents troubles cognitifs comme dans l'autisme et la schizophrénie [MURRAY et collab., 2014; YIZHAR et collab., 2011]. Les différentes observations de situations perturbées suggèrent que le rôle essentiel des neurones inhibiteurs serait de stabiliser l'activité neuronale. Dans un système purement excitateur l'activité pourrait trop facilement s'emballer et se propager sur tout le cerveau [AVIEL et collab., 2003].

Des expérimentations avec des cultures devraient également aider à clarifier comment ce ratio affecte la dynamique spontanée ou induite d'une population neuronale. De telles expériences peuvent se réaliser à l'aide de bloqueurs de synapses de chaque type, mais depuis peu aussi au laboratoire d'Elisha Moses par l'incubation de cultures avec de neurones triées préalablement selon leur type inhibiteur ou excitateurs. Ces systèmes permettront de directement observer des conséquences d'une altération artificielle du rapport et également étudier comment la plasticité neuronale intervient. Même si ces hypothèses sur les neurones inhibiteurs restent à confirmer, la constatation de leur présence cruciale dans les cerveaux doit en tout cas être prise en compte dans le cadre de la conception de dispositifs synthétiques neuronaux.

6.6 Programmer les cultures neuronales

« Du reste, les réminiscences se produisent parce que tel mouvement vient naturellement à la suite de tel autre. Si cette succession de mouvements est nécessaire, il est évident que quand tel mouvement aura lieu, il déterminera l'autre aussi. Si cette succession n'est pas nécessaire, mais simplement habituelle, il est seulement probable que le second mouvement aura lieu après le premier. »

[ARISTOTELES], cité dans [MARKRAM, 2011]

6.6.1 Codage neuronal et dynamique collective

Neurosciences, informatique et mathématiques ont un long historique d'influences croisées. En particulier depuis l'origine des ordinateurs généraux électroniques. L'article de 1943 « A logical calculus of the ideas immanent in nervous activity » par le biologiste Warren S. McCulloch et le mathématicien Walter Pitts est connu pour être le premier article avec un modèle mathématique de réseau neuronal [MCCULLOCH et PITTS, 1943]. McCulloch et Pitts ont les premiers suggéré un fonctionnement des neurones comme des portes logiques. Ayant ensuite démontré qu'un réseau d'opérateurs logiques peut servir de calculateur universel McCulloch et Pitts ont montré, avec la prémisse précédente, l'équivalence d'un point de vue calculatoire d'un tel réseau à une machine de Turing.

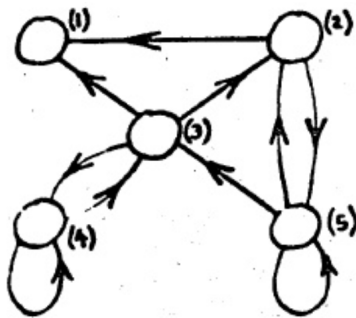


FIGURE 6.17 – L'« **unorganized machine** » de Turing. Émouvant premier schéma par TURING, de système de calcul construit sur un réseau complexe dans son rapport intitulé « Intelligent Machinery » de 1948. L'activité de chaque nœud est déterminée par celle provenant des liens en entrée. Ce principe de calcul distribué sur un réseau complexe est celui des réseaux génétiques et des réseaux neuronaux.

Peu de temps après, en 1948, Alan Turing lui-même s'est intéressé aux propriétés calculatoires du cerveau vu comme un système distribué d'éléments interconnectés capable de remodeler leurs connexions [BIOWALL; TEUSCHER, 2001; TURING, 1948]. Dans son rapport à la diffusion très confidentielle datant de 1948 il suggère la création d'ordinateurs s'inspirant des systèmes cérébraux bien avant tout essai d'ordinateur neuromorphique, et même avant la mise en opération du premier ordinateur électronique généraliste, voir figure 6.17.

L'activité neuronale spontanée est le comportement de fond sur lequel s'établit tout traitement d'une information en réponse à un signal en entrée. Elle affecte en conséquence les propriétés du codage neuronal et il est important de l'étudier. Il en va de même dans le contexte de systèmes en culture. De nombreux thèmes d'étude sont alors d'actualité pour lesquels les cultures sont un outil privilégié.

6.6.2 Codage dynamique de l'information

La manière par laquelle les réseaux neuronaux biologiques stockent et traitent l'information de manière fiable est encore largement mal connue. Plusieurs principes ont été proposés basés sur l'observation de la plasticité cérébrale qui se manifeste avec la création ou la destruction de connexions mais également par la modulation des intensités des interactions entre neurones. Une des théories les plus influentes du traitement et de la mémorisation de l'information par le

cerveau est celle proposée par Donald Hebb [HEBB, 1949] selon laquelle ces phénomènes correspondent à la dynamique des neurones et de leurs interactions dont les propriétés évoluent sous l'effet rétro-actif de la dynamique elle-même. Hebb a supposé l'existence d'« assemblées de cellules neuronales » dont les comportements statiques et dynamiques en réponse à des entrées spécifiques constitueraient les représentations neuronales des sensations, images et concepts. Les différents états de mémoire correspondraient alors à différents attracteurs spatio-temporels stables possibles de la dynamique.

De nombreuses observations électrophysiologiques ou par imagerie fonctionnelle ont mis en évidence l'existence des groupes de neurones dotés d'activités spatio-temporelles répétitives particulières [ABELES et collab., 1993; AERTSEN et collab., 1996; BIENENSTOCK, 1996]. Ces ensembles donnent corps aux « assemblées de cellules avec des comportements d'ensemble spécifiques » suggérées par Hebb dans le cadre de sa théorie. Des ensembles qui s'interprètent comme des cellules reliées dynamiquement [GÜTIG et SOMPOLINSKY, 2006] convergeant vers des états organisés particuliers qui manifestent un apprentissage et un processus cérébral. Les neurones de telles assemblées synchroniseraient leurs activités en réponse à certains stimuli ou bien aussi spontanément en absence de stimulation.

Expressivité des réseaux et dynamiques spatio-temporelles.

EYTAN et MAROM [2006] assimilent les bouffées de réseau des cultures à l'activité orchestrée d'assemblées neuronales au sens de Hebb. Les temps caractéristiques d'activation d'une bouffée de réseau synchronisée, qu'ils ont très précisément mesuré pour des cultures avec des densités d'incubation différentes, sont cohérents avec les durées (0,1-0,2 ms) observées *in vivo* d'assemblées de cellules neuronales et semblent invariants avec la taille de la population. Ces résultats sur des cultures suggèrent ainsi la pertinence de l'intuition de Hebb d'un codage d'information dans l'activité collective de populations de neurones. Un codage de l'information traitée par le cerveau donc lié à l'auto-organisation de l'activité des neurones plutôt qu'à la mise en place de schéma anatomiques ou structuraux des neurones et leurs circuits dans les différentes zones cervicales. Ce sont des patrons d'activité collective qui manifesteraient l'information mémorisée et ses transformations. D'autres résultats sur des cultures par WAGENAAR et collab. [2006b] et BONIFAZI et collab. [2005] vont aussi dans ce sens en illustrant comment différents schémas d'activité temporels stables sont possibles. Chaque mode d'activité spontanée correspondrait à un état cognitif du réseau. La multiplicité des modes montre la possibilité de représenter un grand nombre d'informations.

L'existence de schémas récurrents d'activité pour le traitement et stockage d'information observée dans les cultures suggère que cette propriété est un résultat auto-organisé de populations de neurones. Qu'en est-il alors de la capacité de l'ensemble des cellules neuronales à coder de l'information dans la dynamique des bouffées de potentiels d'actions et de la capacité de changer d'état? En d'autres termes, quelle est la capacité de cultures neuronales d'admettre une multitude d'états stables différents pouvant changer sous l'effet de la plasticité?

Les observations dans des cultures de neurones de comportements réguliers récurrents à l'occasion de bouffées d'activité suggèrent que l'apparition de schémas réguliers serait une propriété générale d'auto-organisation de réseaux [ROLSTON et collab., 2007] ne dépendant pas de la structure précise anatomique des circuits cérébraux *in vivo*. Cependant la possibilité d'une activité synchronisée récurrente d'une population donnée de neurones activée suffisamment rapidement avec une précision de l'ordre de la milliseconde n'est qu'un des ingrédients nécessaires pour les codage et stockage d'information dans des schémas spatio-temporels cervicaux. Dans leur publication sur la dynamique critique neuronale BEGGS et PLENZ [2004] ont argumenté de la nécessité d'une diversité suffisante des schémas possibles d'activité collective d'une population neuronale pour que ces schémas spatio-temporels stables puissent être les éléments des représentations de l'information. Si la seule dynamique collective possible était une synchronisation collective par-

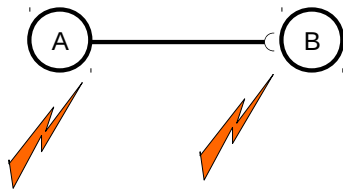


FIGURE 6.18 – **Modèle classique de renforcement synaptique suivant Hebb.** Un stimulus active les cellules A et B. La réponse « simultanée » des deux cellules, c'est-à-dire dans un intervalle temporel suffisamment proche, conduit à un renforcement de la synapse entre A et B (ou à une réduction si le neurone A est inhibiteur). La règle de Hebb ne tient pas compte de l'ordre d'activation des deux neurones.

faite de tous les neurones alors les capacités de stockage d'information seraient très limitées. L'idée d'un état critique, au sens de la physique statistique, du fonctionnement cérébral, dans lequel l'activité collective des neurones présenterait une statistique invariante d'échelle sur toutes les tailles possibles est pour cela très attractive. Les corrélations à toutes les échelles spatiales et temporelles qui seraient le cas garantiraient le maximum d'expressivité possible au système.

Mais la criticalité n'est peut-être pas une nécessité. EYTAN et MAROM [2006] ont par exemple mis en évidence la richesse des comportements collectifs possibles même dans la situation de bouffées globales de réseau en culture. Les différents modes correspondent alors à des zones d'initiation récurrentes différentes, ainsi qu'à différents historiques de recrutement hiérarchique, en cascade lors de la formation de bouffées. Si plusieurs populations neuronales sont couplées entre elles, l'historique de formation d'une bouffée au sein de l'une peut affecter les délais temporels pour la mise en communication des deux populations. En fonction des neurones particuliers connectant les populations, qu'ils soient des neurones précurseurs ou tardifs dans une bouffée, l'écart temporel entre les réponses collectives des deux populations peut être très différent. Il serait y compris possible d'obtenir une inversion de l'ordre temporel d'activité macroscopique des deux populations par rapport à l'ordre de causalité de communication de l'une vers l'autre (dans le cas d'une connexion d'une population X vers une population Y via les projections d'un neurone précurseur de X, la population cible Y peut collectivement répondre avant la réponse collective de la population X).

Différents schémas possibles d'activité collective spatio-temporelle dans des populations neuronales en culture ou *in vivo*, peuvent donc probablement être la manifestation de l'encodage d'information dans ces assemblées. Se pose alors la question de la manière de réaliser sur une population le stockage d'une information sous cette forme dynamique. Autrement dit l'apprentissage par la population neuronale. Plus spécifiquement la situation qui nous intéresse est l'apprentissage dans les cultures.

6.6.3 Apprentissage dans les cultures

Plasticité neuronale

La théorie la plus influente est celle proposée déjà il y a longtemps par Donald Hebb, suivant laquelle le stockage et le traitement de l'information se produit dynamiquement entre des assemblées de cellules reliées en fonction de leur activité combinée. Les règles de Hebb stipulent que l'activité récurrente des mêmes connections conduit à un renforcement persistant dans le temps du lien dont l'intensité s'affaiblit sans activation conjointe. Ce principe où l'information dépend d'un schéma spatio-temporel d'activité précis a été repris par de nombreux auteurs depuis son introduction par Hebb en 1949 et démontré dans de nombreux contextes. Il s'appuie sur les mécanismes maintenant bien établis de plasticité synaptique à long terme, que ce soit pour une potentialisation à long terme renforçant le poids des synapses, ou de dépression à long terme pour une diminution de celui-ci, figure 6.18. La découverte par Henri Markram d'une dépendance de ces effets de l'écart temporel entre l'activité des neurones pré- et post-synaptiques affine les règles de Hebb en précisant davantage encore l'importance de l'agencement temporel précis de l'activité dans les circuits neuronaux. (On pourra se référer à [MARKRAM, 2011] pour une revue historique passionnante de la plasticité synaptique.)

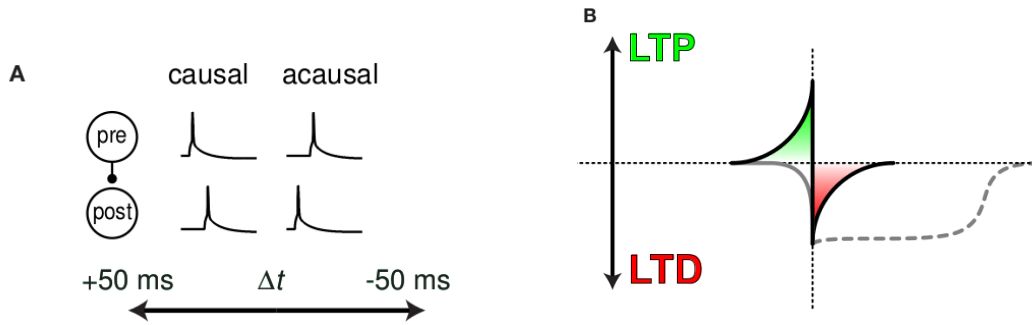


FIGURE 6.19 – **Spike Time Dependent Plasticity (STDP)**. A) L'activité de manière répétitive d'un neurone présynaptique juste avant celle d'un neurone postsynaptique auquel il est connecté contribue à renforcer l'activation du second. Le contraire dans le sens non causal si le neurone post-synaptique est de manière répétée actif avant celui pré-synaptique. B) Dans la plasticité STDP classique, l'activation causale induit une potentialisation à long terme (long-term potentiation –LTP–) et l'activation non causale induit une dépression à long terme (long-term depression –LDP–). L'intensité des effet dépend de l'intervalle de temps entre les deux activations, jusqu'à quelques dizaines de millisecondes. Source : [MARKRAM, 2011]

On peut relier une telle exigence sur le codage de l'information neuronale à une exigence similaire dans la précision du câblage des neurones dans des circuits finement agencés pour assurer la précision temporelle nécessaire. Cependant même si les circuits neuronaux dans le cerveau semblent obéir à une structuration régulière importante, au niveau cellulaire les détails des ramifications neuronales et des localisations des synapses sont aléatoires. La précision des connections entre neurones dans le cerveau est variable entre différentes régions. Comprendre la relation entre variabilité des connections dans le cerveau d'une part et les propriétés de codage d'information d'autre part est particulièrement intéressant. Cette question traite en effet de la balance entre la fiabilité du codage d'information et les capacités d'adaptation et d'apprentissage par la plasticité des synapses et du réseau.

L'étude du stockage et du traitement de l'information dans des cultures est un moyen pour aborder ces questions. Dans quelle mesure le désordre inhérent aux cultures permet de préserver des fonctionnalités permettant un codage d'information neuronal? La structuration précise des circuits cérébraux est-elle indispensable? L'est-elle pour le schéma de Hebb? D'autres mécanismes neuronaux de codage sont-ils possibles?

Plasticité synaptique dans les cultures neuronales

« The aim is to study neural mechanisms of learning, it is important to be clear about what exactly one means by « learning ». This is a loaded term. »

S. Marom, D. Eytan [MAROM et EYTAN, 2005]

Les mécanismes fondamentaux de la mémorisation et de l'apprentissage sont maintenant bien établis. Il s'agit pour l'essentiel de processus qui modifient l'efficacité des synapses suivant l'activité des neurones connectés. Ces mécanismes ont été largement étudiés dans toute sorte de configurations naturelles et également dans des cultures. Les changements dans les réseaux neuronaux peuvent être de nature à créer de nouvelles connections ou d'en déduire par le biais de la croissance ou la décroissance de neurites et de boutons synaptiques. En préservant les connections existantes, les changements prennent la forme de renforcement ou d'affaiblissement des synapses existantes correspondant aux phénomènes de Potentialisation à Long Terme –PLT– (« Long Term Potentiation » –LTP–) et la Dépression à Long Terme (« Long Term Depression » –LTP–) [BEAR et MALENKA, 1994].

Overview of studies on changes in network activity by electrical stimulation. Tetani are trains of stimuli. A volley is like a train in which a different electrode is stimulated with each pulse.

Study	Number of electrons	Stimulation parameters	Evaluated aspect	Effect
Maeda et al. (1998) [1]	5	Tetani: (20 × , 20 Hz; ITI 10–15 s) 5–10 ×	Burst rate and size, spontaneous and induced by test stimuli.	Increase in burst rate and spikes per burst. Both spontaneous and as induced by test stimuli.
Jimbo et al. (1998) [2]	1 or 8	Tetani: (11 × , 20 Hz; ITI 5 s) 10 ×	Cell voltage	Increase in EPSC currents.
Jimbo et al. (1999) [3]	1	Tetani: (11 × , 20 Hz; ITI 5 s) 10 ×	Number and timing of spikes induced per neuron in response to test stimuli.	Increase or decrease of EPSC's and number of induced spikes depending on pathway.
Tateno et al. (1999) [3]	1 or 2	100 μs pulses	Reliability of action potentials after test stimuli, time between test stimuli and first spike.	Shortening of latency and reduction of jitter of first spike in some neurons. Increase of effect from 1 to 2 electrodes.
Shahaf et al. (2001) [5]	2	Bipolar stimulation, ISI 1–3 s until desired response, 10 min max.	Increase in number of spikes on 1 evaluation electrode.	Increase in spikes induced in window.
Ruaro et al. (2005) [6]	15	Tetani: (100 × , 250 Hz, ITI 2 s) 40 ×	Responses to test stimuli.	Increase inset FR between 1 and 50 ms.
Wagenaar et al. (2006) [7]	2	Tetani: 20 × , 20 Hz; ITI 6 s, 150 × , shift 5ms	Responses to test stimuli, 10–50 ms. after test stimuli.	Increase in number of spikes induced by leading electrode; decrease for following electrode.
Madhavan et al. (2007) [8]	2	Tetani: 18,000 × , 20 Hz (15 min), shift 10 ms	Spontaneous expression rate of burst types (BT).	Generation of new BTs, increase or decrease in expression rate of BTs.
Chao et al. (2007) [9]	2	Tetani: 18,000 × , 20 Hz (15 min), shift 10 ms	Center of Activity Trajectory (CAT) of evoked activity.	CAT's change significantly on some of the test electrodes.
Brewer et al. (2010) [10]	30	Biphasic paired pulses, 50ms ISI, 5 s ITI, 0–3 h 30 μA, 100 μs phase	Evoked responses, spike activity.	Increase of evoked responses and in spike frequency.
Martinoia et al. (2010) [11]	8	Biphasic pulse 36 × , 0.2 Hz sequentially to each electrode, 200 μs/phase, 1.5 Vpp	Spontaneous firing rate, burst rate.	Stimulation significantly changes the mean firing rate of the networks.

TABLEAU 6.1 – Récapitulatif des études sur les changements d'activité électrique de neurones suite à des stimuli électriques. « Tetani » sont des trains de stimuli. Une volée (« volley ») est comme un train mais avec une électrode différente stimulée à chaque impulsion. Source : [FERRÁNDEZ et collab., 2013]

La plasticité dans les cultures neuronales a suscité un grand intérêt en raison de la praticité des cultures comme plateforme expérimentale, mais aussi dans la perspective technologique. C'est en effet le phénomène principal auquel on songe qui pourrait conduire à la programmation de populations de neurones. Les nombreux travaux cherchant à induire de la plasticité dans les cultures se basent sur l'hypothèse que des schémas particuliers de stimulation peuvent causer des changements de l'efficacité de la connectivité du réseau conduisant à une l'altération persistante de l'activité de la culture. La plupart des expériences dans cet objectif sont résumées dans le tableau 6.1 [FERRÁNDEZ et collab., 2013] qui fait ressortir la diversité des méthodes appliquées.

Des résultats positifs historiques ont été obtenus dès les années 1990 au moyen de matrices d'électrodes [JIMBO et collab., 1998, 1999; MAEDA et collab., 1998]. Avec des stimulations téaniques d'intensités moyennes à fortes, ces publications rapportent des modifications des probabilités de déclenchement de bouffées et de taux des bouffées spontanées faisant suite à des stimuli.

D'autre part, c'est dans des expériences sur des cultures d'hippocampe de rat que BI et POO ont confirmé en 1998 en utilisant des méthodes électrophysiologiques le phénomène de plasticité dépendante du temps (« spike timing dependent plasticity » STDP) découvert par Henri l'année précédente sur des neurones pyramidaux du neocortex [MARKRAM et collab., 1997], figure 6.19. Cette découverte a pour conséquence fondamentale d'indiquer que la célèbre règle d'apprentissage de Hebb doit être étendue et prendre aussi en compte des critères temporels [BI et POO, 2001]. La plasticité d'une synapse individuelle dépend non seulement de la répétition des activations des neurones pré- et post-synaptiques, mais également l'intervalle de temps précis entre les déclenchements respectifs [DAN et POO, 2006].

Malgré ces résultats positifs la plasticité dans les cultures de neurones reste un sujet confus lorsqu'elle est recherchée au niveau d'une population plutôt que pour des synapses individuelles. Une conclusion que l'on peut tirer avec une lecture critique de la littérature stimulée par les résultats négatifs obtenus dans le cadre de la thèse de Renaud Renault. Et à laquelle sont aussi arrivés plusieurs groupes experts de cultures neuronales et plasticité, qui va dans le sens également de résultats obtenus dans le cadre de la thèse de Renaud Renault. Ainsi dans un article de 2006 dans le Journal of Negative Results le groupe de Steve Potter a rapporté [WAGENAAR et collab., 2006c] ne

pas être parvenu à reproduire les résultats positifs des principaux protocoles publiés auparavant et conclue :

« At least, it appears that the conditions in which plasticity can be induced in dissociated cortical cultures using extra-cellular electrical stimuli are subtle and not very well understood. »

D'autres répétitions des protocoles historiques de [JIMBO et collab. \[1998, 1999\]](#) ainsi que de [SHAHAF et MAROM \[2001b\]](#) ont été effectuées par l'équipe de W.L.C. Rutten avec également des résultats peu concluants [[VAN STAVEREN et collab., 2005](#)]. Des séries de stimulations peuvent en effet amener à un changement significatif de la réponse immédiate du réseau, mais sans persistance dans le temps de l'entraînement. Des travaux ultérieurs du même groupe rapportent cependant d'autres résultats positifs obtenus avec le protocole de [SHAHAF et MAROM \[2001b\]](#), mais seulement dans 5 expériences sur 16 réalisées. Ce protocole qui laisse entrevoir le fonctionnement de la plasticité dans les cultures au-delà de la synapse unique mets en œuvre des électrodes et est focalisé sur la stimulation particulière d'un neurone spécifique. Au niveau d'action sur une population entière de neurones la situation semble être encore plus confuse. Dans une revue de [2011](#), [EYTAN](#) du groupe de Shimon Maron fait le commentaire suivant qui résume très bien à mon sens cette situation complexe :

« What these studies mainly show is that such changes can indeed be achieved, but there are no simple “plasticity rules” at the network level, such as those discovered for single synapse in the sense of long-term potentiation (LTP), long-term depression (LTD), or spike-timing-dependent plasticity (STDP) [...] It seems that stimulation drives changes in connectivity, but the direction and amplitude of change is not easily predicted and varies between different protocols and laboratories. »

Les résultats rapportés par [FERRÁNDEZ et collab. \[2013\]](#) dans autre tentative de plasticité en culture visant d'imposer un schéma d'apprentissage « Hebbien » par des répétitions de stimuli sur des paires d'électrodes sont également contrastés. Des modifications persistances de corrélations entre électrodes s'établissent dans certains cas avec un contrôle qui apparaît faible et peu reproductible.

Cette difficulté pour caractériser la plasticité dans les cultures a été également constatée durant la thèse de Renaud Renaud ([[RENAULT, 2015](#)] chapitre 5). L'objectif de ce chapitre était d'étudier comment la plasticité dépendante du temps (STDP) caractérisée au niveau de neurones individuels, se répercuterait dans une plasticité au niveau de l'activité de cultures. Un processus que nous avons nommé « Burst Time Dependent Plasticity » pour le distinguer de la plasticité proprement synaptique. Malheureusement aucun des efforts réalisés avec une multiplicité de protocoles rapportés dans le chapitre de la thèse, n'a permis de caractériser un effet persistant que ce soit de dépression ou de potentialisation.

L'apprentissage dans les cultures est clairement un processus délicat. De nombreux auteurs ont proposé des méthodes différentes pour tenter de produire de la plasticité avec des résultats contradictoires en raison de la diversité des approches, de l'absence de critères clairs pour valider la réalisation d'un apprentissage, et des difficultés de reproductibilité des expériences.

Les raisons de ces échecs à reproduire ou obtenir des manifestations reproductible de plasticité dans les cultures neuronales peuvent être multiples. Les préparations en laboratoire semblent très sensibles aux détails des protocoles et de la manière de les mettre en œuvre qui conduisent les cellules à réagir différemment. Une autre explication pourrait tenir à la nature désordonnée des réseaux en culture. Leur forte interconnexion mais sans la structuration modulaire des réseaux *in vivo* pourrait aussi perturber suffisamment et de manière contradictoire les modifications des poids synaptiques annihilant la plasticité au niveau des populations. Enfin, étant donné l'immensité de l'espace des phases des paramètres et des conditions possible des protocoles d'apprentissage on peut aussi concevoir que les conditions d'opération pour des manifestations significatives n'ont pas encore été établies. Des schémas d'apprentissages alternatifs peuvent également

offrir de nouvelles voies d'exploration [SINAPAYEN et collab., 2017] à tester en laboratoire. Les expériences avec les cultures sont notoirement délicates et lentes ce qui réduit naturellement les capacités d'exploration expérimentale de ces effets.

L'utilisation d'une culture neuronale apprenant à interagir dans un environnement physique pour contrôler les mouvements d'un robot a suscité l'attention de plusieurs groupes souhaitant étendre le champ de l'intelligence artificielle dans le domaine biologique. Quelques publications [DEMARSE et collab., 2001; LI et collab., 2016; NOVELLINO et collab., 2007; SHULTZ et collab., 2017; TESSADORI et collab., 2012] rapportent la mise en œuvre de tels dispositifs, donnant lieu également à des communications publiques médiatisées. À la lumière de la discussion précédente ces résultats interpellent. L'apprentissage de comportements semble effectivement obtenu avec certains choix de codage et décodage des signaux envoyé et déchiffrés à la puce neuronale au cœur du dispositif [LI et collab., 2016].

Les systèmes robotiques employés font l'usage de nombreux composants artificiels dans la boucle entre la lecture de l'activité de la cultures par les électrodes, le contrôle des mouvements, et la perception de signaux extérieurs pour envoyer des entrées aux neurones via les électrodes. Il n'est pas du tout clair qu'elle est la nature du traitement des informations et du contrôle qui se passe vraiment dans les neurones biologiques. Une étude approfondie des différentes publications et des détails de leurs méthodes serait nécessaire pour établir le comportement effectif des cultures dans les dispositifs proposés et les propriétés de l'apprentissage obtenu.

En conclusion la question de la plasticité dans les cultures paraît encore grande ouverte. La littérature est en tout cas encore trop contradictoire pour définir des conditions opérationnelles pour l'exploitation de la plasticité synaptique dans des dispositifs neuronaux artificiels. Des explorations plus systématiques sont nécessaires en perspective des recherches et des applications très convoitées d'apprentissage dans les cultures.

La capacité calculatoire ne semble pas être une propriété spontanée des cultures avec des neurones largement désordonnés. Leur activité est dominée par des bouffées de déclenchements à grandes échelles apparemment peu informatives. Quelles sont alors les conditions nécessaires pour amener une population de neurones à posséder une capacité calculatoire ?

Cultures structurées pour organiser l'activité.

Grâce aux possibilités de conception des géométries particulières qui rompent l'homogénéité des cultures, on peut explorer les conditions permettant aux cultures d'avoir une capacité calculatoire. Comme on le développera plus loin, dans les cultures de neurones homogènes, où les structures neuronales et les connexions se construisent aléatoirement, aucun comportement de nature calculatoire ne semble émerger des activités spontanées qui y sont observées. C'est avec une structuration, l'introduction d'hétérogénéité, que certaines explorations avec cultures neuronales comme celles de l'équipe de Elisha Moses discutées à la section 6.4.4 voient apparaître des comportements s'apparentant à des propriétés calculatoires. Cette organisation d'ensemble du réseau pour permettre le calcul n'est pas loin de la situation universelle des systèmes neuronaux naturels, qui loin d'une masse indifférenciée de neurones aléatoirement interconnectés, sont hautement organisés dans des structures sophistiquées.

La possibilité de fabriquer des milieux de cultures avec des formes à souhait permet à ce titre d'explorer les relations entre structure des populations neuronales et leurs propriétés. C'est là une approche synthétique, inspirée de l'ingénierie, pour investiguer les principes de calcul neuronal en cherchant à construire les structures neuronales capables de le réaliser. L'approche est différente et complémentaire de la dissection plus usuelle de systèmes neuronaux naturels en neurosciences.

L'évolution naturelle a conduit à la formation de circuits neuronaux dans les organismes vivants capables de réalisation d'opérations de calcul complexes pour traiter les informations né-

cessaires à la survie et liées à la fois aux interactions avec l'environnement qu'à la physiologie interne de l'animal. Avec les connectivités et les types de neurones appropriés ces circuits ont la propriété de produire rapidement des réponses sophistiquées avec fiabilité et grande robustesse. En dépit de toutes les variations physiques et chimiques auxquels est soumis le vivant, les circuits de neurones mêmes des organismes les plus simples qui en sont dotés accomplissent les opérations pour lesquels ils sont dédiés avec un taux d'erreur faible.

Cette fiabilité est étonnante étant donné ce qui est connu de la variabilité des circuits neuronaux naturels. Au niveau individuel, pour des conditions déterminées le fonctionnement d'un neurone est très reproductible. La source de variation principale dans les systèmes nerveux semble se situer au niveau des connections synaptiques. Les plans de répartition des neurones et de projection de leurs arborescences synaptiques semblent très reproductibles. Le niveau des connections synaptiques inter-neurones est sujet à plus de variabilité. Il peut s'agir de variabilité liée à l'établissement des synapses mais aussi des fluctuations induites par la plasticité de l'intensité synaptique.

Une explication proposée pour la fiabilité des circuits est la redondance qu'on y retrouve la plupart du temps. Sauf dans des organismes très élémentaires avec un nombre de neurones bien précis et réduit comme *C. Elegans*, ce sont toujours des grands nombres de neurones qui concourent à un circuit. La mise en œuvre de nombreuses unités en parallèle semble être une solution pour aboutir à une réponse d'ensemble fiable indépendamment des liens synaptiques incertains.

6.7 Modélisation des cultures neuronales

6.7.1 Modélisation de la structure des cultures

La morphologie des neurones est une des caractéristiques principales de leur nature. Les propriétés fonctionnelles d'une population neuronale, ses capacités de traitement d'information, de production d'une réponse en fonction d'une entrée dépendent fondamentalement de la forme des neurites de ses neurones et des recouvrements qui s'y retrouvent entre dendrites et axones permettant l'établissement des synapses. Dans le cerveau les morphologies des neurones sont très diverses. Schémas d'arborisation, longueurs de branches, présence de bourgeons synaptiques, sont autant de facteurs morphologiques qui varient fortement entre les différentes classes de neurones. Des différences morphologiques connectées à des différences fonctionnelles dans des relations encore mal caractérisées. Un florilège de types neuronaux différents est connu depuis longtemps. Les facteurs conduisant à la différenciation de neurones vers les différentes variantes sont mal connus [CUNTZ et collab., 2014; STIEFEL et SEJNOWSKI, 2007]. Les formes achevées et les propriétés biologiques doivent être le résultat d'un processus de morphogénèse de tissus complexe, combinant des mécanismes de régulation génétiques internes mais sensibles à des signaux de l'environnement physique et chimique où se développent les cellules en interactions avec d'autres.

Malgré cette place centrale de la forme, de la géométrie et du développement la majorité des efforts des simulations de systèmes neuronaux sont essentiellement focalisés sur l'activité électrique de collectifs de neurones. Quand elle est prise en compte la morphologie neuronale elle est une donnée en entrée paramétrant la simulation [MARKRAM et collab., 2015]. Cette utilisation de morphologies réelles de neurones pour des simulations détaillées de modèles électrophysiologiques est ancienne. Les simulations utilisent comme support des informations anatomiques détaillées issues de reconstructions numériques à partir d'images par microscopie. Des outils bien établis comme GENESIS [BOWER et BEEMAN, 2007] et NEURON [HINES et CARNEVALE, 2001] et peuvent être utilisés pour intégrer l'activité de neurones avec des morphologies réalistes et des localisations précises de connexions.

La caractérisation quantitative de la morphologie de neurones fait l'objet d'études depuis longtemps mais reste encore partielle et soumise à de nombreuses contraintes pratiques. La préparation des tissus, l'identification des fragments de neurites dans les images rendent la production

de ces données de structure très laborieuse. Des logiciels pour venir en aide à la reconstitution manuelle et différentes initiatives y compris de science ouverte ont vu le jour dans le but de faciliter ces démarches pour lesquelles on peut encore se référer à la revue de 2011 par Duncan et Ascoli [DONOHUE et ASCOLI, 2011]. De tels efforts ont conduit à la constitution de bases de données de structures neuronales et de caractéristiques de neurites nécessitant la conceptualisation de modèles de représentation des structures et de données. Le livre édité par G. Ascoli en 2002 « Computational Neuroanatomy » [ASCOLI, 2002] donne encore une des visions plus larges sur les différentes thématiques associées. Le champ de la reconstruction des formes neuronales a conduit notamment la base de données NeuroMorpho.org [GEORGE MASON] est la plus grande collection publique de reconstructions de formes 3D de neurones contribué es par des centaines de laboratoires du monde entier. L'acquisition de données morphologiques de neurones a connu un regain d'activité portée par les méga projets de neuroscience qui ont vu le jour ces dernières années avec de nouvelles initiatives d'envergure comme BigNeuron [PENG et collab., 2015] qui combine de nouvelles techniques d'imagerie et des avancées en marquage et microscopie. Cependant la connaissance des formes neuronale dont on vient de discuter reste descriptive. C'est une caractérisation, mesure et catégorisation de structure déduites à partir d'observations qui pourra servir de support pour des études fonctionnelles. À ma connaissance ces études structurales portent toutes sur des neurones dans des contextes *in vivo*. Probablement plusieurs facteurs expliquent l'absence de reconstruction de morphologies neuronales en culture. D'une part l'entrelacement des neurones dans une population en culture rend impossible de suivre les trajets des neurites d'un neurone autrement que sur quelques segments. D'autre part, le substrat des cultures et leur nature planaire fine ne rend pas possible la fabrication la segmentation d'un volume contenant les neurones à reconstruire en coupes microtomiques entre lesquelles on pourrait suivre la continuité cellulaire des neurites.

La morphologie des neurones dans des cultures reste donc très mal connue et la structure des réseaux établie. La connaissance de ces réseaux serait cependant précieuse pour les applications biotechnologiques des cultures. Des modèles et simulations de croissance de neurones et de réseaux seraient utiles pour guider la conception des dispositifs et l'exploration d'hypothèses sur les mécanismes en jeu. Positionner ces études de morphogénèse neuronale dans le contexte particulier des cultures neuronales offre aussi l'avantage considérable, contrairement aux contextes *in vivo*, de pouvoir effectivement observer expérimentalement la croissance du réseau neuronal.

6.7.2 Modélisation de la croissance neuronale : état de l'art

Comprendre et modéliser le développement neuronal et l'établissement des réseaux est clairement un défi considérable tant les processus impliqués sont nombreux et variés. La modélisation théorique de la génération des formes neuronales n'est pas une question vraiment nouvelle. Ce sujet a connu une certaine vague durant les années 1990 qui se manifeste dans deux ouvrages collectifs parus au début des années 2000 [ASCOLI, 2002; VAN OUYEN, 2003] aussi discutées dans la revue de 2006 par Gramah et van Ooyen [GRAHAM et VAN OUYEN, 2006]. J'aborderai plus bas les différents points sur lesquels des modélisations ont été entreprises, elles portent souvent sur des mécanismes particuliers impliqués dans le développement des neurones sans un cadre mathématique et numérique intégré. Durant la dernière décennie la thématique a été moins développées, et a surtout fait l'objet d'applications' informatiques. De tels modèles ont été utilisés dans le cadre d'études théoriques du comportement de cultures neuronales afin de générer les réseaux spatiaux servant de support à des modèles d'activité [GRITSUN et collab., 2012; ORLANDI et collab., 2013].

La morphogénèse des neurones et la croissance des réseaux est le résultat de l'action combinée de nombreux processus moléculaires, génétiques, chimiques et biophysiques au niveau de la cellule et de son environnement. Un modèle mathématique unique uniforme n'est pas vraisemblable pour capturer cette complexité. Comprendre et modéliser le développement neuronal et l'établissement des réseaux est clairement un défi considérable tant les processus impliqués sont

nombreux et variés. La modélisation théorique de la génération des formes neuronales n'est pas une question vraiment nouvelle. Ce sujet a connu une certaine vague durant les années 1990 qui se manifeste dans deux ouvrages collectifs parus au début des années 2000 [ASCOLI, 2002; VAN OOYEN, 2003] aussi discutées dans la revue de 2006 par Gramah et van Ooyen [GRAHAM et VAN OOYEN, 2006]. Malgré cet historique, la compréhension des mécanismes de morphogénèse des neurones et des réseaux est encore partielle. Une raison qui l'explique est la difficulté de l'étude expérimentale de ces processus nécessaire à la théorisation. Le cours du développement de cerveau est en effet difficile à observer *in vivo* en temps réel. Et ce sont justement les nouvelles possibilités d'expérimentation avec des cultures en microfluidique qui renouvellent aussi les possibilités théoriques.

Durant la dernière décennie la thématique a été moins développée, mais les grandes ambitions de simulation globale de cerveau des méga projets de neurosciences Human Brain Project en Europe et BRAIN aux USA nécessitent la génération de réseaux neuronaux réalistes de grandes taille appellent à de nouveaux développements dans cette direction [SHILLCOCK et collab., 2016]. Des recherches se sont cependant toujours poursuivies notamment dans des cercles autour de chercheurs impliqués de longue date sur la question, comme à l'université d'Amsterdam et autour de Arjen van Ooyen. Cet auteur a notamment publié en 2011 une revue encore très actuelle des modèles théoriques pour l'analyse du développement neuronal à toutes ses étapes depuis la structuration embryonnaire précoce des tissus cérébraux à la maturation dépendante de l'activité des réseaux [VAN OYEN et OYEN, 2011]. Une autre discussion critique un peu plus récente sur l'état de ces approches est l'article de Torben-Nielsen et Erick De Schutter de 2014 [TORBEN-NIELSEN et DE SCHUTTER, 2014] qui développe plusieurs idées proches des développements en cours dans notre équipe.

Ces auteurs mettent notamment en avant un grand nombre de modèles sont basés sur des processus stochastiques d'élongation de neurites et des processus de branchement sans modélisation des mécanismes biologiques de croissance subjacents [ASCOLI et collab., 2001; VAN PELT et UYLINGS, 2002]. Plusieurs programmes de génération de formes neuronales sont ainsi disponibles L-Neuron [ASCOLI], Neugen [EBERHARD et collab., 2006] et d'autres [CUNTZ et collab., 2014; LÓPEZ-CRUZ et collab., 2011; STIEFEL et SEJNOWSKI, 2007] listés avec plus de références dans [HERMAN CUNTZ, MICHIEL W.H.REMME, 2009]. Ces approches se basent sur des données de variables morphologiques relevées dans les structures neuronales reconstruites à partir d'observation (par exemple de distributions des longueurs de segments, nombres et angles des branchements...) pour paramétrer des algorithmes stochastiques de génération de structures géométriques. Ces méthodes s'inspirent généralement des systèmes de Lyndenmayer communs pour la description de processus de développement et de croissance dans le vivant avec des processus stochastiques de branchement de type Galton–Watson. Chacune a ses avantages spécifiques et se distinguent par des hypothèses et mécanismes pour tenir compte d'effets différents tels la diffusion et de compétition pour des ressources, des principes auto-référentiel d'évitement des branches, de tropisme par rapport au soma, de rigidité membranaire des contraintes d'occupation d'espace (voir [MEMELLI et collab., 2013]). Il s'agit pour l'essentiel d'algorithmes, capables de générer efficacement des structures géométriques avec des caractéristiques morphologiques réalistes de différentes classes de neurones, mais sans correspondre au résultat d'une simulation des mécanismes biologiques de croissance opérant dans la réalité.

Cependant, si ces algorithmes permettent la génération de morphologies neuronales réalistes, ils restent la plupart au niveau de la formation d'un neurone. Quelques travaux seulement ont explicitement pris en compte la formation d'un réseau [EVANS et POLAVARAM, 2013], les programmes NetMorph [KOENE et collab., 2009], CX3D [ZUBLER, 2009] et NeuroMac [TORBEN-NIELSEN et DE SCHUTTER, 2014] qui seront discutés plus bas.

Outre ces approches visant une modélisation réaliste de morphologies neuronales on trouve aussi dans la littérature différents algorithmes de génération de réseaux neuronaux spatiaux appliqués dans le cadre d'études théoriques du comportement population des cultures, afin de générer

les réseaux spatiaux servant de support à des modèles d'activité. Il s'agit d'un pendant spatial de modèles de graphes aléatoires de théorie des graphes utiles pour des études théoriques. Des exemples non limitatifs peuvent être retrouvés dans les références ici [LENK et collab., 2016; MAHESWARANATHAN et collab., 2012; ORLANDI et collab., 2013; ZBINDEN, 2011]. Je discuterai brièvement des modélisation et approches de la littérature se rapprochant le plus du projet suivi dans l'équipe.

Netmorph L'outil NetMorph [KOENE et collab., 2009] développé dans l'équipe de van Ooyen simule la croissance en parallèle de nombreux neurones et la création de synapses. La topologie des neurones est générée dans ce modèle par des règles de croissance mécaniques correspondant à une modélisation phénoménologique pour l'élongation et les branchements paramétrés par des données d'observations. Une approche intermédiaire entre l'algorithme purement mathématique de génération de forme et une modélisation des mécanismes des processus biologiques. Cependant dans NETMORPH les neurones sont indépendants les uns des autres et leur morphogénèse s'effectue sans interactions dans l'environnement et ne correspond donc pas à une modélisation de croissance de réseau. Une évolution de Netmorph tenant compte le guidage chimique a été implémentée dans [GRITSUN et collab., 2012], où les dynamiques d'activités sur des réseaux générées avec ou sans signaux chimiques son comparés.

CX3D CX3D [ZUBLER, 2009; ZUBLER et collab., 2013] est un autre outil informatique permettant de traiter ed parallèle de nombreux neurones qui a été conçu pour la simulation du développement du cortex dans son ensemble, incluant division cellulaire, migration de cellules et croissance des neurites. Les neurones simulés sont implantés physiquement dans l'espace, peuvent sécréter et réagir à des composés chimiques avec des règles stochastiques d'élongation et de branchement configurables par l'utilisateur. Si CX3D intègre bien des interactions, son implémentation informatique est cependant sérielle et ne permet pas la simulation en parallèle d'un grand nombre de neurones (une version parallèle apparait mentionnée en ligne). Le système implémente modèle d'élongation et de branchement qualitatif peu détaillé dans les publications mais qui implémente un guidage chimique et de contraintes physiques. La conception du programme permet cependant l'implémentation de modèles différents.

Simulateur de morphologies neuronale d'IBM La publication de 2013 par Memmelli *et al.* [MEMELLI et collab., 2013] présente un modèle phénoménologique et un algorithme de génération de formes dendritiques implémenté pour des simulations sur super ordinateurs. Le modèle comprend le branchement des arbres suivant un processus de Galton-Watson, et des trajectoires des neurites déterminées par des forces auto-référentielles qui influences la direction de la croissance aléatoire dans un espace limité. Les trois forces empiriques qui déterminent l'orientation de la croissance sont une inertie venant de la rigidité du cytosquelette, une force tropique orienté par rapport au soma, et une force d'auto-évitement entre les branches. L'algorithme est capable de générer des morphologies reproduisant plusieurs exemples naturels. Cependant il dépend essentiellement de 4 paramètres difficiles à relier a priori à des observables biologiques, dont le traitement analytique est délicat ainsi que l'ajustement. Ce modèle n'inclut pas des interactions avec d'autres neurones et des obstacles, mais un des auteurs, B. Torben-Nielsen est également auteur du système NeuroMac présenté ci-dessous. Enfin, le code de ce simulateur n'est pas disponible, quoique des possibilités de collaborations à son sujet sont mentionnées dans l'article.

NeuroMac Les deux programmes NETMORPH et CX3D n'ont plus été maintenus depuis les publications correspondantes et à ce jour (janvier 2018) les sites web correspondant ne sont plus accessibles. Ils avaient cependant fait l'objet d'une étude comparative [AĆIMOVIĆ et collab., 2011] (avec des applications dans le contexte de l'étude de la croissance de cultures corticales). Étant

donné que le second ne repose pas sur une modélisation particulière et l'étude a repris le modèle de NETMORPH cette comparaison est celle de programmes de simulation. Même si les programmes ne sont plus disponibles cette analyse permet de conclure sur la nécessité de combiner deux facteurs pour des simulations efficaces. La nature des modèles implémentés est naturellement centrale et notamment pour pouvoir effectivement simuler le développement de réseaux dans leur environnement les interactions doivent être prises en compte. D'autre par la mise en œuvre informatique doit suivre une implémentation permettant le calcul haute performance et une parallélisation pour pouvoir gérer de nombreux neurones simultanément avec des temps d'exécutions possibles.

Le développement de cet environnement était motivé par les préoccupations mentionnées d'avoir un système de simulation parallèle efficace permettant de gestion de la croissance de nombreux neurones en parallèles en interactions mutuelles et avec l'environnement basés sur des modèles de mécanismes biologiques plutôt que de principes algorithmiques [TORBEN-NIELSEN et DE SCHUTTER, 2014]. Le modèle de croissance est centré sur les cônes de croissance représentés par des agents informatiques qui suivent des règles de mise à jour inspirées de la vision biologique de ces structures et de leur dynamique avec des filopodes qui explorent de l'environnement physique et chimique pour déterminer une direction d'élongation ou la possibilité d'un branchement. Comme l'illustre le cas de CX3D une grande attention doit être portée à la programmation en lien avec la construction du modèle lui-même. En particulier pour gérer les interactions dans un environnement fait d'obstacles en évolution (les neurones) l'exécution en parallèle exige de gérer très efficacement les échanges de messages entre les cœurs de calcul gérant les différents cônes de croissance. La seule publication sur ce système de 2014 [TORBEN-NIELSEN et DE SCHUTTER, 2014] est une version préliminaire d'étude pour valider la pertinence de l'approche. Ce système est très proche de la modélisation que nous souhaiterions pouvoir mettre en œuvre pour l'étude des cultures. Nous avons contacté les auteurs pour envisager d'utiliser le code ou en vue d'une collaboration. Malheureusement le projet s'est arrêté en 2014 et le code écrit en Python est uniquement à l'état de prototype de principe [TORBEN-NIELSEN] et devrait être complètement repris pour du calcul à haute performance.

6.7.3 Modélisation de l'activité des neurones

La recherche de modèles minimaux de dynamique a été particulièrement active en neurobiologie théorique [GHIGLIAZZA et HOLMES, 2004]. Le modèle Hodgkin–Huxley a fourni l'explication toujours actuelle de l'activité électrique des neurones éclairant les contributions de canaux ioniques multiples pour la dynamique du potentiel membranaire. De nombreuses extensions et spécialisation du modèle ont été introduites depuis son introduction. Notamment l'application des équations du câble électrique par Wilfrid All dès 1962 a permis la description de la propagation des potentiels d'actions le long des neurites. Au prix de la description de la structure d'un neurone en de multiples compartiments reproduisant sa structure morphologique.

Pour l'activité d'un ensemble de neurones en interactions on peut cependant supposer que c'est le comportement global des cellules qui déterminent le comportement collectif de la population. La modélisation des neurones est simplifiée en le considérant comme un compartiment unique, dont la décharge communique un signal aux neurones connectés.

Les modèles « à la Hodgkin Huxley » représentent un compartiment unique. Ils peuvent être ajustés avec grande précision, quoique laborieusement, à des mesures précises des potentiels membranaires enregistrés par électrophysiologie. Il s'agit cependant de modèles très complexes avec des équations additionnelles pour chaque nouveau canal que l'on voudrait inclure et plusieurs paramètres pour décrire aussi bien l'activation et que l'arrêt de chaque canal. Ces modèles permettent d'être très précis pour un type de neurone donné, mais leur complexité cache les raisons du comportement temporel particulier des potentiels. Comment l'action de tous les canaux ioniques contribue aux variations d'activité forcements non linéaires observées? Quels sont les

effets des courants ioniques sur l'existence, la stabilité d'équilibres ou de cycles périodiques d'activité. Quels sont les régimes dynamiques et les transitions possibles?

Les modèles de type Hodgkin-Huxley ont trop de termes, et trop complexes, pour permettre de systématiser de telles analyses y compris numériquement. Son espace des phases est en effet à 4 dimensions et le modèle original possède de nombreux paramètres. Pour répondre aux questions précédentes il est indispensable de disposer de modèles simplifiés qui reproduisent suffisamment la complexité des comportements observés avec moins de variables et de relations pour en permettre l'étude au moins partiellement analytique ou numérique mais extensivement. Les modèles réduits de neurones sont essentiels pour la simulation de réseaux neuronaux de grande taille, inaccessible avec des modèles à la Hodgkin-Huxley.

Il existe une grande variété de modèles mathématiques utilisés pour décrire l'activité d'un neurone à un compartiment. Leur construction est généralement fondée sur la reproduction de caractéristiques dynamiques du modèle à un compartiment de Hodgkin & Huxley. Des approches mathématiques différentes ont aussi été utilisées comme des automates cellulaires [VIANA et collab., 2014] dont le modèle de Courbage & Nekorkin [COURBAGE et collab., 2007] et le mapping à deux dimensions de Rulkov [RULKOV, 2001]. Et surtout des modèles d'équations différentielles avec un nombre de variables réduit dont les plus connus sont Hindmarsh & Rose [HINDMARSH et ROSE, 1984] et FitzHugh-Nagumo [FITZHUGH, 1961; NAGUMO et collab., 1962] dont les motivations principales sont d'isoler les propriétés mathématiques essentielles de l'excitation membranaire à partir des propriétés des flux d'ions calcium et potassium. Le modèle FitzHugh-Nagumo introduit une représentation qui se révélera très efficace avec deux variables, l'une pour décrire le voltage de la membrane et dotée d'une activité auto-référentielle positive qui amène à des décharges, l'autre une variable de variation lente qui capture de manière efficace la dynamique lente des canaux ioniques. Un autre modèle très connu, le plus simple et introduit même bien avant les travaux de Hodgkin & Huxley est le modèle « integrate & fire » [LAPICQUE, 1907] basé sur une équation différentielle linéaire, permettant des applications faciles utiles dans certains contextes en dépit du manque de réalisme. L'idée d'une dynamique à seuil « integrate & fire » est à la base des travaux sur la percolation dans les populations neuronales présentés dans le prochain chapitre.

Une version plus proche de comportements observés, est le modèle adaptatif exponentiel integrate and fire AEIF avec lequel nous avons effectué la plupart de nos travaux de simulation dynamique présentés dans le prochain chapitre. Ce modèle introduit par Brette and Gerstner en 2005 [BRETTE et GERSTNER, 2005] utilise comme celui de FitzHugh-Nagumo deux variables, une variable de potentiel avec un terme exponentiel pour la décharge et une variable capurant les effets d'adaptation de courant ioniques lents. Ce modèle est capable de reproduire remarquablement la diversité des comportements temporels et des dynamiques de décharges observés sur des neurones réels [CLOPATH et collab., 2007; NAUD et collab., 2008]. Touboul and Brette [FEIGE et collab., 2011] ont étudié en détail le diagramme de phase de ce modèle caractérisant l'existence de bifurcations Andronov-Hopf et saddle-node qui permettent de comprendre des transitions entre comportements observés expérimentalement.

6.8 Conclusions et perspectives : vers la neurobiologie synthétique.

Dans ce chapitre j'ai introduit plusieurs aspects concernant les cultures primaires de neurones dissociées qui constituent un objet d'étude particulier dans le champ des neurosciences et encore plus pour la biophysique. Ces dispositifs sont utilisés dans de nombreuses applications diverses. L'accessibilité des neurones en culture à des traitements en font en premier lieu une plate-forme expérimentale très utilisée pour des recherches appliquées sur des troubles neurologiques et la sensibilité des cellules nerveuses à toute sorte de drogues et produits chimiques. La meilleure compréhension de l'objet « culture neuronale » lui-même est dès lors intéressante afin d'améliorer leurs possibilités d'utilisation et peut suffire pour motiver des recherches appliquées.

Pour le biophysicien, ces objets possèdent cependant d'autres attraits fascinants. Les cultures neuronales permettent d'observer la croissance d'un tissu excitable exhibant une grande richesse de comportements difficiles à résumer, et formés des mêmes éléments conduisant, dans les cerveaux naturels aux processus cognitifs supérieurs. Dans cet esprit les cultures sont un objet d'étude des phénomènes dynamiques et d'auto-organisation qui conduisent potentiellement aux processus cognitifs. En mettant en place des dispositifs pour manipuler et contrôler les cultures on peut espérer tenter de découvrir les conditions nécessaires aux fonctions calculatoires nécessaires au traitement de l'information. L'ambition scientifique de la biologie synthétique discutée dans la partie II de ce mémoire est de comprendre les systèmes génétiques et biochimiques et apprenant à les fabriquer. La création de processeurs de calcul à base de neurones vivants peut alors se concevoir comme l'activité d'une « neurobiologie synthétique » visant à mieux comprendre la cognition via une démarche originale d'ingénierie.

Les dynamiques collectives ayant lieu dans les cultures neuronales, d'une part durant la structuration du tissu cellulaire lors de la croissance de celles-ci, d'autre part concernant l'activité électrique qui s'y déroulent nécessitent des études relevant de la physique statistique et des systèmes dynamiques. Les contributions dans ce sens auxquelles j'ai participé font l'objet du chapitre suivant.

6.9 Références

- ABELES, M., H. BERGMAN, E. MARGALIT et E. VAADIA. 1993, «Spatiotemporal firing patterns in the frontal cortex of behaving monkeys», *Journal of Neurophysiology*, vol. 70, n° 4, doi :10.1152/jn.1993.70.4.1629, p. 1629–1638, ISSN 0022-3077. [256](#)
- AĆIMOVIĆ, J., T. MÄKI-MARTTUNEN, R. HAVELA, H. TEPPOLA et M.-L. LINNE. 2011, «Modeling of Neuronal Growth In Vitro : Comparison of Simulation Tools NETMORPH and CX3D.», *EURASIP journal on bioinformatics & systems biology*, vol. 2011, doi :10.1155/2011/616382, p. 616 382, ISSN 1687-4153. [234](#), [265](#)
- ACIMOVIC, J., T. MAKI-MARTTUNEN et M. L. LINNE. 2015, «The effects of neuron morphology on graph theoretic measures of network connectivity : the analysis of a two-level statistical model», *Front Neuroanat*, vol. 9, n° June, doi :10.3389/fnana.2015.00076, p. 76, ISSN 1662-5129. [238](#)
- AERTSEN, A., M. DIEMANN et M. O. GEWALTIG. 1996, «Propagation of synchronous spiking activity in feedforward neural networks.», *Journal of physiology, Paris*, vol. 90, n° 3-4, p. 243–7, ISSN 0928-4257. [243](#), [256](#)
- ALBUQUERQUE, C., D. J. JOSEPH, P. CHOUDHURY et A. B. MACDERMOTT. 2009, «Dissection, plating, and maintenance of cortical astrocyte cultures», *Cold Spring Harbor Protocols*, vol. 4, n° 8, doi : 10.1101/pdb.prot5273, p. pdb.prot5273, ISSN 15596095. [223](#)
- ALCANTARA, S. L., P. GARAY, L. MCGILLICUDDY et collaborateurs. «Kinetic and label-free , live content imaging assays for neurite outgrowth in primary , iPSC-derived and immortalised neurons», . [221](#)
- AMERICAN FRIENDS OF TEL AVIV UNIVERSITY. 2011, «Understanding the human neurosystem by researching locust brains», URL <https://www.sciencedaily.com/releases/2010/11/101129131812.htm>. [224](#)
- ARISTOTELES. «De la mémoire et de la réminiscence; Traduction de Jules Barthélemy-Saint-Hilaire, 1866», URL <http://www.documentacatholicaomnia.eu/03d/-384{-}322, {-}Aristoteles, {-}Memoire{-}et{-}reminiscence, {-}FR.pdf>. [255](#)
- ASCOLI, G. «L-Neuron Download Page», URL <http://krasnow1.gmu.edu/cn3/L-Neuron/index.htm>. [264](#)

- ASCOLI, G. A. 2002, «Computational Neuroanatomy : Principles and Methods», doi :10.1385/1592592759, p. 469, ISSN 1941-6016. [263](#), [264](#)
- ASCOLI, G. A., J. L. KRICHMAR, R. SCORCIONI et collaborateurs. 2001, «Computer generation and quantitative morphometric analysis of virtual neurons», *Anatomy and Embryology*, vol. 204, n° 4, doi :10.1007/s004290100201, p. 283–301, ISSN 03402061. [264](#)
- AVIEL, Y., C. MEHRING, M. ABELES et D. HORN. 2003, «On embedding synfire chains in a balanced network.», *Neural computation*, vol. 15, n° 6, doi :10.1162/089976603321780290, p. 1321–1340, ISSN 0899-7667. [254](#)
- BAKER, B. J., E. K. KOSMIDIS, D. VUCINIC et collaborateurs. 2005, «Imaging brain activity with voltage- and calcium-sensitive dyes», *Cell Mol Neurobiol*, vol. 25, n° 2, doi :10.1007/s10571-005-3059-6, p. 245–282, ISSN 0272-4340. [229](#), [230](#)
- BANERJEE, A. et T. J. ELLENDER. 2009, «Oscillations in the Developing Cortex : A Mechanism for Establishing and Synchronizing an Early Network?», *Journal of Neuroscience*, vol. 29, n° 48, doi : 10.1523/JNEUROSCI.4567-09.2009, p. 15 029–15 030, ISSN 0270-6474. [233](#)
- BARLOW, H. B. 1961, «Possible Principles Underlying the Transformations of Sensory Messages», dans *Sensory Communication*, The MIT Press, ISBN 9780262518420, p. 216–234, doi :10.7551/mitpress/9780262518420.003.0013. [243](#)
- BEAR, M. F. et R. C. MALENKA. 1994, «Synaptic plasticity : LTP and LTD», *Current Opinion in Neurobiology*, vol. 4, n° 3, doi :10.1016/0959-4388(94)90101-5, p. 389–399, ISSN 09594388. [258](#)
- BÉDARD, C. et A. DESTEXHE. 2009, «Macroscopic models of local field potentials and the apparent 1/f noise in brain activity», *Biophysical Journal*, vol. 96, n° 7, doi :10.1016/j.bpj.2008.12.3951, p. 2589–2603, ISSN 15420086. [254](#)
- BÉDARD, C., H. KRÖGER et A. DESTEXHE. 2006, «Does the 1/f frequency scaling of brain signals reflect self-organized critical states?», *Physical Review Letters*, vol. 97, n° 11, doi :10.1103/PhysRevLett.97.118102, p. 118 102, ISSN 00319007. [254](#)
- BEGGS, J. M. et D. PLENZ. 2003, «Neuronal Avalanches in Neocortical Circuits», *The Journal of Neuroscience*, vol. 23, n° 35, doi :23/35/11167[pii], p. 11 167–77, ISSN 0270-6474, 1529-2401. [239](#), [251](#), [253](#)
- BEGGS, J. M. et D. PLENZ. 2004, «Neuronal avalanches are diverse and precise activity patterns that are stable for many hours in cortical slice cultures.», *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 24, n° 22, doi :10.1523/JNEUROSCI.0540-04.2004, p. 5216–29, ISSN 1529-2401. [256](#)
- BETTENCOURT, L. M. A., G. J. STEPHENS, M. I. HAM et G. W. GROSS. 2007, «Functional structure of cortical neuronal networks grown in vitro», *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 75, n° 2, doi :10.1103/PhysRevE.75.021915, p. 021 915, ISSN 15393755. [232](#)
- BI, G. Q. et M. M. POO. 1998, «Synaptic modifications in cultured hippocampal neurons : dependence on spike timing, synaptic strength, and postsynaptic cell type.», *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 18, n° 24, p. 10 464–72, ISSN 0270-6474. [259](#)
- BI, G.-Q. et M.-M. POO. 2001, «Synaptic Modification by Correlated Activity : Hebb's Postulate Revisited», *Annual Review of Neuroscience*, vol. 24, n° 1, doi :10.1146/annurev.neuro.24.1.139, p. 139–166, ISSN 0147-006X. [259](#)
- BIALEK, W. et A. ZEE. 1990, «Coding and computation with neural spike trains», *Journal of Statistical Physics*, vol. 59, n° 1-2, doi :10.1007/BF01015565, p. 103–115, ISSN 00224715. [242](#)

- BIENENSTOCK, E. 1996, «On the dimensionality of cortical graphs.», *Journal of physiology, Paris*, vol. 90, n° 3-4, p. 251–6, ISSN 0928-4257. [256](#)
- BIFFI, E. 2015, «Microfluidic and compartmentalized platforms for neurobiological research», *Microfluidic and Compartmentalized Platforms for Neurobiological Research*, vol. 39, n° 3, doi : 10.1007/978-1-4939-2510-0, p. 1–248, ISSN 0278-940X. [226](#)
- BIOWALL, E. «Turing Neuronal Networks», URL <http://lslwww.epfl.ch/biowall/VersionE/ApplicationsE/TuringE.html>. [255](#)
- BLANKENSHIP, A. G. et M. B. FELLER. 2009, «Mechanisms underlying spontaneous patterned activity in developing neural circuits.», *Nature Reviews Neuroscience*, vol. 11, n° 1, doi :10.1038/nrn2759, p. 18–29, ISSN 1471-0048. [247](#)
- BONIFAZI, P., M. E. RUARO et V. TORRE. 2005, «Statistical properties of information processing in neuronal networks», *European Journal of Neuroscience*, vol. 22, n° 11, doi :10.1111/j.1460-9568.2005.04464.x, p. 2953–2964, ISSN 0953816X. [256](#)
- BOTTANI, S. 1995, «Pulse-Coupled Relaxation Oscillators : From Biological Synchronization to Self-Organized Criticality», *Physical Review Letters*, vol. 74, n° 21, doi :10.1103/PhysRevLett.74.4189, p. 4189–4192, ISSN 0031-9007. [250](#)
- BOWER, J. M. et D. BEEMAN. 2007, «Constructing realistic neural simulations with GENESIS.», *Methods in molecular biology (Clifton, N.J.)*, vol. 401, doi :10.1007/978-1-59745-520-6_7, p. 103–125, ISSN 1064-3745. [262](#)
- BRETTE, R. et W. GERSTNER. 2005, «Adaptive exponential integrate-and-fire model as an effective description of neuronal activity.», *Journal of neurophysiology*, vol. 94, n° 5, p. 3637–3642. [267](#)
- BRUNEL, N. et N. BRUNEL. 2000, «Dynamics of sparsely connected networks of excitatory and inhibitory neurons», *Computational Neuroscience*, vol. 8, p. 183–208. [238](#)
- BRUNEL, N. et V. HAKIM. 1999, «Fast global oscillations in networks of integrate-and-fire neurons with low firing rates», *Neural computation*, vol. 11, n° 7, doi :10.1162/089976699300016179, p. 1621–71, ISSN 0899-7667. [238](#)
- BUGNICOURT, G., J. BROCARD, A. NICOLAS et C. VILLARD. 2014, «Nanoscale Surface Topography Reshapes Neuronal Growth in Culture», *Langmuir*, vol. 30, n° 15, doi :10.1021/la5001683, p. 4441–4449, ISSN 0743-7463. [226](#)
- BULLOCK, T. H. et D. H. PERKEL. 1968, «Neutral coding - A report based on an NRP work session», . [242](#)
- CARTER, M. et J. SHIEH. 2015a, «Manipulating Neural Activity», dans *Guide to Research Techniques in Neuroscience*, Elsevier, ISBN 9780128005118, p. 185–201, doi :10.1016/B978-0-12-800511-8.00008-3. [230](#)
- CARTER, M. et J. SHIEH. 2015b, «Visualizing Neural Function», dans *Guide to Research Techniques in Neuroscience*, Elsevier, ISBN 9780128005118, p. 167–183, doi :10.1016/B978-0-12-800511-8.00007-1. [230](#)
- CARTER, M., J. C. SHIEH, N. FARRA et G. HARRIS. *Guide to research techniques in neuroscience*, ISBN 9780128005118, 417 p.. [229](#)
- CERQUEIRA, S. R., Y.-S. LEE et M. B. BUNGE. 2018, «A Culture Model to Study Neuron-Schwann Cell-Astrocyte Interactions», dans *Methods in molecular biology (Clifton, N.J.)*, vol. 1739, p. 269–279, doi :10.1007/978-1-4939-7649-2_17. [239](#)

- CHALUPA, J., P. L. LEATH et G. R. REICH. 1979, «Bootstrap percolation on a Bethe lattice», *Journal of Physics C : Solid State Physics*, vol. 12, n° 1, doi :10.1088/0022-3719/12/1/008, p. L31–L35, ISSN 00223719. [235](#)
- CHIALVO, D. R. 2010, «Emergent complex neural dynamics», *Nature Physics*, vol. 6, n° 10, doi : 10.1038/nphys1803, p. 744–750, ISSN 17452473. [253](#)
- CHIAPPALONE, M., M. BOVE, A. VATO, M. TEDESCO et S. MARTINOIA. 2006, «Dissociated cortical networks show spontaneously correlated activity patterns during in vitro development», *Brain Research*, vol. 1093, n° 1, doi :10.1016/j.brainres.2006.03.049, p. 41–53, ISSN 00068993. [224](#), [236](#)
- CLAUSET, A., C. R. SHALIZI et M. E. J. NEWMAN. 2009, «Power-Law Distributions in Empirical Data», *SIAM Review*, vol. 51, n° 4, doi :10.1137/070710111, p. 661–703, ISSN 0036-1445. [253](#)
- CLOPATH, C., R. JOLIVET, A. RAUCH, H.-R. LÜSCHER et W. GERSTNER. 2007, «Predicting neuronal activity with simple models of the threshold type : Adaptive Exponential Integrate-and-Fire model with two compartments», *Neurocomputing*, vol. 70, n° 10-12, doi :10.1016/J.NEUCOM.2006.10.047, p. 1668–1673, ISSN 0925-2312. [267](#)
- COHEN, E., M. IVENSHITZ, V. AMOR-BAROUKH, V. GREENBERGER et M. SEGAL. 2008, «Determinants of spontaneous activity in networks of cultured hippocampus», *Brain Research*, vol. 1235, doi :10.1016/j.brainres.2008.06.022, p. 21–30, ISSN 00068993. [249](#)
- COHEN, O., A. KESELMAN, E. MOSES et collaborateurs. 2010, «Quorum percolation in living neural networks», *EPL*, vol. 89, n° 1, doi :10.1209/0295-5075/89/18008, p. 1–6, ISSN 02955075. [235](#)
- COREY, J. M. et E. L. FELDMAN. 2003, «Substrate patterning : An emerging technology for the study of neuronal behavior», *Experimental Neurology*, vol. 184, n° SUPPL. 1, doi :10.1016/S0014-4886(03)00392-3, p. S89–96, ISSN 00144886. [225](#)
- CORNER, M. A., J. VAN PELT, P. S. WOLTERS, R. E. BAKER et R. H. NUYTINCK. 2002, «Physiological effects of sustained blockade of excitatory synaptic transmission on spontaneously active developing neuronal networks - An inquiry into the reciprocal linkage between intrinsic biorhythms and neuroplasticity in early ontogeny», *Neuroscience and Biobehavioral Reviews*, vol. 26, n° 2, doi :10.1016/S0149-7634(01)00062-8, p. 127–185, ISSN 01497634. [249](#)
- COURBAGE, M., V. I. NEKORKIN et L. V. VDOVIN. 2007, «Chaotic oscillations in a map-based model of neural activity», *Chaos*, vol. 17, n° 4, doi :10.1063/1.2795435, p. 043 109, ISSN 10541500. [267](#)
- CRONIER, S., V. BERINGUE, A. BELLON, J.-M. PEYRIN et H. LAUDE. 2007, «Prion Strain- and Species-Dependent Effects of Antiprion Molecules in Primary Neuronal Cultures», *Journal of Virology*, vol. 81, n° 24, doi :10.1128/JVI.01502-07, p. 13 794–13 800, ISSN 0022-538X. [239](#)
- CULLEN, D. K., M. E. GILROY, H. R. IRONS et M. C. LAPLACA. 2010, «Synapse-to-neuron ratio is inversely related to neuronal density in mature neuronal cultures», *Brain Research*, vol. 1359, doi :10.1016/j.brainres.2010.08.058, p. 44–55, ISSN 00068993. [234](#)
- CULLEN, D. K., J. A. WOLF, V. N. VERNEKAR, J. VUKASINOVIC et M. C. LAPLACA. 2011, «Neural Tissue Engineering and Biohybridized Microsystems for Neurobiological Investigation In Vitro (Part 1)», *Critical Reviews™ in Biomedical Engineering*, vol. 39, n° 3, doi :10.1615/CritRevBiomedEng.v39.i3.30, p. 201–240, ISSN 0278-940X. [226](#)
- CUNTZ, H., M. W. REMME et B. TORBEN-NIELSEN. 2014, «The Computing Dendrite From Structure to Function», dans *The Computing Dendrite*, ISBN 978-1-4614-8093-8, p. 1–110, doi :10.1007/978-1-4614-8094-5. [262](#), [264](#)

- DAMBACH, H., D. HINKEROHE, N. PROCHNOW et collaborateurs. 2014, «Glia and epilepsy : Experimental investigation of antiepileptic drugs in an astroglia/microglia co-culture model of inflammation», *Epilepsia*, vol. 55, n° 1, doi :10.1111/epi.12473, p. 184–192, ISSN 00139580. [239](#)
- DAN, Y. et M.-M. POO. 2006, «Spike Timing-Dependent Plasticity : From Synapse to Perception», *Physiological Reviews*, vol. 86, n° 3, doi :10.1152/physrev.00030.2005, p. 1033–1048, ISSN 0031-9333. [259](#)
- DAUTH, S., B. M. MAOZ, S. P. SHEEHY et collaborateurs. 2017, «Neurons derived from different brain regions are inherently different in vitro : a novel multiregional brain-on-a-chip», *Journal of Neurophysiology*, vol. 117, n° 3, doi :10.1152/jn.00575.2016, p. 1320–1341, ISSN 0022-3077. [239](#)
- DE SANTOS-SIERRA, D., I. SENDIÑA-NADAL, I. LEYVA et collaborateurs. 2014, «Emergence of small-world anatomical networks in self-organizing clustered neuronal cultures», *PLoS ONE*, vol. 9, n° 1, doi :10.1371/journal.pone.0085828, p. e85 828, ISSN 19326203. [222](#), [233](#), [237](#)
- DEHGHANI, N., N. G. HATSOPOULOS, Z. D. HAGA et collaborateurs. 2012, «Avalanche analysis from multielectrode ensemble recordings in cat, monkey, and human cerebral cortex during wakefulness and sleep», *Frontiers in Physiology*, vol. 3 AUG, doi :10.3389/fphys.2012.00302, p. 1–18, ISSN 1664042X. [254](#)
- DELEGLISE, B., B. LASSUS, V. SOUBEYRE et collaborateurs. 2013, «Synapto-Protective Drugs Evaluation in Reconstructed Neuronal Network», *PLoS ONE*, vol. 8, n° 8, doi :10.1371/journal.pone.0071103, p. e71 103, ISSN 19326203. [239](#)
- DELORENZO, R. J., D. A. SUN, R. E. BLAIR et S. SOMBATI. 2007, «An in vitro model of Stroke-Induced Epilepsy : Elucidation of The roles of Glutamate and Calcium in The induction and Maintenance of Stroke-Induced Epileptogenesis», *International Review of Neurobiology*, vol. 81, doi :10.1016/S0074-7742(06)81005-6, p. 59–84, ISSN 00747742. [239](#)
- DEMARSE, T. B., D. A. WAGENAAR, A. W. BLAU et S. M. POTTER. 2001, «The neurally controlled animat : Biological brains acting with simulated bodies», *Autonomous Robots*, vol. 11, n° 3, doi : 10.1023/A:1012407611130, p. 305–310, ISSN 09295593. [261](#)
- DICHTER, M. A. 1978, «Rat cortical neurons in cell culture : culture methods, cell morphology, electrophysiology, and synapse formation.», *Brain research*, vol. 149, n° 2, p. 279–93, ISSN 0006-8993. [223](#)
- DINH, N.-D., Y.-Y. CHIANG, H. HARDELAUF et collaborateurs. 2013, «Microfluidic construction of minimalistic neuronal co-cultures.», *Lab on a chip*, vol. 13, n° 7, doi :10.1039/c3lc41224e, p. 1402–12, ISSN 1473-0189. [239](#)
- DINH, N.-D., Y.-Y. CHIANG, H. HARDELAUF et collaborateurs. 2014, «Preparation of neuronal co-cultures with single cell precision», *Journal of visualized experiments : JoVE*, , n° 87, doi :10.3791/51389, p. e51 389, ISSN 1940-087X. [239](#)
- DLESMANN, M., M. O. GEWALTIG et A. AERTSEN. 1999, «Stable propagation of synchronous spiking in cortical neural networks», *Nature*, vol. 402, n° 6761, doi :10.1038/990101, p. 529–533, ISSN 00280836. [243](#)
- DONOHUE, D. E. et G. A. ASCOLI. 2011, «Automated reconstruction of neuronal morphology : An overview», *Brain Research Reviews*, vol. 67, n° 1-2, doi :10.1016/j.brainresrev.2010.11.003, p. 94–102, ISSN 01650173. [263](#)
- DOWNES, J. H., M. W. HAMMOND, D. XYDAS et collaborateurs. 2012, «Emergence of a small-world functional network in cultured neurons», *PLoS Computational Biology*, vol. 8, n° 5, doi :10.1371/journal.pcbi.1002522, p. e1002 522, ISSN 1553734X. [236](#), [237](#)

- EBERHARD, J. P., A. WANNER et G. WITTUM. 2006, «NeuGen : A tool for the generation of realistic morphology of cortical neurons and neural networks in 3D», *Neurocomputing*, vol. 70, n° 1-3, doi :10.1016/j.neucom.2006.01.028, p. 327–342, ISSN 09252312. [264](#)
- ECKMANN, J. P., O. FEINERMAN, L. GRUENDLINGER et collaborateurs. 2007, «The physics of living neural networks», *Physics Reports*, vol. 449, n° 1-3, doi :10.1016/j.physrep.2007.02.014, p. 54–76, ISSN 03701573. [230](#)
- ECKMANN, J. P., S. JACOBI, S. MAROM, E. MOSES et C. ZBINDEN. 2008, «Leader neurons in population bursts of 2D living neural networks», *New Journal of Physics*, vol. 10, n° 10, doi : 10.1088/1367-2630/10/1/015011, p. 15 011–15 011, ISSN 13672630. [249](#), [250](#)
- ECKMANN, J.-P., E. MOSES, O. STETTER, T. TLUSTY et C. ZBINDEN. 2010, «Leaders of neuronal cultures in a quorum percolation model.», *Frontiers in computational neuroscience*, vol. 4, n° September, doi :10.3389/fncom.2010.00132, p. 1–12, ISSN 1662-5188. [235](#), [236](#)
- ESSEN BIOSCIENCE. 2018, «Neurite Analysis», URL <https://www.essenbioscience.com/en/applications/live-cell-assays/neurite-analysis/>. [221](#)
- EVANS, R. C. et S. POLAVARAM. 2013, «Growing a garden of neurons.», *Frontiers in neuroinformatics*, vol. 7, doi :10.3389/fninf.2013.00017, p. 17, ISSN 1662-5196. [264](#)
- EYTAN, D. 2011, «Representation and Learning in Neuronal Networks : A Conceptual Nervous System Approach», *Rambam Maimonides Medical Journal*, vol. 2, n° 3, doi :10.5041/RMMJ.10054, p. 1–12, ISSN 20769172. [260](#)
- EYTAN, D. et S. MAROM. 2006, «Dynamics and Effective Topology Underlying Synchronization in Networks of Cortical Neurons», *Journal of Neuroscience*, vol. 26, n° 33, doi :10.1523/JNEUROSCI.1627-06.2006, p. 8465–8476, ISSN 0270-6474. [247](#), [249](#), [250](#), [252](#), [253](#), [256](#), [257](#)
- FEIGE, U., V. S. MIRROKNI et J. A. N. V. AK. 2011, «BIFURCATION ANALYSIS OF A GENERAL CLASS OF NONLINEAR INTEGRATE-AND-FIRE NEURONS», *SIAM J. Appl. Math.*, vol. 40, n° 4, doi : 10.1137/090750688, p. 1133–1153. [267](#)
- FEINERMAN, O., A. ROTEM et E. MOSES. 2008, «Reliable neuronal logic devices from patterned hippocampal cultures», *Nature Physics*, vol. 4, n° 12, doi :10.1038/nphys1099, p. 967–973, ISSN 1745-2473. [245](#), [246](#)
- FELDT, S., P. BONIFAZI et R. COSSART. 2011, «Dissecting functional connectivity of neuronal microcircuits : Experimental and theoretical insights», *Trends in Neurosciences*, vol. 34, n° 5, doi : 10.1016/j.tins.2011.02.007, p. 225–236, ISSN 01662236. [236](#)
- FELLER, M. B. 1999, «Spontaneous correlated activity in developing neural circuits», *Neuron*, vol. 22, n° 4, doi :10.1016/S0896-6273(00)80724-2, p. 653–656, ISSN 08966273. [247](#)
- FERRÁNDEZ, J. M., V. LORENTE, F. DE LA PAZ et E. FERNÁNDEZ. 2013, «Training biological neural cultures : Towards Hebbian learning», *Neurocomputing*, vol. 114, doi :10.1016/j.neucom.2012.09.031, p. 3–8, ISSN 09252312. [259](#), [260](#)
- FITZHUGH, R. 1961, «Impulses and Physiological States in Theoretical Models of Nerve Membrane.», *Biophysical journal*, vol. 1, n° 6, p. 445–66, ISSN 0006-3495. [267](#)
- FITZJOHN, S., L. PICKARD, J. DUCKWORTH et collaborateurs. 2001, «An electrophysiological characterisation of long-term potentiation in cultured dissociated hippocampal neurones», *Neuropharmacology*, vol. 41, n° 6, doi :10.1016/S0028-3908(01)00128-9, p. 693–699, ISSN 00283908. [231](#)

- FREGA, M. 2016a, *Dissociated Neuronal Networks Coupled to Micro-Electrode Arrays Devices*, 9–29 p., doi :10.1007/978-3-319-30237-9_2. [229](#)
- FREGA, M. 2016b, «In Vitro Neuronal Networks», dans *Neuronal Network Dynamics in 2D and 3D in vitro Neuroengineered Systems*, Springer t éd., Springer, Cham, ISBN 978-3-319-30237-9, p. 31–41, doi :https://doi.org/10.1007/978-3-319-30237-9_3. [227](#), [229](#)
- FRIMAT, J.-P., J. SISNAISKE, S. SUBBIAH et collaborateurs. 2010, «The network formation assay : a spatially standardized neurite outgrowth analytical display for neurotoxicity screening», *Lab on a Chip*, vol. 10, n° 6, doi :10.1039/b922193j, p. 701, ISSN 1473-0197. [239](#)
- FURSHPAN, E. J. et D. D. POTTER. 1989, «Seizure-like activity and cellular damage in rat hippocampal neurons in cell culture.», *Neuron*, vol. 3, n° 2, p. 199–207, ISSN 0896-6273. [247](#)
- GAL, A. et S. MAROM. 2013, «Entrainment of the intrinsic dynamics of single isolated neurons by natural-like input.», *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 33, n° 18, doi :10.1523/JNEUROSCI.3763-12.2013, p. 7912–8, ISSN 1529-2401. [239](#)
- GEORGE MASON, U. «NeuroMorpho», URL <http://www.neuromorpho.org/>. [263](#)
- GERSTNER, W., A. K. KREITER, H. MARKRAM et A. V. M. HERZ. 1997, «Neural codes : firing rates and beyond.», *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, n° 24, doi :VL-94, p. 12 740–12 741, ISSN 0027-8424. [242](#), [244](#)
- GEWALTIG, M.-O., M. DIESMANN et A. AERTSEN. 2001, «Propagation of cortical synfire activity : survival probability in single trials and stability in the mean», *Neural Networks*, vol. 14, n° 6-7, doi :10.1016/S0893-6080(01)00070-3, p. 657–673, ISSN 0893-6080. [243](#)
- GHIGLIAZZA, R. M. et P. HOLMES. 2004, «Minimal Models of Bursting Neurons : How Multiple Currents, Conductances, and Timescales Affect Bifurcation Diagrams», *SIAM Journal on Applied Dynamical Systems*, vol. 3, n° 4, doi :10.1137/030602307, p. 636–670, ISSN 1536-0040. [266](#)
- GILSON, M., A. N. BURKITT, D. B. GRAYDEN, D. A. THOMAS et J. L. VAN HEMMEN. 2009, «Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. I. Input selectivity-strengthening correlated input pathways», *Biological Cybernetics*, vol. 101, n° 2, doi :10.1007/s00422-009-0319-4, p. 81–102, ISSN 03401200. [244](#)
- GIREESH, E. D. et D. PLENZ. 2008, «Neuronal avalanches organize as nested theta- and beta/gamma-oscillations during development of cortical layer 2/3», *Proceedings of the National Academy of Sciences*, vol. 105, n° 21, doi :10.1073/pnas.0800537105, p. 7576–7581, ISSN 0027-8424. [252](#)
- GISIGER, T. 2001, «Scale invariance in biology : Coincidence or footprint of a universal mechanism?», *Biological Reviews of the Cambridge Philosophical Society*, vol. 76, n° 2, doi :10.1017/S1464793101005607, p. 161–209, ISSN 14647931. [254](#)
- GRAHAM, B. P. et A. VAN OUYEN. 2006, «Mathematical modelling and numerical simulation of the morphological development of neurons.», *BMC neuroscience*, vol. 7 Suppl 1, doi :10.1186/1471-2202-7-S1-S9, p. S9, ISSN 1471-2202. [263](#), [264](#)
- GRITSUN, T. A., J. LE FEBER et W. L. C. RUTTEN. 2012, «Growth dynamics explain the development of spatiotemporal burst activity of young cultured neuronal networks in detail.», *PloS one*, vol. 7, n° 9, doi :10.1371/journal.pone.0043352, p. e43 352, ISSN 1932-6203. [238](#), [263](#), [265](#)
- GROSS, G. 2011, «Multielectrode arrays», *Scholarpedia*, vol. 6(3), n° 1994, doi :10.4249/scholarpedia.5749, p. 5749, ISSN 1941-6016. [228](#), [232](#)

- GROSS, G. W., E. RIESKE, G. W. KREUTZBERG et A. MEYER. 1977, «A new fixed-array multi-microelectrode system designed for long-term monitoring of extracellular single unit neuronal activity in vitro», *Neuroscience Letters*, vol. 6, n° 2-3, doi :10.1016/0304-3940(77)90003-9, p. 101–105, ISSN 03043940. [229](#)
- GROSS, P. G., E. P. KARTALOV, A. SCHERER et L. P. WEINER. 2007, «Applications of microfluidics for neuronal studies», *Journal of the Neurological Sciences*, vol. 252, n° 2, doi :10.1016/j.jns.2006.11.009, p. 135–143, ISSN 0022510X. [226](#)
- GROSS, P. G., L. P. WEINER, E. P. KARTALOV et A. SCHERER. 2005, «Microfluidic techniques for studying the nervous system.», *Critical reviews in neurobiology*, vol. 17, n° 3-4, p. 119–44, ISSN 0892-0915. [226](#)
- GÜTIG, R. et H. SOMPOLINSKY. 2006, «The tempotron : A neuron that learns spike timing-based decisions», *Nature Neuroscience*, vol. 9, n° 3, doi :10.1038/nn1643, p. 420–428, ISSN 10976256. [256](#)
- HARRISON, R. G. 1910, «The outgrowth of the nerve fiber as a mode of protoplasmic movement», *Journal of Experimental Zoology*, vol. 9, n° 4, doi :10.1002/jez.1400090405, p. 787–846, ISSN 1097010X. [222](#)
- HARRISON, R. G., M. J. GREENMAN, F. P. MALL et C. M. JACKSON. 1907, «Observations of the living developing nerve fiber», *The Anatomical Record*, vol. 1, n° 5, doi :10.1002/ar.1090010503, p. 116–128, ISSN 10970185. [222](#)
- HEBB, D. O. 1949, *The Organization of Behavior; A Neuropsychological Theory*, vol. 63, 2002^e éd., Francis and Taylor, ISBN 0805843000, 633 p., doi :10.2307/1418888. [256](#)
- HERMAN CUNTZ, MICHEL W.H.REMME, B. T.-N. 2009, *The Computing Dendrite*, vol. 8, ISBN 9783540604525, 25–42 p., doi :10.1007/978-0-387-93797-7. [264](#)
- HERNÁNDEZ-NAVARRO, L., J. G. ORLANDI, B. CERRUTI, E. VIVES et J. SORIANO. 2017, «Dominance of Metric Correlations in Two-Dimensional Neuronal Cultures Described through a Random Field Ising Model», *Physical Review Letters*, vol. 118, n° 20, doi :10.1103/PhysRevLett.118.208101, p. 208 101, ISSN 10797114. [238](#)
- HINDMARSH, J. L. et R. M. ROSE. 1984, «A Model of Neuronal Bursting Using Three Coupled First Order Differential Equations», *Proceedings of the Royal Society B : Biological Sciences*, vol. 221, n° 1222, doi :10.1098/rspb.1984.0024, p. 87–102, ISSN 0962-8452. [267](#)
- HINES, M. L. et N. T. CARNEVALE. 2001, «NEURON : A tool for neuroscientists», *Neuroscientist*, vol. 7, n° 2, doi :10.1177/107385840100700207, p. 123–135, ISSN 10738584. [262](#)
- HOCQUEMILLER, M., S. VITRY, S. BIGOU et collaborateurs. 2010, «GAP43 overexpression and enhanced neurite outgrowth in mucopolysaccharidosis type IIIB cortical neuron cultures», *Journal of Neuroscience Research*, vol. 88, n° 1, doi :10.1002/jnr.22190, p. 202–213, ISSN 03604012. [239](#), [240](#)
- HONEY, C. J., O. SPORNS, L. CAMMOUN et collaborateurs. 2009, «Predicting human resting-state functional connectivity from structural connectivity», *Proceedings of the National Academy of Sciences*, vol. 106, n° 6, doi :10.1073/pnas.0811168106, p. 2035–2040, ISSN 0027-8424. [236](#)
- JIMBO, Y., H. ROBINSON et A. KAWANA. 1998, «Strengthening of synchronized activity by tetanic stimulation in cortical cultures : application of planar electrode arrays», *IEEE Transactions on Biomedical Engineering*, vol. 45, n° 11, doi :10.1109/10.725326, p. 1297–1304, ISSN 00189294. [259](#), [260](#)

- JIMBO, Y., T. TATENO et H. ROBINSON. 1999, «Simultaneous Induction of Pathway-Specific Potentiation and Depression in Networks of Cortical Neurons», *Biophysical Journal*, vol. 76, n° 2, doi : 10.1016/S0006-3495(99)77234-6, p. 670–678, ISSN 00063495. [259](#), [260](#)
- JOVANOVIĆ, S. et S. ROTTER. 2016, «Interplay between Graph Topology and Correlations of Third Order in Spiking Neuronal Networks», *PLOS Computational Biology*, vol. 12, n° 6, doi :10.1371/journal.pcbi.1004963, p. e1004963, ISSN 1553-7358. [236](#)
- KELAVA, I. et M. A. LANCASTER. 2016, «Dishing out mini-brains : Current progress and future prospects in brain organoid research.», *Developmental biology*, vol. 420, n° 2, doi :10.1016/j.ydbio.2016.06.037, p. 199–209, ISSN 1095-564X. [227](#), [228](#)
- KHAZIPOV, R. et H. J. LUHMANN. 2006, «Early patterns of electrical activity in the developing cerebral cortex of humans and rodents.», *Trends in neurosciences*, vol. 29, n° 7, doi :10.1016/j.tins.2006.05.007, p. 414–418, ISSN 0166-2236. [247](#)
- KIM, H. J. et J. MAGRANÉ. 2011, «Isolation and Culture of Neurons and Astrocytes from the Mouse Brain Cortex», dans *Methods in molecular biology (Clifton, N.J.)*, vol. 793, p. 63–75, doi :10.1007/978-1-61779-328-8_4. [223](#)
- KIRKBY, L. A., G. S. SACK, A. FIRL et M. B. FELLER. 2013, «A role for correlated spontaneous activity in the assembly of neural circuits», *Neuron*, vol. 80, n° 5, doi :10.1016/j.neuron.2013.10.030, p. 1129–1144, ISSN 08966273. [247](#)
- KLAUS, A., S. YU et D. PLENZ. 2011, «Statistical analyses support power law distributions found in neuronal avalanches», *PLoS ONE*, vol. 6, n° 5, doi :10.1371/journal.pone.0019779, p. e19779, ISSN 19326203. [252](#)
- KOENE, R. A., B. TIJMS, P. VAN HEES et collaborateurs. 2009, «NETMORPH : a framework for the stochastic generation of large scale neuronal networks with realistic neuron morphologies.», *Neuroinformatics*, vol. 7, n° 3, doi :10.1007/s12021-009-9052-3, p. 195–210, ISSN 1559-0089. [264](#), [265](#)
- KRIEGSTEIN, A. R. et M. A. DICHTER. 1983, «Morphological classification of rat cortical neurons in cell culture.», *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 3, n° 8, p. 1634–1647, ISSN 0270-6474. [223](#)
- KUMAR, A., S. ROTTER et A. AERTSEN. 2010, «Spiking activity propagation in neuronal networks : Reconciling different perspectives on neural coding», *Nature Reviews Neuroscience*, vol. 11, n° 9, doi :10.1038/nrn2886, p. 615–627, ISSN 1471003X. [244](#)
- LÁBOS, E. 2000, «Codes, operations, measurements and neural networks», dans *BioSystems*, vol. 58, ISBN 0303-2647 (Print), ISSN 03032647, p. 9–18, doi :10.1016/S0303-2647(00)00101-5. [242](#), [244](#)
- LANCASTER, M. A., M. RENNER, C.-A. MARTIN et collaborateurs. 2013, «Cerebral organoids model human brain development and microcephaly», *Nature*, vol. 501, n° 7467, doi :10.1038/nature12517, p. 373–379, ISSN 0028-0836. [227](#)
- LAPICQUE, L. 1907, «Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation», *J. Physiol. Pathol. Gen.*, vol. 9, p. 620–635. [267](#)
- LEE, C.-T., R. M. BENDRIEM, W. W. WU et R.-F. SHEN. 2017, «3D brain Organoids derived from pluripotent stem cells : promising experimental models for brain development and neurodegenerative disorders», *Journal of Biomedical Science*, vol. 24, n° 1, doi :10.1186/s12929-017-0362-8, p. 59, ISSN 1423-0127. [228](#)

- LENK, K., B. PRIWITZER, L. YLÄ-OUTINEN et collaborateurs. 2016, «Simulation of developing human neuronal cell networks», *BioMedical Engineering OnLine*, vol. 15, n° 1, doi :10.1186/s12938-016-0226-6, p. 105, ISSN 1475-925X. [265](#)
- LI, Y., R. SUN, Y. WANG, H. LI et X. ZHENG. 2016, «A Novel Robot System Integrating Biological and Mechanical Intelligence Based on Dissociated Neural Network-Controlled Closed-Loop Environment», *PLOS ONE*, vol. 11, n° 11, doi :10.1371/journal.pone.0165600, p. e0165 600, ISSN 1932-6203. [261](#)
- LINDEN, M. L., A. J. HEYNEN, R. H. HASLINGER et M. F. BEAR. 2009, «Thalamic activity that drives visual cortical plasticity», *Nature Neuroscience*, vol. 12, n° 4, doi :10.1038/nn.2284, p. 390–392, ISSN 10976256. [233](#)
- LITT, B. et J. ECHAUZ. 2002, «Prediction of epileptic seizures.», *The Lancet. Neurology*, vol. 1, n° 1, p. 22–30, ISSN 1474-4422. [247](#)
- LITVAK, V., H. SOMPOLINSKY, I. SEGEV et M. ABELES. 2003, «On the transmission of rate code in long feedforward networks with excitatory-inhibitory balance.», *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 23, n° 7, doi :23/7/3006[pii], p. 3006–3015, ISSN 1529-2401. [243](#)
- LIU, M. G., X. F. CHEN, T. HE, Z. LI et J. CHEN. 2012, «Use of multi-electrode array recordings in studies of network synaptic plasticity in both time and space», *Neuroscience Bulletin*, vol. 28, n° 4, doi :10.1007/s12264-012-1251-5, p. 409–422, ISSN 16737067. [229](#)
- LONARDONI, D., H. AMIN, S. DI MARCO et collaborateurs. 2017, «Recurrently connected and localized neuronal communities initiate coordinated spontaneous activity in neuronal networks», *PLoS Computational Biology*, vol. 13, n° 7, doi :10.1371/journal.pcbi.1005672, p. e1005 672, ISSN 15537358. [238](#)
- LÓPEZ-CRUZ, P. L., C. BIELZA, P. LARRAÑAGA, R. BENAVIDES-PICCIONE et J. DEFELIPE. 2011, «Models and simulation of 3D neuronal dendritic trees using Bayesian networks», *Neuroinformatics*, vol. 9, n° 4, doi :10.1007/s12021-011-9103-4, p. 347–369, ISSN 15392791. [264](#)
- MADHAVAN, R., Z. C. CHAO et S. M. POTTER. 2007, «Plasticity of recurring spatiotemporal activity patterns in cortical networks.», *Physical biology*, vol. 4, n° 3, doi :10.1088/1478-3975/4/3/005, p. 181–93, ISSN 1478-3975. [253](#)
- MAEDA, E., Y. KURODA, H. P. C. ROBINSON et A. KAWANA. 1998, «Modification of parallel activity elicited by propagating bursts in developing networks of rat cortical neurones», *European Journal of Neuroscience*, vol. 10, n° 2, doi :10.1046/j.1460-9568.1998.00062.x, p. 488–496, ISSN 0953816X. [259](#)
- MAEDA, E., H. P. ROBINSON et A. KAWANA. 1995, «The mechanisms of generation and propagation of synchronized bursting in developing networks of cortical neurons.», *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 15, n° 10, p. 6834–45, ISSN 0270-6474. [239](#)
- MAHESWARANATHAN, N., S. FERRARI, A. M. J. VANDONGEN et C. S. HENRIQUEZ. 2012, «Emergent bursting and synchrony in computer simulations of neuronal cultures.», *Frontiers in computational neuroscience*, vol. 6, doi :10.3389/fncom.2012.00015, p. 15, ISSN 1662-5188. [236](#), [238](#), [265](#)
- MÄKI-MARTTUNEN, T., J. AČIMOVIĆ, K. RUOHONEN et M. L. LINNE. 2013, «Structure-Dynamics Relationships in Bursting Neuronal Networks Revealed Using a Prediction Framework», *PLoS ONE*, vol. 8, n° 7, doi :10.1371/journal.pone.0069373, ISSN 19326203. [238](#)

- MARI, J. F., J. H. SAITO, A. F. NEVES et collaborateurs. 2015, «Quantitative Analysis of Rat Dorsal Root Ganglion Neurons Cultured on Microelectrode Arrays Based on Fluorescence Microscopy Image Processing», *International Journal of Neural Systems*, vol. 25, n° 08, doi :10.1142/S0129065715500331, p. 1550 033, ISSN 0129-0657. [233](#)
- MARKRAM, H. 2011, «A history of spike-timing-dependent plasticity», *Frontiers in Synaptic Neuroscience*, vol. 3, n° August, doi :10.3389/fnsyn.2011.00004, p. 1–24, ISSN 16633563. [255](#), [257](#), [258](#)
- MARKRAM, H., J. LÜBKE, M. FROTSCHER et B. SAKMANN. 1997, «Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs», *Science*, vol. 275, n° 5297, doi :10.1126/science.275.5297.213, p. 213–215, ISSN 00368075. [259](#)
- MARKRAM, H., E. MULLER, S. RAMASWAMY et E. AL. 2015, «Reconstruction and Simulation of Neocortical Microcircuitry», *Cell*, vol. 163, n° 2, doi :10.1016/j.cell.2015.09.029, p. 456–492, ISSN 1097-4172. [262](#)
- MAROM, S. et D. EYTAN. 2004, «Learning in ex-vivo developing networks of cortical neurons», *Progress in Brain Research*, vol. 147, n° SPEC. ISS., doi :10.1016/S0079-6123(04)47014-9, p. 189–199, ISSN 00796123. [239](#)
- MAROM, S. et D. EYTAN. 2005, «Learning in ex-vivo developing networks of cortical neurons», dans *Progress in brain research*, vol. 147, p. 189–199, doi :10.1016/S0079-6123(04)47014-9. [258](#)
- MAROM, S. et G. SHAHAF. 2002, «Development, learning and memory in large random networks of cortical neurons : lessons beyond anatomy.», *Quarterly reviews of biophysics*, vol. 35, n° 1, doi :10.1017/S0033583501003742, p. 63–87, ISSN 0033-5835. [229](#), [233](#), [249](#)
- MASQUELIER, T., G. DECO, P. WOLTERS, R. BAKER et R. NUYTINCK. 2013, «Network bursting dynamics in excitatory cortical neuron cultures results from the combination of different adaptive mechanisms.», *PloS one*, vol. 8, n° 10, doi :10.1371/journal.pone.0075824, p. e75 824, ISSN 1932-6203. [238](#)
- MATSUDA, N., A. ODAWARA, H. KATOH et collaborateurs. 2018, «Detection of synchronized burst firing in cultured human induced pluripotent stem cell-derived neurons using a 4-step method», *Biochemical and Biophysical Research Communications*, vol. 497, n° 2, doi :10.1016/j.bbrc.2018.02.117, p. 612–618, ISSN 10902104. [239](#)
- MATURANA, H. R. et F. J. VARELA. 1994, *L'arbre de la connaissance. Racines biologiques de la compréhension humaine*, Addison-Wesley France, ISBN 287908072X, 256 p.. [221](#)
- MAZZONI, A., F. D. BROCCARD, E. GARCIA-PEREZ et collaborateurs. 2007, «On the dynamics of the spontaneous activity in neuronal networks.», *PloS one*, vol. 2, n° 5, doi :10.1371/journal.pone.0000439, p. e439, ISSN 1932-6203. [247](#), [249](#), [250](#), [253](#)
- MCCULLOCH, W. S. et W. PITTS. 1943, «A logical calculus of the ideas immanent in nervous activity», *The Bulletin of Mathematical Biophysics*, vol. 5, n° 4, doi :10.1007/BF02478259, p. 115–133, ISSN 00074985. [255](#)
- MEMELLI, H., B. TORBEN-NIELSEN et J. KOZLOSKI. 2013, «Self-referential forces are sufficient to explain different dendritic morphologies», *Frontiers in Neuroinformatics*, vol. 7, doi :10.3389/fninf.2013.00001, p. 1, ISSN 1662-5196. [264](#), [265](#)
- MERCATI, O., A. DANCKAERT, G. ANDRÉ-LEROUX et collaborateurs. 2013, «Contactin 4, -5 and -6 differentially regulate neuritogenesis while they display identical PTPRG binding sites.», *Biology open*, vol. 2, n° 3, doi :10.1242/bio.20133343, p. 324–34, ISSN 2046-6390. [240](#)

- MILLET, L. J. et M. U. GILLETTE. 2012, «Over a century of neuron culture : From the hanging drop to microfluidic devices», *Yale Journal of Biology and Medicine*, vol. 85, n° 4, p. 501–521, ISSN 00440086. [222](#), [225](#)
- MILLET, L. J., M. E. STEWART, R. G. NUZZO et M. U. GILLETTE. 2010, «Guiding neuron development with planar surface gradients of substrate cues deposited using microfluidic devices», *Lab on a Chip*, vol. 10, n° 12, doi :10.1039/c001552k, p. 1525, ISSN 1473-0197. [225](#)
- MOREFIELD, S. I., E. W. KEEFER, K. D. CHAPMAN et G. W. GROSS. 2000, «Drug evaluations using neuronal networks cultured on microelectrode arrays.», *Biosensors & bioelectronics*, vol. 15, n° 7-8, p. 383–96, ISSN 0956-5663. [239](#)
- MURRAY, J. D., A. ANTICEVIC, M. GANCSOS et collaborateurs. 2014, «Linking Microcircuit Dysfunction to Cognitive Impairment : Effects of Disinhibition Associated with Schizophrenia in a Cortical Working Memory Model», *Cerebral Cortex*, vol. 24, n° 4, doi :10.1093/cercor/bhs370, p. 859–872, ISSN 1047-3211. [254](#)
- NAGUMO, J., S. ARIMOTO et S. YOSHIKAWA. 1962, «An active pulse transmission line simulating nerve axon», *Proc IRE.*, vol. 50, p. 2061–2070. [267](#)
- NAKANISHI, K. et F. KUKITA. 1998, «Functional synapses in synchronized bursting of neocortical neurons in culture», *Brain Research*, vol. 795, n° 1-2, doi :10.1016/S0006-8993(98)00283-2, p. 137–146, ISSN 00068993. [234](#)
- NAPOLI, A., J. XIE et I. OBEID. 2014, «Understanding the temporal evolution of neuronal connectivity in cultured networks using statistical analysis», *BMC Neuroscience*, vol. 15, n° 1, doi : 10.1186/1471-2202-15-17, p. 17, ISSN 14712202. [236](#), [249](#)
- NAUD, R., N. MARCILLE, C. CLOPATH et W. GERSTNER. 2008, «Firing patterns in the adaptive exponential integrate-and-fire model», *Biological Cybernetics*, vol. 99, n° 4-5, p. 335–347. [267](#)
- NETO, E., L. LEITÃO, D. M. SOUSA et collaborateurs. 2016, «Compartmentalized Microfluidic Platforms : The Unrivaled Breakthrough of In Vitro Tools for Neurobiological Research», *The Journal of Neuroscience*, vol. 36, n° 46, doi :10.1523/JNEUROSCI.1748-16.2016, p. 11 573–11 584, ISSN 0270-6474. [239](#)
- NOVELLINO, A., P. D'ANGELO, L. COZZI et collaborateurs. 2007, «Connecting Neurons to a Mobile Robot : An In Vitro Bidirectional Neural Interface», *Computational Intelligence and Neuroscience*, vol. 2007, doi :10.1155/2007/12725, p. 1–13, ISSN 1687-5265. [261](#)
- OCKER, G. K., Y. HU, M. A. BUICE et collaborateurs. 2017a, «From the statistics of connectivity to the statistics of spike times in neuronal networks», *Current Opinion in Neurobiology*, vol. 46, doi :10.1016/j.conb.2017.07.011, p. 109–119, ISSN 18736882. [236](#)
- OCKER, G. K., K. JOSIĆ, E. SHEA-BROWN et M. A. BUICE. 2017b, «Linking structure and activity in nonlinear spiking networks», *PLOS Computational Biology*, vol. 13, n° 6, doi :10.1371/journal.pcbi.1005583, p. e1005 583, ISSN 1553-7358. [236](#)
- ONG, C. S., P. YESANTHARAO, C. Y. HUANG et collaborateurs. 2018, «3D bioprinting using stem cells», *Pediatric Research*, vol. 83, n° 1-2, doi :10.1038/pr.2017.252, p. 223–231, ISSN 0031-3998. [228](#)
- VAN OUYEN, A. 2003, *Modeling neural development*, MIT Press, ISBN 0262220660, 339 p.. [263](#), [264](#)
- ORENSTEIN, O. et H. KEREN. 2017, «Development of Cortical Networks under Continuous Stimulation», *Frontiers in Molecular Neuroscience*, vol. 10, doi :10.3389/fnmol.2017.00018, p. 18, ISSN 1662-5099. [236](#)

- ORLANDI, J. G., J. SORIANO, E. ALVAREZ-LACALLE, S. TELLER et J. CASADEMUNT. 2013, «Noise focusing : the emergence of coherent activity in neuronal cultures supplementary information», *Nature Physics*. [238](#), [249](#), [250](#), [263](#), [265](#)
- ORZÓ, L. et E. LÁBOS. 2000, «Effects of the synaptic transmission's dynamics on possible neural codes», dans *BioSystems*, vol. 58, ISSN 03032647, p. 75–81, doi :10.1016/S0303-2647(00)00109-X. [242](#), [244](#)
- PARK, J., S. KIM, J. LI et A. HAN. 2014, «Axon length quantification microfluidic culture platform for growth and regeneration study», *Methods in Molecular Biology*, vol. 1162, doi :10.1007/978-1-4939-0777-9_7, p. 85–95, ISSN 10643745. [226](#)
- PASQUALE, V., P. MASSOBRIO, L. L. BOLOGNA, M. CHIAPPALONE et S. MARTINOIA. 2008, «Self-organization and neuronal avalanches in networks of dissociated cortical neurons», *Neuroscience*, vol. 153, n° 4, doi :10.1016/j.neuroscience.2008.03.050, p. 1354–1369, ISSN 03064522. [231](#), [249](#), [253](#)
- VAN PEL, D. M., K. HARADA, D. SONG, C. C. NAUS et W. C. SIN. 2018, «Modelling glioma invasion using 3D bioprinting and scaffold-free 3D culture», *Journal of Cell Communication and Signaling*, doi :10.1007/s12079-018-0469-z, ISSN 1873-9601. [228](#)
- VAN PELT, J. et H. B. M. UYLINGS. 2002, «Branching rates and growth functions in the outgrowth of dendritic branching patterns.», *Network (Bristol, England)*, vol. 13, n° 3, p. 261–81, ISSN 0954-898X. [264](#)
- PENG, H., M. HAWRYLYCZ, J. ROSKAMS et collaborateurs. 2015, «BigNeuron : Large-Scale 3D Neuron Reconstruction from Optical Microscopy Images», *Neuron*, vol. 87, n° 2, doi :10.1016/j.neuron.2015.06.036, p. 252–256, ISSN 10974199. [263](#)
- PENN, Y., M. SEGAL et E. MOSES. 2016, «Network synchronization in hippocampal neurons», *Proceedings of the National Academy of Sciences*, doi :10.1073/pnas.1515105113, p. 201515 105, ISSN 0027-8424. [237](#), [249](#), [250](#)
- PETERMANN, T., T. C. THIAGARAJAN, M. A. LEBEDEV et collaborateurs. 2009, «Spontaneous cortical activity in awake monkeys composed of neuronal avalanches», *Proceedings of the National Academy of Sciences*, vol. 106, n° 37, doi :10.1073/pnas.0904089106, p. 15 921–15 926, ISSN 0027-8424. [252](#)
- PEYRIN, J.-M., B. DELEGLISE, L. SAIAS et collaborateurs. 2011, «Axon diodes for the reconstruction of oriented neuronal networks in microfluidic chambers.», *Lab on a Chip*, vol. 11, n° 21, doi : 10.1039/c1lc20014c, p. 3663–3673, ISSN 14730189. [239](#)
- PINE, J. 2006, «A history of MEA development», dans *Advances in Network Electrophysiology : Using Multi-Electrode Arrays*, Springer US, ISBN 9780387258584, p. 3–23, doi :10.1007/0-387-25858-2_1. [229](#)
- POLI, D., V. P. PASTORE, S. MARTINOIA et P. MASSOBRIO. 2016, «From functional to structural connectivity using partial correlation in neuronal assemblies», *Journal of Neural Engineering*, vol. 13, n° 2, doi :10.1088/1741-2560/13/2/026023, p. 026 023, ISSN 17412552. [236](#)
- RAICHEL, M. E. 2006, «Neuroscience : The brain's dark energy», *Science*, vol. 314, n° 5803, doi : 10.1126/science., p. 1249–1250, ISSN 00368075. [247](#)
- RAMIREZ, J.-M., A. K. TRYBA et F. PEN. 2004, «Pacemaker neurons and neuronal networks : an integrative view ~», doi :10.1016/j.conb.2004.10.011. [250](#)

- REED, W. J. et B. D. HUGHES. 2002, «From gene families and genera to incomes and internet file sizes : Why power laws are so common in nature», *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 66, n° 6, doi :10.1103/PhysRevE.66.067103, p. 4, ISSN 1063651X. [254](#)
- RENAULT, R. 2015, *Emergent design of neuronal devices*, thèse de doctorat, Université Paris-diderot. URL <http://www.theses.fr/2015USPCC161>. [226](#), [231](#), [260](#)
- RENAULT, R., J.-B. DURAND, J.-L. VIOVY et C. VILLARD. 2016, «Asymmetric axonal edge guidance : a new paradigm for building oriented neuronal networks», *Lab Chip*, vol. 16, n° 12, doi :10.1039/C6LC00479B, p. 2188–2191, ISSN 1473-0197. [226](#)
- RENAULT, R., N. SUKENIK, S. DESCROIX et collaborateurs. 2015, «Combining microfluidics, optogenetics and calcium imaging to study neuronal communication in vitro», *PLoS ONE*, vol. 10, n° 4, doi :10.1371/journal.pone.0120680, ISSN 19326203. [226](#), [231](#)
- ROBERT, F., J. F. CLOIX et T. HEVOR. 2012, «Ultrastructural characterization of rat neurons in primary culture», *Neuroscience*, vol. 200, doi :10.1016/j.neuroscience.2011.10.002, p. 248–260, ISSN 03064522. [223](#)
- ROLSTON, J. D., D. A. WAGENAAR et S. M. POTTER. 2007, «Precisely timed spatiotemporal patterns of neural activity in dissociated cortical cultures.», *Neuroscience*, vol. 148, n° 1, doi :10.1016/j.neuroscience.2007.05.025, p. 294–303, ISSN 0306-4522. [249](#), [253](#), [256](#)
- VAN ROSSUM, M. C. W., G. G. TURRIGIANO et S. B. NELSON. 2002, «Fast propagation of firing rates through layered networks of noisy neurons.», *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 22, n° 5, doi :22/5/1956[pil], p. 1956–1966, ISSN 1529-2401. [243](#)
- ROTH, S., M. BISBAL, J. BROCARD et collaborateurs. 2012a, «How morphological constraints affect axonal polarity in mouse neurons», *PLoS ONE*, vol. 7, n° 3, doi :10.1371/journal.pone.0033623, p. e33623, ISSN 19326203. [226](#)
- ROTH, S., G. BUGNICOURT, M. BISBAL et collaborateurs. 2012b, «Neuronal architectures with axodendritic polarity above silicon nanowires», *Small*, vol. 8, n° 5, doi :10.1002/sml.201102325, p. 671–675, ISSN 16136810. [226](#)
- RULKOV, N. F. 2001, «Regularization of synchronized chaotic bursts», *Physical Review Letters*, vol. 86, n° 1, doi :10.1103/PhysRevLett.86.183, p. 183–186, ISSN 00319007. [267](#)
- DE SANTOS-SIERRA, D., I. SENDIÑA-NADAL, I. LEYVA et collaborateurs. 2015, «Graph-based unsupervised segmentation algorithm for cultured neuronal networks' structure characterization and modeling», *Cytometry Part A*, vol. 87, n° 6, doi :10.1002/cyto.a.22591, p. 513–523, ISSN 15524930. [233](#), [237](#)
- SCHROETER, M. S., P. CHARLESWORTH, M. G. KITZBICHLER, O. PAULSEN et E. T. BULLMORE. 2015, «Emergence of Rich-Club Topology and Coordinated Dynamics in Development of Hippocampal Functional Networks In Vitro», *Journal of Neuroscience*, vol. 35, n° 14, doi :10.1523/JNEUROSCI.4259-14.2015, p. 5459–5470, ISSN 0270-6474. [236](#)
- SEGEV, R. et E. BEN-JACOB. 2001, «Spontaneous synchronized bursting in 2D neural networks», dans *Physica A : Statistical Mechanics and its Applications*, vol. 302, North-Holland, ISSN 03784371, p. 64–69, doi :10.1016/S0378-4371(01)00441-1. [239](#)
- SELINGER, J. V., J. J. PANCAZIO et G. W. GROSS. 2004, «Measuring synchronization in neuronal networks for biosensor applications.», *Biosensors & bioelectronics*, vol. 19, n° 7, p. 675–83, ISSN 0956-5663. [241](#)

- SEPEHRI RAD, M., Y. CHOI, L. B. COHEN et collaborateurs. 2017, «Voltage and Calcium Imaging of Brain Activity», *Biophysical Journal*, vol. 113, n° 10, doi :10.1016/j.bpj.2017.09.040, p. 2160–2167, ISSN 15420086. [230](#)
- SHADLEN, M. N. et W. T. NEWSOME. 1994, «Noise, neural codes and cortical organization.», *Current opinion in neurobiology*, vol. 4, n° 4, p. 569–79, ISSN 0959-4388. [243](#)
- SHADLEN, M. N. et W. T. NEWSOME. 1998, «The variable discharge of cortical neurons : implications for connectivity, computation, and information coding.», *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 18, n° 10, p. 3870–96, ISSN 0270-6474. [243](#)
- SHAHAF, G. et S. MAROM. 2001a, «Learning in networks of cortical neurons.», *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 21, n° 22, doi :0270-6474/01/218782-07\$15.00/0, p. 8782–8, ISSN 1529-2401. [239](#)
- SHAHAF, G. et S. MAROM. 2001b, «Learning in networks of cortical neurons.», *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 21, n° 22, doi :0270-6474/01/218782-07\$15.00/0, p. 8782–8, ISSN 1529-2401. [260](#)
- SHEFI, O., E. BEN-JACOB et A. AYALI. 2002a, «Growth morphology of two-dimensional insect neural networks», *Neurocomputing*, vol. 44-46, doi :10.1016/S0925-2312(02)00451-4, p. 635–643, ISSN 09252312. [233](#), [237](#)
- SHEFI, O., I. GOLDING, R. SEGEV, E. BEN-JACOB et A. AYALI. 2002b, «Morphological characterization of in vitro neuronal networks», *Physical Review E*, vol. 66, n° 2, doi :10.1103/PhysRevE.66.021905, p. 021 905, ISSN 1063-651X. [222](#)
- SHEFI, O., I. GOLDING, R. SEGEV, E. BEN-JACOB et A. AYALI. 2002c, «Morphological characterization of in vitro neuronal networks», *Physical Review E*, vol. 66, n° 2, doi :10.1103/PhysRevE.66.021905, p. 021 905, ISSN 1063-651X. [237](#)
- SHILLCOCK, J. C., M. HAWRYLYCZ, S. HILL et H. PENG. 2016, «Reconstructing the brain : from image stacks to neuron synthesis», *Brain Informatics*, vol. 3, n° 4, doi :10.1007/s40708-016-0041-7, p. 205–209, ISSN 2198-4018. [264](#)
- SHULTZ, A. M., S. LEE, M. GUARALDI, T. B. SHEA et H. C. YANCO. 2017, «Robot-Embodied Neuronal Networks as an Interactive Model of Learning», *The Open Neurology Journal*, vol. 11, n° 1, doi :10.2174/1874205X01711010039, p. 39–47, ISSN 1874-205X. [261](#)
- SINAPAYEN, L., A. MASUMORI et T. IKEGAMI. 2017, «Learning by stimulation avoidance : A principle to control spiking neural networks dynamics», *PLoS ONE*, vol. 12, n° 2, doi :10.1371/journal.pone.0170388, p. e0170 388, ISSN 19326203. [261](#)
- SKAPER, S. D., C. ARGENTINI et M. BARBIERATO. 2012, «Culture of Neonatal Rodent Microglia, Astrocytes, and Oligodendrocytes from Cortex and Spinal Cord», dans *Methods in molecular biology (Clifton, N.J.)*, vol. 846, p. 67–77, doi :10.1007/978-1-61779-536-7_7. [223](#)
- SMETTERS, D., A. MAJEWSKA et R. YUSTE. 1999, «Detecting action potentials in neuronal populations with calcium imaging», *Methods*, doi :10.1006/meth.1999.0774, p. 215–21. [230](#)
- SOE, A. K., S. NAHAVANDI et K. KHOSHMANESH. 2012, «Neuroscience goes on a chip», doi : 10.1016/j.bios.2012.02.012. URL <http://www.ncbi.nlm.nih.gov/pubmed/22387040><http://linkinghub.elsevier.com/retrieve/pii/S0956566312000723>. [226](#)
- SOMERS, D. et N. KOPELL. 1993, «Rapid synchronization through fast threshold modulation», *Biol. Cybern.*, vol. 68, p. 393–407. [250](#)

- SORIANO, J., I. BRESKIN, E. MOSES et T. TLUSTY. 2007, «Percolation approach to study connectivity in living neural networks», dans *AIP Conference Proceedings*, vol. 887, ISBN 0735403902, ISSN 0094243X, p. 96–106, doi :10.1063/1.2709591. [235](#)
- SORIANO, J., M. RODRÍGUEZ MARTÍNEZ, T. TLUSTY et E. MOSES. 2008, «Development of input connections in neural cultures.», *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, n° 37, doi :10.1073/pnas.0707492105, p. 13 758–63, ISSN 1091-6490. [235](#)
- SPORNS, O., D. R. CHIALVO, M. KAISER et C. C. HILGETAG. 2004, «Organization, development and function of complex brain networks», *Trends Cogn Sci*, vol. 8, n° 9, doi :10.1016/j.tics.2004.07.008, p. 418–425, ISSN 1364-6613 (Print) 1364-6613 (Linking). [236](#)
- STENGER, D. A., G. W. GROSS, E. W. KEEFER et collaborateurs. 2001, «Detection of physiologically active compounds using cell-based biosensors.», *Trends in biotechnology*, vol. 19, n° 8, p. 304–9, ISSN 0167-7799. [241](#)
- STENGER, D. A., J. J. HICKMAN, K. E. BATEMAN et collaborateurs. 1998, «Microlithographic determination of axonal/dendritic polarity in cultured hippocampal neurons», *Journal of Neuroscience Methods*, vol. 82, n° 2, doi :10.1016/S0165-0270(98)00047-8, p. 167–173, ISSN 01650270. [226](#)
- STIEFEL, K. M. et T. J. SEJNOWSKI. 2007, «Mapping function onto neuronal morphology.», *Journal of neurophysiology*, vol. 98, n° 1, doi :10.1152/jn.00865.2006, p. 513–26, ISSN 0022-3077. [262](#), [264](#)
- STOECKLI, E. T. 2004, «Ig superfamily cell adhesion molecules in the brain.», *Handbook of experimental pharmacology*, vol. 165, n° 165, doi :10.1007/978-3-540-68170-0_12, p. 373–401, ISSN 0171-2004. [240](#)
- SURESH, J., M. RADOJICIC, L. L. PESCE et collaborateurs. 2016, «Network burst activity in hippocampal neuronal cultures : the role of synaptic and intrinsic currents», *Journal of Neurophysiology*, vol. 115, n° 6, doi :10.1152/jn.00995.2015, p. 3073–3089, ISSN 0022-3077. [250](#)
- TAYLOR, A. M., M. BLURTON-JONES, S. W. RHEE et collaborateurs. 2005, «A microfluidic culture platform for CNS axonal injury, regeneration and transport», *Nature Methods*, vol. 2, n° 8, doi :10.1038/nmeth777, p. 599–605, ISSN 15487091. [226](#)
- TESSADORI, J., M. BISIO, S. MARTINOIA et M. CHIAPPALONE. 2012, «Modular Neuronal Assemblies Embodied in a Closed-Loop Environment : Toward Future Integration of Brains and Machines», *Frontiers in Neural Circuits*, vol. 6, doi :10.3389/fncir.2012.00099, p. 99, ISSN 1662-5110. [261](#)
- TEUSCHER, C. 2001, *Turing's Connectionism. An Investigation of Neural Network Architectures. Integrating Connectionism and Cognitive Science*, Springer, ISBN 1852334754, 200 p.. [255](#)
- THOMAS, C. A., P. A. SPRINGER, G. E. LOEB, Y. BERWALD-NETTER et L. M. OKUN. 1972, «A miniature microelectrode array to monitor the bioelectric activity of cultured cells», *Experimental Cell Research*, vol. 74, n° 1, doi :10.1016/0014-4827(72)90481-8, p. 61–66, ISSN 00144827. [229](#)
- THOMPSON, E. J. 1977, «Culture of Neurons and Glial Cells», dans *The Cultured Cell and Inherited Metabolic Disease*, Springer Netherlands, Dordrecht, p. 69–76, doi :10.1007/978-94-011-6627-0_6. [222](#)
- TIMOFEEV, I. 2000, «Origin of Slow Cortical Oscillations in Deafferented Cortical Slabs», *Cerebral Cortex*, vol. 10, n° 12, doi :10.1093/cercor/10.12.1185, p. 1185–1199, ISSN 14602199. [233](#)

- TLUSTY, T. et J.-P. ECKMANN. 2009, «Remarks on bootstrap percolation in metric networks», *Journal of Physics A : Mathematical and Theoretical*, vol. 42, n° 20, doi :10.1088/1751-8113/42/20/205004, p. 205 004, ISSN 1751-8113. 236, 238
- TOMBA, C., C. BRAINI, B. WU, N. S. GOV et C. VILLARD. 2014, «Tuning the adhesive geometry of neurons : length and polarity control», *Soft Matter*, vol. 10, n° 14, doi :10.1039/c3sm52342j, p. 2381, ISSN 1744-683X. 226
- TORBEN-NIELSEN, B. «NeuroMaC : Neuronal Morphologies & Circuits», URL <https://github.com/btorb/neuromac>. 266
- TORBEN-NIELSEN, B. et E. DE SCHUTTER. 2014, «Context-aware modeling of neuronal morphologies.», *Frontiers in Neuroanatomy*, vol. 8, n° September, doi :10.3389/fnana.2014.00092, p. 92, ISSN 1662-5129. 264, 266
- TORRE, V. et Y. CHEN. 2018, «Neuroscaffolds», URL <http://www.neuroscaffolds.eu/site/>. 228
- TOUBOUL, J. et A. DESTEXHE. 2010, «Can power-law scaling and neuronal avalanches arise from stochastic dynamics?», *PLoS ONE*, vol. 5, n° 2, doi :10.1371/journal.pone.0008982, p. e8982, ISSN 19326203. 254
- TURING, A. 1948, «Intelligent Machinery», cahier de recherche, National Physical Laboratory. URL <http://www.npl.co.uk/about/history/notable-individuals/turing/intelligent-machinery>. 255
- ULLOA SEVERINO, F. P., J. BAN, Q. SONG et collaborateurs. 2016, «The role of dimensionality in neuronal network dynamics», *Scientific Reports*, vol. 6, n° 1, doi :10.1038/srep29640, p. 29 640, ISSN 2045-2322. 226
- VAN KESSENICH, M. L., L. DE ARCANGELIS et H. J. HERRMANN. 2016, «Synaptic plasticity and neuronal refractory time cause scaling behaviour of neuronal avalanches», *Scientific Reports*, vol. 6, doi :10.1038/srep32071, p. 32 071, ISSN 20452322. 252
- VAN OOYEN, A. et A. V. OOYEN. 2011, «Using theoretical models to analyse neural development», *Nature Reviews Neuroscience*, vol. 12, n° 6, doi :10.1038/nrn3031, p. 311–326, ISSN 1471003X. 264
- VAN PELT, J., I. VAJDA, P. S. WOLTERS, M. A. CORNER et G. J. RAMAKERS. 2004a, «Dynamics and plasticity in developing neuronal networks in vitro», *Progress in Brain Research*, vol. 147, n° SPEC. ISS., doi :10.1016/S0079-6123(04)47013-7, p. 173–188, ISSN 00796123. 249
- VAN PELT, J., P. S. WOLTERS, M. A. CORNER et collaborateurs. 2004b, «Long-term characterization of firing dynamics of spontaneous bursts in cultured neural networks», *IEEE Transactions on Biomedical Engineering*, vol. 51, n° 11, doi :10.1109/TBME.2004.827936, p. 86–89, ISSN 03043940. 224, 239, 247, 249
- VAN STAVEREN, G. W., J. R. BUITENWEG, E. MARANI et W. L. C. RUTTEN. 2005, «The effect of training of cultured neuronal networks, can they learn?», dans *2nd International IEEE EMBS Conference on Neural Engineering*, vol. 2005, ISBN 0-7803-8710-4, p. 328–331, doi :10.1109/CNE.2005.1419624. 260
- VAN VREESWIJK, C. et D. HANSEL. 2001, «Patterns of synchrony in neural networks with spike adaptation», *Neural Computation*, vol. 13, n° 5, doi :10.1162/08997660151134280, p. 959–992, ISSN 0899-7667. 238

- VIANA, R. L., F. S. BORGES, K. C. IAROSZ et collaborateurs. 2014, «Dynamic range in a neuron network with electrical and chemical synapses», *Communications in Nonlinear Science and Numerical Simulation*, vol. 19, n° 1, doi :10.1016/j.cnsns.2013.06.003, p. 164–172, ISSN 10075704. [267](#)
- VIGNES, M. 2013, *Development and of Activity of In Vitro Neuronal Networks*, thèse de doctorat, Université Paris Descartes. [224](#), [225](#)
- VOGELS, T. P. et L. F. ABBOTT. 2005, «Signal Propagation and Logic Gating in Networks of Integrate-and-Fire Neurons», *Journal of Neuroscience*, vol. 25, n° 46, doi :10.1523/JNEUROSCI.3508-05.2005, p. 10 786–10 795, ISSN 0270-6474. [242](#), [243](#)
- WAGENAAR, D. A. 2005, «Controlling Bursting in Cortical Cultures with Closed-Loop Multi-Electrode Stimulation», *Journal of Neuroscience*, vol. 25, n° 3, doi :10.1523/JNEUROSCI.4209-04.2005, p. 680–688, ISSN 0270-6474. [233](#), [248](#), [249](#)
- WAGENAAR, D. A., J. PINE et S. M. POTTER. 2006a, «An extremely rich repertoire of bursting patterns during the development of cortical cultures», *BMC Neuroscience*, vol. 7, n° 1, doi : 10.1186/1471-2202-7-11, p. 11, ISSN 14712202. [233](#), [239](#), [247](#), [248](#), [249](#)
- WAGENAAR, D. A., J. PINE et S. M. POTTER. 2006b, «Searching for plasticity in dissociated cortical cultures on multi-electrode arrays.», *Journal of negative results in biomedicine*, vol. 5, doi :10.1186/1477-5751-5-16, p. 16, ISSN 1477-5751. [256](#)
- WAGENAAR, D. A., J. PINE et S. M. POTTER. 2006c, «Searching for plasticity in dissociated cortical cultures on multi-electrode arrays», *Journal of Negative Results in BioMedicine*, vol. 5, n° 1, doi : 10.1186/1477-5751-5-16, p. 16, ISSN 14775751. [259](#)
- WANG, H. 2018, «Modeling Neurological Diseases With Human Brain Organoids», *Frontiers in Synaptic Neuroscience*, vol. 10, doi :10.3389/fnsyn.2018.00015, p. 15, ISSN 1663-3563. [228](#)
- WANG, J., L. REN, L. LI et collaborateurs. 2009, «Microfluidics : A new cosset for neurobiology», *Lab Chip*, vol. 9, n° 5, doi :10.1039/B813495B, p. 644–652, ISSN 1473-0197. [226](#)
- WEARNE, S. L., A. RODRIGUEZ, D. B. EHLENBERGER et collaborateurs. 2005, «New techniques for imaging, digitization and analysis of three-dimensional neural morphology on multiple scales», *Neuroscience*, vol. 136, n° 3, doi :10.1016/j.neuroscience.2005.05.053, p. 661–680, ISSN 03064522. [233](#)
- WHEELER, B. C. et G. J. BREWER. 2010, «Designing neural networks in culture», *Proceedings of the IEEE*, vol. 98, n° 3, doi :10.1109/JPROC.2009.2039029, p. 398–406, ISSN 00189219. [226](#)
- WHEELER, B. C., Y. NAM et B. C. WHEELER. 2011, «In vitro microelectrode array technology and neural recordings.», *Critical Reviews™ in Biomedical Engineering*, vol. 39, n° 1, doi :10.1615/CritRevBiomedEng.v39.i1.40, p. 45–61, ISSN 0278-940X. [228](#)
- YAMADA, A., M. VIGNES, C. BUREAU et collaborateurs. 2016, «In-mold patterning and actionable axo-somatic compartmentalization for on-chip neuron culture», *Lab Chip*, vol. 16, n° 11, doi : 10.1039/C6LC00414H, p. 2059–2068, ISSN 1473-0197. [226](#), [239](#)
- YAN, Y., L. SONG, J. BEJOY et collaborateurs. 2018, «Modeling Neurodegenerative Microenvironment Using Cortical Organoids Derived from Human Stem Cells.», *Tissue engineering. Part A*, doi :10.1089/ten.TEA.2017.0423, ISSN 1937-335X. [228](#)
- YIZHAR, O., L. E. FENNO, M. PRIGGE et collaborateurs. 2011, «Neocortical excitation/inhibition balance in information processing and social dysfunction», *Nature*, vol. 477, n° 7363, doi :10.1038/nature10360, p. 171–178, ISSN 0028-0836. [254](#)

- ZAPPERI, S., K. B. LAURITSEN et H. E. STANLEY. 1995, «Self-organized branching processes : Mean-field theory for avalanches», *Physical Review Letters*, vol. 75, n° 22, doi :10.1103/PhysRevLett.75.4071, p. 4071–4074, ISSN 00319007. [252](#)
- ZBINDEN, C. 2011, «Leader neurons in leaky integrate and fire neural network simulations», *Journal of Computational Neuroscience*, vol. 31, n° 2, doi :10.1007/s10827-010-0308-6, p. 285–304, ISSN 09295313. [236](#), [265](#)
- ZHOU, S. et Y. YU. 2018, «Synaptic E-I Balance Underlies Efficient Neural Coding», *Frontiers in Neuroscience*, vol. 12, doi :10.3389/fnins.2018.00046, p. 46, ISSN 1662-453X. [254](#)
- ZUBLER, F. 2009, «A framework for modeling the growth and development of neurons and networks», *Frontiers in Computational Neuroscience*, vol. 3, doi :10.3389/neuro.10.025.2009, p. 25, ISSN 16625188. [264](#), [265](#)
- ZUBLER, F., A. HAURI, S. PFISTER et collaborateurs. 2013, «Simulating Cortical Development as a Self Constructing Process : A Novel Multi-Scale Approach Combining Molecular and Physical Aspects», *PLoS Computational Biology*, vol. 9, n° 8, doi :10.1371/journal.pcbi.1003173, p. e1003 173, ISSN 1553734X. [265](#)

Chapitre 7

Neurophysique des cultures neuronales : contributions

*« L'incendio suo seguiva ogni scintilla;
ed eran tante, che 'l numero loro
più che 'l doppiar delli scacchi s'immilla »^a*

Leur embrasement se reproduisait en chaque étincelle,
et tant elles étaient, que leur nombre
en mille surpasse le doubler des échecs
[Traduction Félicité Robert de Lamennais, Flammarion
(1910)]

Dante Alighieri, Paradiso XXVIII, 91-93

^a. Allusion de Dante au problème de l'échiquier de Sissa et la croissance d'une suite géométrique de raisons 2 partant de l'unité.

Sommaire

7.1 Introduction au chapitre	288
7.1.1 Cultures de neurones en bref	288
7.2 Modèles de percolation	289
7.2.1 Motivation et contexte	289
7.2.2 Étude précise du point critique de la percolation à quorum	292
7.2.3 Percolation à quorum avec dégradation	295
7.2.4 Percolation à quorum avec désordre d'excitabilité	298
7.2.5 Percolation à quorum et inhibition	300
7.2.6 Percolation à quorum dans des modèles dynamiques	302
7.2.7 Contribution au contrôle et à la mesure expérimentaux des cultures	306
7.2.8 Résumé et perspective « Études de la stimulation et de la percolation dans les cultures »	308
7.3 Bouffées synchronisées de potentiels d'actions (Network bursting)	308
7.3.1 Motivation et contexte	308
7.3.2 Résumé des résultats	310
7.3.3 Synapses continues	315
7.3.4 Bouffées de réseau : conclusions et perspectives	315
7.4 Développement neuronal et croissance de réseau	319
7.4.1 Modèle empirique de Renaud Renault	321
7.4.2 Modèle de croissance étendu et simulateur	324
7.4.3 Évaluation et extensions futures	331
7.5 Conclusion du chapitre	333
7.6 Références	335

7.1 Introduction au chapitre

Les travaux présentés dans ce chapitre concernent un axe que j'ai démarré en 2012 suite à ma rencontre avec Elisha Moses de l'Institut Weizmann, et ma découverte de ses recherches relatives aux conditions nécessaires pour l'émergence de propriétés calculatoires dans des populations de neurones et le développement de processeurs neuronaux. Des recherches basées sur les nouvelles possibilités discutées au chapitre précédent pour réaliser et manipuler des cultures *in vitro* de neurones. Ces systèmes mettent en œuvres des réseaux complexes de neurones qui se développent spontanément et font apparaître des comportements collectifs spontanés de différentes natures. Ces objets combinaient donc les propriétés de réseaux complexes biologiques, sur lesquelles j'avais travaillé dans le contexte génomique, et des intérêts beaucoup plus anciens datant mon doctorat durant lequel je m'étais occupé de phénomènes de synchronisation collective et de criticalité auto-organisée. Progressivement avec le renfort de Pascal Monceau et ensuite de Stéphane Métens au laboratoire Matières et Systèmes Complexes, qui eux aussi ont été captivés par la problématique, nous avons démarré au laboratoire MSC une thématique *Neurophysique des cultures de neurones*, tirant fortement parti de collaborations avec les groupe expérimentaux d'Elisha Moses ainsi que de Jean-Louis Viovy à l'Institut Curie. Depuis nous travaillons aussi régulièrement avec Catherine Villard qui a rejoint l'équipe de l'Institut Curie et Jordi Soriano de l'université de Barcelone. Ce chapitre rassemble mes différents travaux, essentiellement théoriques sur les thèmes suivants qui feront l'objet des différentes sections plus bas :

- l'extension de modèles de percolation pour la propagation d'une vague d'activité dans une culture;
- la recherche d'une explication au comportement de bouffées d'activité synchronisées générales dans les cultures;
- la modélisation biophysique et la simulation du développement de neurones en culture et la croissance du réseau.

Auparavant, pour un accès direct possible à ce chapitre sur mes travaux, je résume ci-dessous brièvement les éléments principaux concernant les cultures de neurones déjà discutés en détail dans le chapitre précédent 6. Les lecteurs qui en ont déjà pris connaissance pourront passer directement à la section suivante 7.1.7.2.

7.1.1 Cultures de neurones en bref

Grâce aux possibilités de construction d'environnements expérimentaux artificiels, les cultures de neurones constituent un outil particulièrement puissant pour explorer des questions neurobiologiques avec une approche physique. De telles cultures sont préparées typiquement à partir de neurones isolés extraits de certains tissus biologiques et incubés sur des substrats et des milieux nutritifs compatibles. Avec des conditions adéquates les neurones se développent, établissent des connexions synaptiques et forment après quelques semaines un réseau électriquement actif. Les cultures présentent une activité spontanée dominée par de courtes séquences de déclenchements quasiment synchronisés sur toute la population et espacées de plus longues périodes de repos. De nombreux facteurs déterminent ce comportement collectif particulier dont les traits sont très variés et complexes en dépit de ses caractéristiques générales communes. Les neurones des cultures semblent avoir les mêmes propriétés individuelles que leurs contreparties *in vivo* permettant de les utiliser comme modèles d'études de pathologies. Les cultures fournissent aussi un système de référence pour comparer des comportements de réseaux neuronaux entre ceux des structures ordonnées cérébrales biologiques et les constructions artificielles désordonnées obtenues *in vitro*. Enfin, ces systèmes offrent la possibilité d'explorer avec de la matière biologique excitable active, la construction de dispositifs artificiels biologiques dotés de propriétés complexes,

peut-être d'apprentissage et traitement d'information. Il s'agit d'explorer par ce moyen quelles sont les conditions nécessaires à des capacités cognitives minimales et leurs propriétés.

L'essor des techniques de micro-fabrication de la microfluidiques permettent de concevoir des milieux de cultures structurés. Ces dispositifs ouvrent la voie à des modèles expérimentaux sophistiqués pour l'étude de pathologies neuronales mais aussi d'envisager l'exploration à l'aide d'approches originales inspirées par la physique, l'émergence de fonctions cognitives élémentaires dans des populations neuronales.

Les travaux présentés dans les sections ci-dessous concernent en premier des modèles de propagation de l'activité dans un réseau en culture suite à une excitation initiale. Une question en lien avec l'excitation du système qui reprend et étend un modèle de percolation introduit par nos collaborateurs de l'Institut Weizmann voici une dizaine d'années. Ce thème a ensuite conduit à l'étude de bouffées d'activité dans les cultures, pour lesquelles nous avons proposé un modèle efficace permettant de formuler des hypothèses sur les mécanismes biologiques en jeu. Enfin la dernière partie de ces travaux sur les cultures concerne la modélisation du développement du réseau qui aborde des modélisations biophysiques des mécanismes de croissances des neurones.

7.2 Modèles de percolation

7.2.1 Motivation et contexte

Dans la perspective de dispositifs artificiels exploitant les neurones pour des tâches de traitement d'information on peut penser se placer à l'échelle des neurones individuels et viser la construction de systèmes, de réseaux, déterminés au niveau des connexions entre neurones individuels. Cette approche est envisageable mais technologiquement encore à un stade très préliminaire, voir la discussion au chapitre précédent, section 6.2.2. La plupart des cultures neuronales réalisées jusqu'à présent sont des dispositifs impliquant une ou plusieurs populations de neurones. L'unité de base pour un processeur à base de neurones est donc plutôt à l'échelle d'une population, et le dispositif consisterait en plusieurs populations en communication. C'est l'activité de chaque sous-population qui serait le signal pour un codage d'information.

Dans cette optique, il est essentiel d'apprendre à contrôler le comportement de chaque population. Comme on l'a vu précédemment le comportement le plus caractéristique des populations de neurones en cultures est de produire des bouffées soudaines d'activité neuronale synchronisée, macroscopique, impliquant l'ensemble des neurones. Son ubiquité dans les cultures en fait un élément essentiel à comprendre afin de le contrôler et si possible de l'exploiter dans des applications dans les dispositifs.

Ne pouvant pas en faire abstraction, il s'agirait d'apprendre à utiliser ces bouffées collectives d'activité comme signal pour le codage d'information. Ces événements sont hautement non linéaires; on peut leur associer à tout instant une interprétation binaire, signal ou absence de signal, exploitables pour du codage d'information. Les bouffées de réseau seraient des analogues au niveau d'une population des déclenchements de neurones individuels. Et de la même manière que ceux-ci, permettre des schémas de code d'information mais au niveau des activités des bouffées des populations. Pour apprendre à utiliser les cultures dans des dispositifs contrôlables il est donc nécessaire de comprendre le fonctionnement des bouffées et la dépendance de leurs propriétés en fonction des paramètres expérimentaux manipulables. Les travaux théoriques présentés dans cette section visent à contribuer à cet objectif.

Plus spécifiquement je me suis intéressé à étendre l'étude du modèle de percolation à quorum introduit par Elisha Moses et ses collègues afin de décrire la propagation d'une vague d'activation dans une culture suite à une stimulation extérieure. Cette étape initiale d'action sur une culture était particulièrement intéressante pour aborder l'étude du contrôle d'une culture. Il s'agit naturellement en premier lieu de la première action qu'il faut savoir faire pour agir sur un dispositif :

savoir le stimuler en entrée et connaître la réponse immédiate d'une population. Pour le développement de dispositifs neuronaux il convient de caractériser et comprendre cette réponse et ne pas se contenter de savoir techniquement déclencher une bouffée de toute la population avec un stimulus très puissant. Il est ainsi intéressant de déterminer justement quel est le seuil de la puissance du stimulus nécessaire pour que le système réponde. Et savoir aussi si un tel seuil de déclenchement soudain pour un stimulus suffisant existe toujours, ou dans quelles conditions c'est le cas. Cette caractérisation était importante, elle était aussi possible expérimentalement grâce à l'imagerie calcique qui permet la mesure de l'activité de toute une population directement au microscope par fluorescence. En entrée un pulse de champ électrique appliqué au milieu de culture stimule la dépolarisation d'une fraction des neurones présents proportionnellement à l'intensité du champ. Avec un tel dispositif Elisha Moses et ses collègues ont pu mesurer des courbes de réponse de cultures en fonction de l'intensité du stimulus (voir figure 7.1 et [BRESKIN et collab., 2006]). Finalement, outre son importance et sa réalisation expérimentale cette caractérisation a également eu ceci d'intéressant d'avoir pu être modélisée théoriquement sous une forme permettant l'interprétation des expériences et des prédictions.

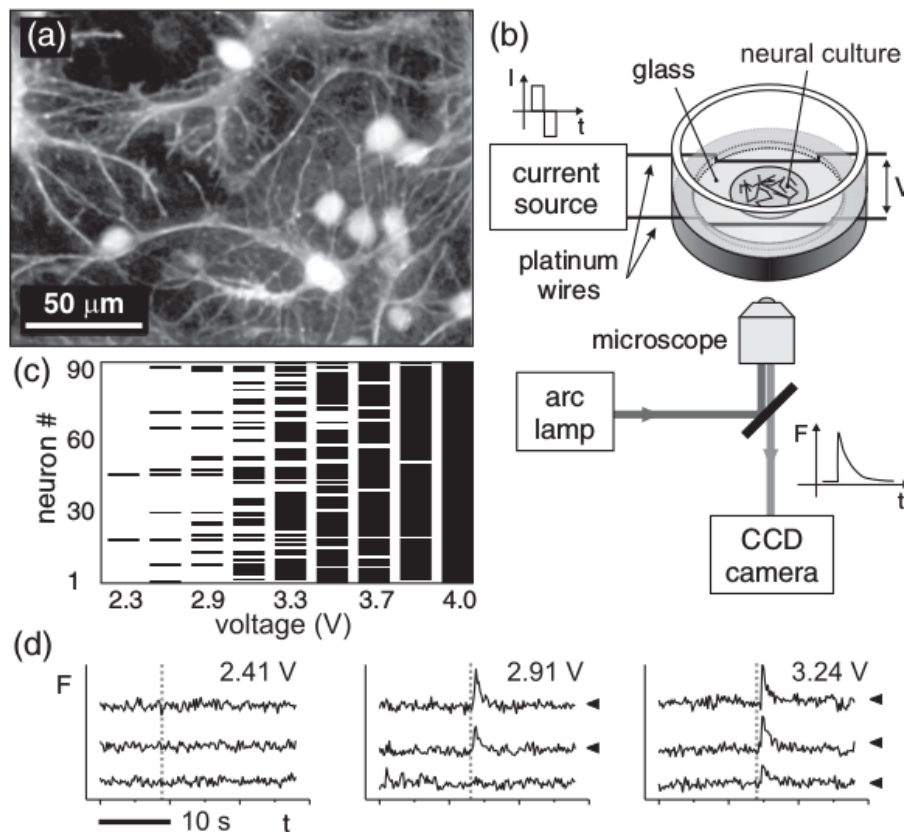


FIGURE 7.1 – Dispositif expérimental de l'équipe d'Elisha Moses pour la mesure de la réponse d'une culture à des stimuli d'intensité croissante. (a) Image de fluorescence d'une petite région de la culture neuronale, les tâches claires sont les corps cellulaires, les connexions neuronales, essentiellement des dendrites sont visibles. (b) Schéma du dispositif expérimental. La stimulation est produite par une impulsion bipolaire de courant électrique. (c) Traçé de la réponse neuronale, une ligne noire indique les neurones qui ont répondu au stimulus. (d) Signal de fluorescence au cours du temps pour 3 neurones à des voltage de stimulation croissants; en pointillé l'instant de stimulation; les flèches indiquent les neurones activés. Source : [BRESKIN et collab., 2006]

Le résultat expérimental remarquable de ces auteurs a été d'observer sur des cultures de neurones vivantes les courbes de transitions de phase en fonction de l'intensité du stimulus d'une part et d'autre part d'un second paramètre de contrôle que l'on peut associer à l'excitabilité des neurones individuels (dont je ne détaillerai pas ici la procédure expérimentale via l'antagoniste CNQX des récepteurs synaptiques AMPA) [BRESKIN et collab., 2006]. La surface décrivant le diagramme

de transition de phase a ceci de fascinant d'un point de vue physique de faire apparaître dans un système vivant l'existence d'un point critique et d'une ligne de transitions du premier ordre discontinues. Un comportement de transition de phase que le modèle théorique reproduit parfaitement. Un succès qui a permis d'envisager l'usage du modèle pour déduire des informations sur la connectivité du réseau neuronal subjacent (figure 7.2).

Résumé des résultats

Le modèle d'activation proposé correspond à un processus de percolation sur un graphe aléatoire dirigé avec des nœuds dotés d'un seuil pour la propagation [COHEN et collab., 2010]. Il s'agit d'un modèle de percolation avec amorçage (« bootstrap ») qui représente de manière minimale la propagation de signaux par des neurones interconnectés suivant un modèle de graphe aléatoire : les neurones localisés aux nœuds du graphe, sont des systèmes à deux états dont l'activation est déterminée par une règle à seuil, le quorum, suivant une dynamique à temps discret. Un nœud est actif, soit parce que il est placé dans cet état initialement, soit parce que le nombre de ses entrées connectées à des neurones actifs est au moins égal au quorum m .

Le modèle de percolation à quorum peut facilement être simulé par une approche Monte-Carlo consistant à générer aléatoirement des réseaux, tirer une fraction f donnée de nœuds actifs, et appliquer les règles de propagation jusqu'à la stabilisation. Ce processus peut aussi être capturé dans une seule équation auto-consistante de champ moyen : la probabilité Φ pour un neurone d'être actif une fois l'équilibre atteint correspond à la probabilité f d'être actif initialement, ou bien à la probabilité d'avoir été activé durant la propagation discrète. Étant donné que l'activation d'un neurone a lieu seulement si au moins m liens entrants viennent de voisins actifs cette probabilité est la probabilité $(1 - f)$ de ne pas avoir fait partie des neurones actifs initialement fois la probabilité pour un neurone d'être connecté à au moins m neurones actifs $\sum_{l=m}^{\infty} \binom{k}{l} \Phi^l (1 - \Phi)^{k-l}$ où p_k est la distribution des connectivités entrants, soit la probabilité pour un nœud d'avoir k liens entrants. Finalement on a l'équation auto-consistante suivante [BRESKIN et collab., 2006] :

$$\frac{\Phi - f}{1 - f} = \sum_{m=1}^{\infty} \mathcal{P}_m \sum_{k=m}^{\infty} p_k \sum_{l=m}^k \binom{k}{l} \Phi^l (1 - \Phi)^{k-l} = \sum_{k,m} \Pi(m, k) \sum_{l=m}^{\infty} \binom{k}{l} \Phi^l (1 - \Phi)^{k-l} \quad (7.1)$$

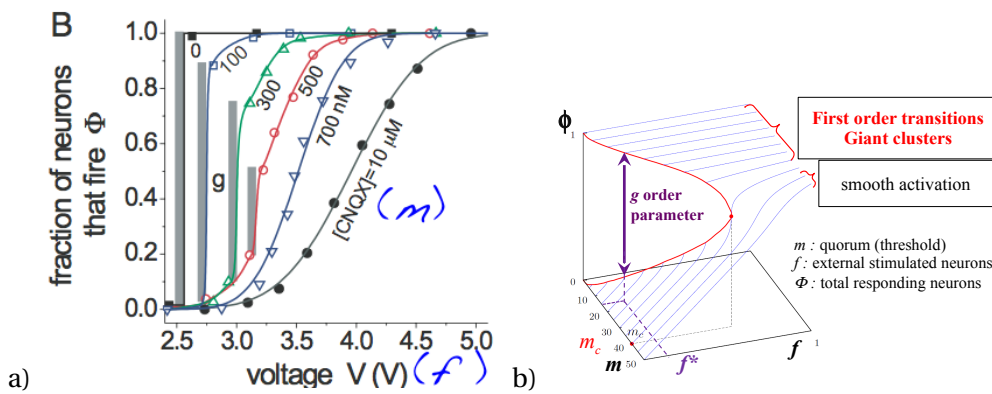


FIGURE 7.2 – **Diagramme de phase de la réponse à une stimulation.** a) Courbes de réponses expérimentales pour 6 concentrations de l'antagoniste CNQX des récepteurs AMPA, assimilées à 6 situations d'excitabilité neuronale différentes. Source : [BRESKIN et collab., 2006]. b) Diagramme de phase théorique du modèle de percolation à quorum pour un réseau avec une distribution Gaussienne des connectivités entrantes de connectivité moyenne $\bar{k} = 50$ et écart type $\sigma_k = 10$. Quand le quorum m est plus petit que la valeur critique m_c on observe un saut de la fraction des neurones activés Φ lorsqu'on augmente la fraction f des neurones initialement activés. La hauteur du saut $\langle g \rangle$ à la discontinuité est la dimension normalisée de cluster de percolation géant.

Cette équation de dérivation très simple relie les paramètres propres au réseau neuronal, que

sont la connectivité p_k et l'excitabilité m , à la fonction de réponse $\Phi(f)$ qui nous intéresse. Comme traité dans les publications commentées ci-dessous, la résolution de l'équation (7.1), nécessairement numériquement, conduit à deux régimes : pour m inférieur à une valeur critique m_c trois solutions réelles existent, dont seulement une solution basse Φ^- et une solution haute Φ^+ sont physiques (pour la percolation la réponse Φ doit être croissante avec f , ce qui n'est pas le cas autour de la solution intermédiaire). La différence $g = \Phi^+ - \Phi^-$ est la taille du cluster géant de percolation qui apparaît soudainement en augmentant f . Pour $m > m_c$ une seule solution réelle existe, et la fonction de réponse perd sa caractéristique à seuil et devient continue.

Par sa simplicité le modèle de percolation à quorum a fourni un point de départ pour l'exploration de propriétés des cultures. Le comportement de la taille du cluster géant $\langle g \rangle$ a proximité du point critique m_c est relié via (7.1) aux caractéristiques du réseau. Par ce biais, en ajustant, le comportement théorique aux observations il a été possible de déduire des informations sur la connectivité du réseau expérimental [COHEN et collab., 2010; SORIANO et collab., 2008]. Cependant le modèle original est très simplifié et ne tient pas compte de nombreuses caractéristiques biologiques des neurones telles que les conséquences sur l'activation de neurones inhibiteurs, des temps caractéristiques de la propagation et de retour au potentiel de repos des neurones, de la présence de désordre dans la population et aussi de la nature éminemment dynamique de l'activité de chaque nœud. Partant de l'idée de base de la percolation à quorum nous avons dans l'équipe au laboratoire MSC étudié plusieurs raffinements de ce modèle pour incorporer ces ingrédients et en analyser les effets.

Les articles rassemblés ici présentent donc plusieurs extensions et variations du modèle de percolation à quorum : i) une extension des valeurs continues, notamment du quorum qui a permis d'approcher un point critique réel inaccessible dans la version discrète originale du modèle; ii) l'ajout d'une dégradation du potentiel neuronal sous le seuil, correspondant aux fuites de courant ionique par la membrane; iii) l'introduction de variabilité des neurones dans la population au niveau de l'excitabilité; iv) les conséquences sur le point critique de la présence d'une fraction de neurones inhibiteurs et enfin v) une transposition du modèle de percolation à quorum dans le cadre d'une modélisation dynamique. Différents sujets introduits maintenant dans les sections suivantes.

7.2.2 Étude précise du point critique de la percolation à quorum

Les articles de l'équipe [RENAULT et collab., 2014] et [MÉTENS et collab., 2016] discutés ici, correspondent à une analyse détaillée du modèle de percolation à quorum proche du point critique afin de mieux caractériser les propriétés de la transition de phase et son éventuelle exploitation pour l'interprétation de résultats expérimentaux. On discute notamment l'apparition du cluster géant des neurones actifs en termes d'une analyse de bifurcation qui nous permet d'approcher les propriétés de la transition de comportement du système au point critique, quand les sauts d'activation disparaissent. L'analyse théorique de cette transition selon la physique statistique suggérerait un comportement universel donc indépendant de caractéristiques particulières du réseau hormis ses dimension et symétries. L'étude de cette transition, y compris expérimentale, ne servirait donc à rien pour apprendre des informations sur les particularités du réseau des neurones et de leur activation. Ceci peut sembler en contradiction avec les conclusions des articles du groupe de E. Moses [2017; BRESKIN et collab., 2006; COHEN et collab., 2010; SORIANO et collab., 2007] qui utilisent le modèle de percolation de quorum pour interpréter les données expérimentales de l'approche du point critique dans l'activation de cultures et en déduire des caractéristiques du réseau, telles que la distribution gaussienne des liens entrants, la connectivité moyenne et son écart type. En réalité, les conclusions de ces auteurs viennent de l'analyse de l'approche du point critique, et non des propriétés, universelles, de celui-ci. Or un des résultats de notre propre étude a été justement de comprendre que toute situation pratique (système à taille finie) est très vite éloignée des propriétés du point critique.

Ceci a nécessité la réalisation d'une prolongation analytique de l'équation d'auto-consistance du modèle de percolation à quorum introduite par les auteurs originaux du modèle. Elle nous a permis de caractériser précisément la transition de phase et d'étudier comment ses propriétés se comparent avec d'une part les résultats de simulations numériques et d'autre part le comportement universel prédit par la théorie de bifurcation via la forme standard « normale » pour ce type de percolation. Ces comparaisons nous ont conduit à étudier les effets de taille finie et les fluctuations du paramètre d'ordre de la transition. Il en ressort en particulier, que si le modèle de percolation à quorum est bien dans une classe d'universalité, les très grandes non linéarités des équations qui le décrivent font que le comportement critique standard ne se manifeste que dans un voisinage extrêmement étroit du point critique.

Pour des systèmes de taille finie, comme c'est le cas des cultures expérimentales, les propriétés de la transition de phase sont différentes de celles théoriques qui ne sont atteintes que pour des systèmes infinis. En effet les termes supérieurs des développements conduisant à la forme normale ne sont pas négligeables. L'approche du point critique dans les systèmes à taille finie que sont les expériences et les simulations, dépend alors de propriétés du réseau : il est possible de mesurer une valeur de contrôle critique m_c et un exposant critique « effectifs » du paramètre d'ordre et de les relier à la connectivité moyenne et son écart type. Ainsi pour le point critique on a pu caractériser que :

$$\frac{m_c}{\bar{k}} = 1 - A\left(\frac{\sigma}{\bar{k}}\right) - B\left(\frac{\sigma}{\bar{k}}\right)^2 \quad (7.2)$$

avec A et B deux coefficients dépendant légèrement de la connectivité entrante moyenne \bar{k} , σ étant la déviation standard du degré de connectivité entrante. Nous avons également déterminé numériquement les valeurs de l'exposant critique du paramètre d'ordre pour différentes valeurs de \bar{k} et σ . Ces mesures montrent que les différents réseaux ont des exposants proches mais distincts de la valeur asymptotique théorique, et ont une petite dépendance de la valeur avec les paramètres de réseaux. Les paragraphes suivants reprennent de manière un peu plus développée le résumé qui vient d'être fait et d'autres résultats de ces publications.

Une des contributions centrales de nos deux articles sur le sujet est en effet l'introduction d'une extension continue à l'équation d'auto-consistante permettant d'approcher mieux le point critique, les idées principales ont été introduites en 2014 [RENAULT et collab.] et développées plus en détails dans la publication de 2016 [MÉTENS et collab.] développant les effets de taille finie. Le modèle initial de la percolation à quorum est défini pour des états des neurones et d'un seuil nécessairement discret en raison du nombre entier de liens. Les paramètres de contrôle f et m sont donc discrets et les valeurs critiques aussi. Cependant dans un cadre discret on ne peut espérer une description précise du comportement critique. En effet suivant les notions usuelles des transitions de phase si $g(m)$ est le paramètre d'ordre caractérisant l'état du système et m un paramètre de contrôle, à l'approche du point critique $m \rightarrow m_c$, le comportement de g doit suivre une loi de puissance $g(m) \sim \left(\frac{m-m_c}{m_c}\right)^\beta$. La détermination de ce comportement et de l'exposant critique β demande de s'approcher d'une valeur précise de m_c . Comme le modèle original de percolation à quorum est discret, m_c ne peut avoir une précision meilleure d'une unité. Une manière pour améliorer la situation serait d'utiliser de grandes valeurs de la connectivité moyenne k pour augmenter la précision relative sur m_c . Cependant ceci conduit à des difficultés pratiques pour résoudre l'équation d'auto-consistante dans ce cas qui comprendrait des coefficients binomiaux très élevés qui ne peuvent être traités numériquement. Pour cette raison nous avons développé une prolongation continue du modèle de percolation à quorum à des valeurs de m réelles. Comme discuté dans nos deux articles ceci peut se faire relativement simplement en tirant parti des propriétés de la fonction mathématique β donnée par $B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt$ et la fonction β incomplète $B_\phi(x, y) = \int_0^\phi t^{x-1}(1-t)^{y-1} dt$ pour lesquelles existe pour m entier la relation remarquable :

$$\sum_{l=m}^k \binom{k}{l} \phi^l (1-\phi)^{k-l} = \frac{B_\phi(m, k-m+1)}{B(m, k-m+1)}.$$

Étant donné que $\frac{B_\phi(m, k-m+1)}{B(m, k-m+1)}$ est calculable quel que soit $m \in \mathbb{R}^+$ la propriété précédente permet d'obtenir la prolongation réelle recherchée de l'équation auto-consistante :

$$\begin{aligned} \Delta(f, m, k, \sigma) = \phi &= f + (1-f) \int_m^\infty \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(k-\bar{k})^2}{2\sigma^2} \frac{B_\phi(m, k-m+1)}{B(m, k-m+1)}\right) dk - \phi \\ &= f + (1-f)F[\phi; m] - \phi = 0 \end{aligned} \quad (7.3)$$

Comme avec l'équation initiale discrète, il existe une valeur de m notée m_c , telle que pour $m < m_c$, le terme de gauche de l'équation auto-continue admet trois régimes en fonction de f (Figure 7.3) : il existe une valeur $f^*(m)$, telle que pour $f < f^*(m)$ l'équation (7.3) admet trois solutions ϕ_- , ϕ_+ stables et une intermédiaire instable ϕ_{instable} ; en stimulant une fraction $f < f^*(m)$ de neurones le système répondra avec la réponse ϕ_- stable basse. Quand f augmente vers m_c les solutions basses et instable se rapprochent pour disparaître à $f = f^*(m)$. À ce point le système saute de la réponse basse ϕ_- qui disparaît vers la solution haute ϕ_+ qui subsiste. Ce saut est associé à l'apparition d'un cluster géant. Pour $f > f^*(m)$ seule la solution haute persiste. La différence $g(m) = \phi_+ - \phi_-$ pour $f = f^*(m)$ correspond à un paramètre d'ordre, car pour $m \rightarrow m_c$, la hauteur du saut diminue et $g(m)$ devient nul au point dit critique qui correspond à un triplet ($f_c = f(m_c), m_c, \phi_c$). Pour $m > m_c$ l'équation (7.3) n'admet qu'une solution quel que soit f , qui prolonge la valeur haute ϕ_+ du régime précédent.

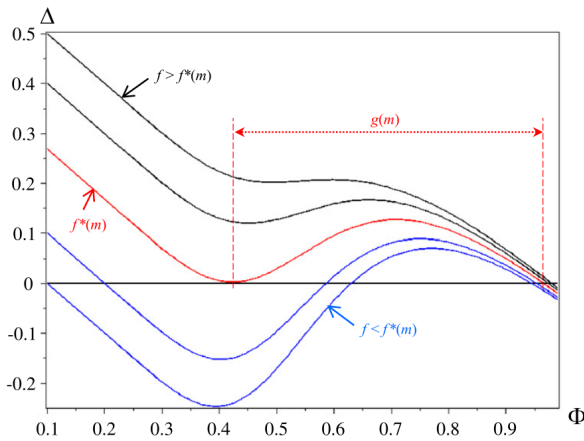


FIGURE 7.3 – ÉVOLUTION DE LA FRACTION DE NEURONES ACTIVÉS DE L'ÉQUATION AUTO-CONSISTANTE. Évolution de la fonction $\Delta(\phi)$ dans l'équation auto-consistante pour $m=30$, $\bar{k} = 50$ et $\sigma = 12$ et 5 valeurs de f . En dessous de $f(m) \simeq 0,368$ (courbes bleues) l'équation auto-consistante a trois solutions ϕ (quand $\Delta = 0$), celle centrale est instable. La solution basse et celle instable se rapprochent pour f croissant et fusionnent à $f^*(m)$ (courbe noire). Pour $f > f^*(m)$ seule la solution stable haute subsiste. Au point critique $m = m_c$ (n'est pas sur la figure), l'équation auto-consistante n'a qu'une solution stable.

L'équation (7.3) permet donc de définir précisément le point critique de la transition de phase de percolation. On peut réaliser une analyse de bifurcation de cette équation et obtenir une forme normale associée à la transition en développant F en série autour du point critique (ϕ_c, m_c). La forme normale correspond dans ce cas à une bifurcation de type pitchfork qui conduit au résultat classique d'un exposant critique $\beta = 1/2$. Cependant la fonction $F(\phi, m)$ qui apparaît dans la prolongation continue du modèle de percolation à quorum est bien davantage non linéaire que la forme normale. Ces non-linéarités qui sont associées à des coefficients binomiaux élevés conduisent à des différences entre la forme normale et le modèle de champ moyen de la percolation à quorum même proche du point critique. La caractérisation précise des différences entre la forme normale de la transition, le champ moyen (7.3) et les simulations Monte-Carlo est l'objet central de notre article de 2016 [MÉTENS et collab., 2016] (Figure 7.4).

En conséquence on peut attendre un comportement critique effectif non universel de ce modèle et on peut espérer pouvoir le relier à des observables expérimentales.

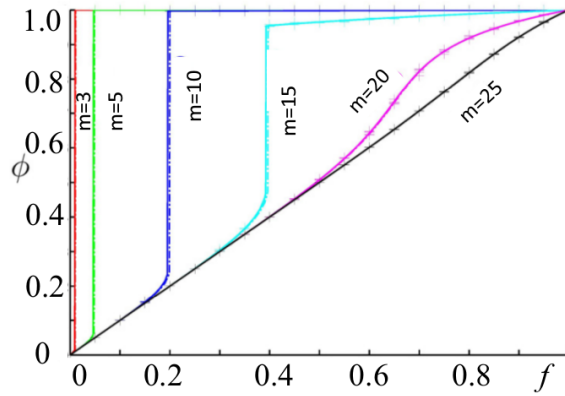


FIGURE 7.4 – **Comparaison entre les solutions Monte Carlo l'équation auto-consistante.** Simulations Monte Carlo directes du modèle de percolation à quorum (croix), solutions de l'équation auto-consistante de champ moyen (lignes) 7.3 pour $k = 25, \sigma = 5, m = \{3, 5, 15, 25\}$. (Source [MÉTENS et collab., 2016]).

7.2.3 Percolation à quorum avec dégradation

La membrane cellulaire des neurones ne se comporte pas comme un condensateur électrique parfait. Des canaux ioniques permettent des échanges entre les milieux intra- et extra-cellulaires. Certains canaux sont actifs comme la pompe potassium/sodium qui assurent en permanence une différence de concentration de ces ions de part et d'autre de la membrane. Un changement du potentiel de membrane par rapport à sa valeur de repos correspond à une déstabilisation de la répartition des charges qui évoluera par le biais des canaux actifs ou passifs vers une restauration de la séparation de charges de repos. Notre publication « Memory decay and loss of criticality in quorum percolation. » [RENAULT et collab., 2013] étudie une extension du modèle de la percolation à quorum qui prend en compte une dégradation des signaux sur chaque nœud avec le temps.

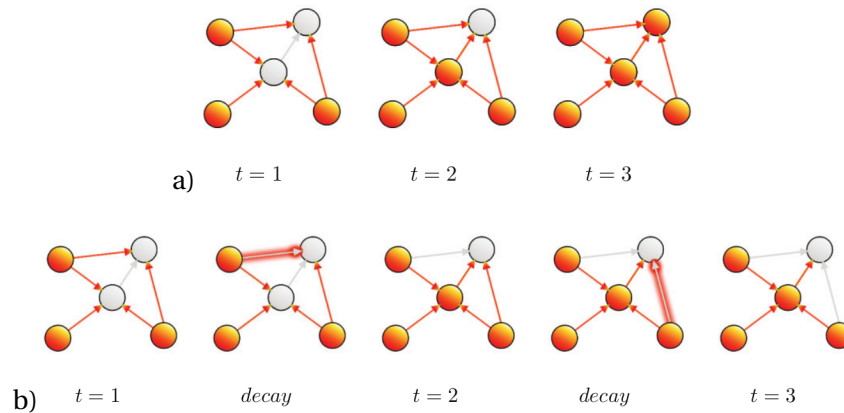


FIGURE 7.5 – **Illustration du processus de percolation à quorum.** Séquence successive montrant comment un neurone inactif (gris) est activé (orange) dès que ses entrées actives dépassent le seuil ici $m=3$. a) Sans dégradation du signal. b) Avec dégradation du signal; la dégradation aléatoire (liens brillants) d'entrées lors d'une étape change l'état final, en dépit d'une situation initiale identique à celle de a).

Dans le modèle de percolation à quorum avec décroissance que nous introduisons (figure 7.5) dans cet article on tient compte pour un neurone excité, mais encore sous le seuil de déclenchement, de la décroissance du potentiel membranaire vers son niveau de repos. Le temps caractéristique pour la propagation des signaux synaptiques au sein du réseau n'étant pas négligeable par rapport au temps caractéristique de décroissance du potentiel, ce dernier effet ne peut être négligé.

Le modèle étendu de percolation à quorum introduit ici incorpore une dégradation des signaux discrets accumulés par un neurone durant la cascade de propagation. Chaque contribution discrète peut se dégrader indépendamment des autres avec une probabilité $d \in [0, 1]$ à chaque étape du processus de percolation. Se rapportant à une estimation de la vitesse de propagation des signaux dans les cultures (~ 1 ms), et en utilisant des déterminations de la constante de temps

caractéristique τ de la membrane neuronale [ISOKAWA, 1997] entre 7 et 120 ms, on peut déterminer une probabilité non négligeable d entre 0,1 et 0,01, nécessaire dans le cadre de la percolation à quorum appliquée à des neurones en culture.

Il est très facile d'implémenter la dégradation dans des simulations Monte Carlo de la percolation à quorum. Cependant, contrairement au cas sans dégradation (7.3) il n'est pas possible d'écrire un modèle champ moyen de l'activation ϕ sous forme d'une équation auto-consistante. Dans un cadre de champ moyen on a pu cependant établir une relation récursive pour écrire la *progression* de l'avalanche d'activation suite à la stimulation d'une fraction initiale f des neurones. On a ainsi établi une relation pour calculer avec le temps discret t de chaque étape de la percolation, la fraction U_t des neurones encore inactifs dans le réseau. Cette fraction converge vers une limite correspondant à la fraction des neurones finalement activés $1 - \phi = \lim_{\infty} U_t$. Le calcul de U_t est itératif à partir de U_{t-1} en partant de $U_0 = 1 - f$. La relation entre U_t et les valeurs précédentes s'obtient par la distribution des fractions $p_{k,s}(t)$ de groupes de neurones non encore actifs de probabilités équivalente pour évoluer, où k est le nombre de liens entrants et s l'état (potentiel). L'évolution de l'état d'un neurone durant la propagation peut avoir lieu soit i) par la réception d'un signal en entrée, soit ii) par dégradation de leur état.

En effet, à tout moment t , l'évolution future d'un neurone est déterminée uniquement par son état s à cet instant, et son nombre de liens entrants non encore activés : l'historique des signaux entrants précédents et des dégradations que ce neurone a pu déjà subir n'a aucune conséquence sur la suite du processus. Avec ces éléments il est possible d'établir des relations entre les fractions $p_{k,s}(t+1)$ et $p_{k,s}(t)$ de cette partition des neurones et d'en déduire U_t .

Pour établir cette relation on distingue dans la progression du processus une première phase de dégradation de l'état (entre t et $t + t_d$) suivie d'une phase de réception des signaux suite à l'activation de liens entrants (entre t_d et $t + 1$). Pour la dégradation on aura donc

$$p_{k,s}(t-d) = \sum_{j=0}^{m-1-s} p_{k+j,s+j}(t) \binom{s+j}{j} d^j (1-d)^s \quad (7.4)$$

Et pour la réception des signaux, on remarque que les neurones caractérisés par les couples $(k, s-i)$ à l'instant t_d qui reçoivent i signaux vont se retrouver dans la fraction $p_{k,s}(t)$. Cette dernière fraction s'obtient par la somme de toutes les contributions :

$$p_{k,s}(t) = \sum_{i=0}^s p_{k,-i}(t_d) \binom{k-(s-i)}{i} \left(1 - \frac{U_t}{U_{t-1}}\right)^i \left(\frac{U_t}{U_{t-1}}\right)^{k-s} \quad (7.5)$$

où $\frac{U_t}{U_{t-1}}$ représente la probabilité qu'un neurone connecté à un des liens entrants inutilisés ait été activé exactement à l'étape t et en conséquence transfère un signal à l'étape courante (entre t et $t+1$).

Les équations (7.4) et (7.5) se combinent dans une longue expression qui permet le calcul itératif des fractions $p_{k,s}(t)$ et par sommation des groupes avec $s > m$ de $U_t(t)$; les fractions initiales ($s = 0$ dans tout le réseau) correspondant à la distribution de connectivité p_k du réseau $p_{k,0}(0) = p_k$. On ne peut en général exprimer la limite infinie $\lim_{t \rightarrow \infty} p_{k,s}(t)$ mais on peut calculer avec un critère de terminaison ($|U_{t+1} - U_t| < \nu \ll 1$) une estimation de cette valeur. À taille finie N le critère le plus strict sera $\nu = 1/N$. On peut calculer la suite précédente seulement dans le cas $d = 0$ et retrouver l'équation auto-consistante (7.3), ce qui valide la démarche.

Les deux méthodes Monte-Carlo et suite du champ moyen sont ensuite utilisées pour calculer la réponse ϕ d'un réseau à des stimuli f croissants en fonction du quorum m et de la décroissance d . Il s'agit d'une étude numérique qui sera notamment sensible à une résolution ϵ arbitrairement choisie pour déterminer si deux valeurs sont distinctes ou pas. Cette attention à la précision numérique est importante car on s'intéresse à des discontinuités de la fonction $\phi_{m,d}(f)$ pour déterminer l'existence d'un point critique m_c , via la fonction $g_\epsilon(m) = \max_f (\phi_{m,d}(f + \epsilon) - \phi_{m,d}(f))$. En

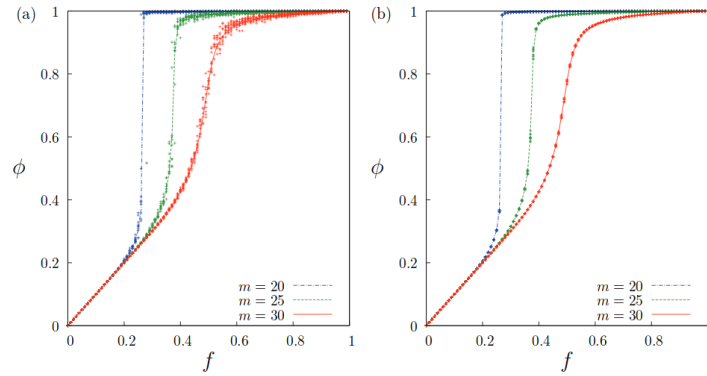


FIGURE 7.6 – **Comparaison de la simulation explicite Monte carlo et du calcul à champ moyen.** Simulation explicite Monte carlo (points, calcul à champ moyen (lignes) pour a) $N=1000$ et b) $N=100\,000$ sur un réseau Gaussien de connectivité moyenne $\bar{k} = 50$, écart type $\sigma = 10$ et dégradation $d = 0, 1$. Source : [RENAULT et collab., 2013]

présence d'une discontinuité pour $m < m_c$ on aura une valeur finie de $g_\epsilon(m)$ correspondant au saut de la transition du premier ordre, alors que pour $m \geq m_c$ la fonction $g_\epsilon(m)$ sera nulle car ϕ est continue. Si analytiquement on définit la continuité dans la limite $\epsilon \rightarrow 0$, numériquement on ne pourra vérifier que la condition $g_\epsilon(m) \geq \epsilon$.

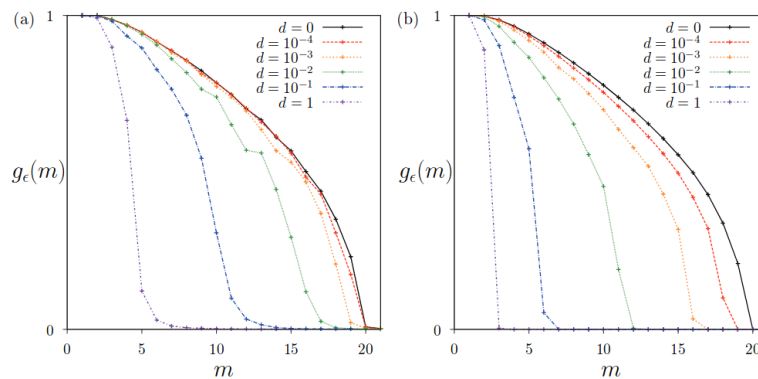


FIGURE 7.7 – **Modèle de percolation à quorum : approche du point critique.** Taille apparente de la discontinuité $g(m)$ en fonction de la dégradation d pour deux valeurs de la précision numérique a) $\epsilon = 10^{-3}$ et b) $\epsilon = 10^{-10}$, sur un réseau Gaussien de connectivité moyenne $\bar{k} = 25$, écart type $\sigma = 5$ et critère de terminaison du calcul champ moyen $\nu = 10^{-10}$. Source : [RENAULT et collab., 2013]

Une première conclusion de l'effet d'une dégradation est bien visible sur les graphes de la figure 7.7) est la constatation que son introduction réduit la taille de la discontinuité et décale la position du point critique m_c . D'autre part, en comparant les deux graphes a) et b) pour deux valeurs différentes de la résolution numérique ϵ , cette figure montre également l'effet de ce dernier critère : pour une résolution plus petite les courbes $d > 0$ de la figure 7.7 b) sont décalées vers des plus petites valeurs de m . Cette sensibilité nous a préoccupé car elle impacte la détermination du point critique m_c et nous a amené à l'étude plus détaillée de la figure 7.8 établie avec une librairie de calcul à précision arbitraire jusqu'à une résolution très petite $\epsilon = 10^{-40}$. En calculant l'évolution de $g_\epsilon(m=2)$ (pour $m=2$ la discontinuité est maximale), ce qu'on interprète comme une très forte discontinuité à faible résolution, disparaît complètement si on la recherche avec une très haute résolution $\epsilon \sim 10^{-40}$.

De manière très étonnante, l'incorporation dans le modèle de percolation à quorum d'un temps de décroissance même infiniment petit détruit la discontinuité de la fonction de réponse observée sans aucune décroissance! Cette conclusion semble paradoxale étant donné l'observation d'une fonction $\Phi(f)$ discontinue qu'interdirait la théorie étant donné l'existence dans les situations expérimentales de la décroissance du potentiel. Toutefois nous avons résolu ce problème

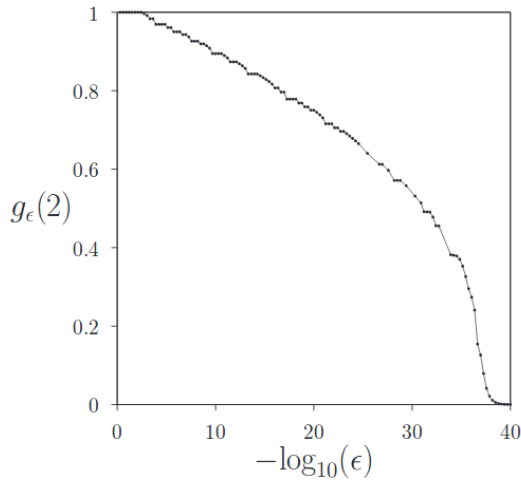


FIGURE 7.8 – **Modèle de percolation à quorum : effet de la résolution sur l’approche du point critique.** Dépendance de la discontinuité $g_\epsilon(m = 2)$ avec la résolution dans un réseau $\bar{k} = 25$, écart type $\sigma = 5$ et dégradation $d = 1$

en montrant que la disparition du point critique en présence de décroissance n’est vraie que à la limite thermodynamique d’un système de taille infinie. Pour un système de taille finie, a fortiori une culture, la fonction de réponse théorique est bien discontinue comme observé dans les expériences.

Cette conclusion est assez intéressante car elle illustre comment dans les applications de modèles de physique statistique à des situations biologiques la vision usuelle de la physique focalisée sur la limite de systèmes de taille infinie n’est pas forcément pertinente. Les modélisations, y compris de type physique statistique doivent tenir compte des effets de taille finie qui peuvent déterminer pratiquement la nature des phénomènes.

7.2.4 Percolation à quorum avec désordre d’excitabilité

Dans la publication « Effect of threshold disorder on the quorum percolation model. » [MONCEAU et collab., 2016] on étudie une autre extension du modèle de percolation à quorum dans laquelle on inclut un désordre Gaussien des seuils (quorum) d’activation de chaque nœud. Ce désordre rend compte de la très probable variabilité dans l’excitabilité des neurones. Le modèle considéré est identique à celui déjà présenté à la section ci-dessus 7.2.2 mais, contrairement au modèle original où tous les nœuds ont la même valeur de seuil, on introduit ici une variabilité aléatoire : chaque neurone i peut avoir une valeur du quorum m_i différente suivant une distribution de probabilité indépendante des propriétés topologiques du réseau.

L’équation de champ moyen précédemment dérivée se généralise facilement avec le même raisonnement en incluant seulement la distribution \mathcal{P}_m des seuils :

$$\Phi = f + (1 - f) \sum_{m=1}^{\infty} \mathcal{P}_m \sum_{k=m}^{\infty} p_k \sum_{l=m}^k \binom{k}{l} \Phi^l (1 - \Phi)^{k-l}. \quad (7.6)$$

En supposant des distributions Gaussiennes des seuils \mathcal{P}_m et des connectivités entrantes p_k avec respectivement les valeurs moyennes et largeurs \bar{m} and \bar{k} et largeurs σ_m et σ_k , l’équation auto-consistante ci-dessus devient :

$$\Phi = f + (1 - f) \frac{1}{2\pi\sigma_k\sigma_m} \sum_{m=1}^{\infty} \sum_{k=m}^{\infty} e^{-\frac{(k-\bar{k})^2}{2\sigma_k^2} - \frac{(m-\bar{m})^2}{2\sigma_m^2}} \sum_{l=m}^k \binom{k}{l} \Phi^l (1 - \Phi)^{k-l}. \quad (7.7)$$

La figure 7.9 montre deux exemples de solution en absence de désordre sur le quorum (haut) et absence (bas). La solution avec désordre est qualitativement similaire à celle sans, avec un décalage de la position du saut à m donné pour des valeurs de f moindres et un décalage du point critique m_c aussi vers une valeur plus faible.

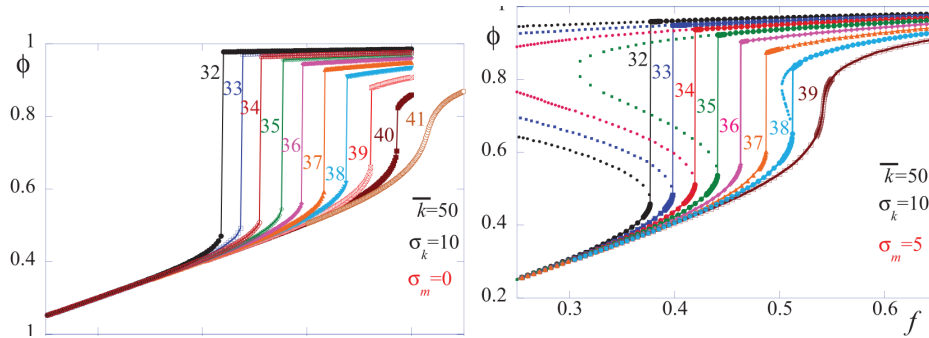


FIGURE 7.9 – **Solutions de l'équation auto-consistante** pour $\bar{k} = 50$ et $\sigma_k = 10$ pour $\sigma_m = 0$ en haut et $\sigma_m = 5$ (en bas). Les pointillés représentent les solutions non physiques.

Le modèle s'étudie également directement par simulations Monte-Carlo. Pour l'estimation de toute observable il faut juste veiller à acquérir des statistiques sur l'ensemble des configurations des différents types de désordre intervenant dans le modèle. À savoir, outre les seuils d'excitabilité, le désordre lié aux configurations des réseaux aléatoires, ainsi qu'aux configurations possibles de choix de la fraction de neurones initialement activée. Ici aussi les deux approches par l'équation auto-consistante et par Monte-Carlo sont remarquablement proches (figure 7.10) en dépit de la taille finie du système dans cette dernière. L'équation auto-consistante capture le comportement champ-moyen valable dans la limite $N \rightarrow \infty$ et dans lequel n'existent pas de boucles alors qu'à taille finie des connexions en boucles fermées sont attendues.

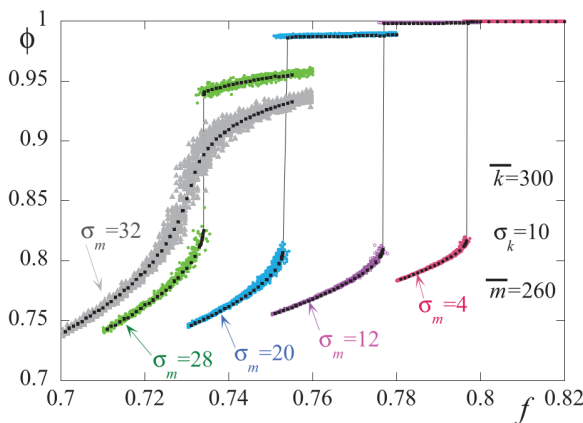


FIGURE 7.10 – **Dépendance de l'activité du réseau pour des désordres du quorum σ_m croissants** à $m = 260$ donné et connectivité constante $\bar{k} = 300$ par simulation Monte Carlo (nuages de points) avec $N = 100000$ nœuds et par l'équation auto-consistante (points noirs).

L'approche Monte-Carlo permet l'étude des effets de taille finie qui demandent une grande attention dans le cas de sauts en raisons des fluctuations des différentes réalisations de désordre sur lesquelles repose chaque exécution. Les simulations Monte-Carlo ne permettent de calculer que des valeurs moyennes $\langle g \rangle$ et $\langle f^* \rangle$ car celles-ci sont déterminées par les fluctuations de taille finie des réalisations de la connectivité aléatoire Gaussienne du réseau à \bar{k} et σ_k fixés, du tirage aléatoire de la fraction f des nœuds initialement activés et du tirage aléatoire des valeurs de l'excitabilité m selon une Gaussienne centrée sur \bar{m} avec une largeur σ_m . La figure 7.11 montre des comportements de $\langle g \rangle$ et $\langle f^* \rangle$ en fonction de σ_m calculées par Monte-Carlo pour un réseau de taille finie grande $N = 100000$ (7.10) pour une valeur moyenne \bar{m} donnée. Une augmentation de σ_m décale toujours la position du saut vers des valeurs plus petites de l'excitation externe correspondant à f^* . Mais on observe aussi qu'une augmentation de σ_m diminue la taille du cluster géant (taille du saut) jusqu'à sa disparition pour une valeur critique $m_c(\sigma_m)$ plus faible qu'en absence de désordre : le désordre peut donc aussi détruire la transition de percolation.

L'effet du désordre du quorum sur la percolation se manifeste de deux manières opposées. Une fraction des nœuds avec un seuil en dessous de \bar{m} devient plus facilement excitable, alors que ceux au-dessus le deviennent moins. Ce sont les nœuds en dessous de \bar{m} qui sont responsables du décalage de $\langle f^* \rangle$ et qui dominant, ce qu'une analyse de l'équation d'auto-consistance permet

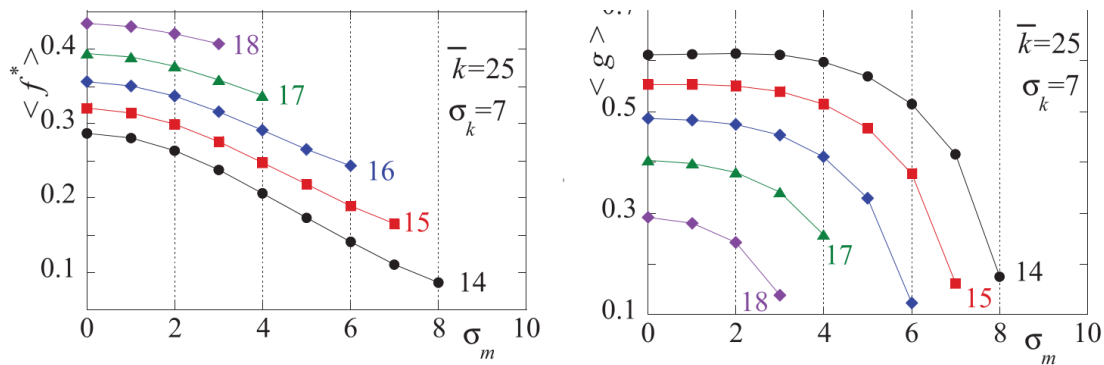


FIGURE 7.11 – Dépendance de la position de la transition du premier ordre (saut) avec le désordre. Position moyenne $\langle f^* \rangle$ et amplitude moyenne $\langle g \rangle$ des sauts en fonction de σ_m pour $\bar{k} = 25$ avec $\sigma_k = 3$ et différentes valeurs de m indiquées sur les figures.

de comprendre.

Une conséquence de cette étude a été de montrer comment le désordre du quorum, donc de l’excitabilité des nœuds, affecte la valeur critique de la transition de percolation. Cette dépendance n’avait pas été prise en compte dans l’exploitation de la percolation à quorum réalisée initialement lors de l’introduction du modèle pour déterminer les propriétés de connectivité du réseau à partir de cette transition.

En absence de désordre la transition de percolation disparaît à une valeur $m_c(0)$ dépendant de la connectivité moyenne \bar{k} et de σ_k selon la relation $m_c(0)/\bar{k} \simeq 1 - A\sigma_k/\bar{k}$ avec $A \simeq 1.3$. Dans les expériences de percolation à quorum effectuées sur des cultures par l’équipe d’Elisha Moses ces auteurs avaient déduit \bar{k} et $\bar{\sigma}_k$ par des ajustements de la relation précédente sur des séries de données manipulant l’excitabilité m à l’aide de bloqueurs synaptiques. Or notre étude montre que la présence d’un désordre σ_m , inconnu mais probable, introduit un paramètre supplémentaire. Les déductions faites en supposant l’absence de désordre sur les seuils doivent conduire à de mauvaises estimations. En l’absence d’informations sur σ_m il ne nous est toutefois pas possible de corriger ces estimations.

7.2.5 Percolation à quorum et inhibition

Le modèle de percolation à quorum se généralise facilement pour inclure une fraction de neurones inhibiteurs dont l’activation a pour conséquence de décrémenter d’une unité leurs voisins en sortie. Dans notre publication [MÉTENS et collab., 2016] discutée plus haut pour l’analyse du point critique nous avons déjà introduit une équation de champ moyen auto-consistante prenant en compte des neurones inhibiteurs. Dans l’article « Effects of inhibitory neurons on the quorum percolation model and dynamical extension with the Brette-Gerstner model » [FARDET et collab., 2018b] nous avons étudié plus en détail les effets de l’inhibition sur la fonction de réponse du réseau à un stimulus. Alors que l’inhibition est généralement considérée comme fondamentale pour l’activité neuronale physiologique, nos résultats amènent à la constatation surprenante qu’en ce qui concerne proprement l’activation de bouffées de potentiels d’action les réseaux avec neurones inhibiteurs sont identiques à des réseaux purement excitateurs avec des connectivités effectives différentes.

La figure 7.12 présente un croquis expliquant le progrès de la percolation du quorum avec neurones inhibiteurs. Il faut noter que, contrairement au modèle excitateur pur, le potentiel d’un neurone n’est plus une fonction monotone croissante du temps de la cinétique à temps discret. La fraction totale de neurones actifs (inhibiteurs et excitateurs) est cependant toujours croissante en raison de la règle du seuil.

Comme dans les études du modèle de percolation à quorum déjà discutées on peut réaliser des

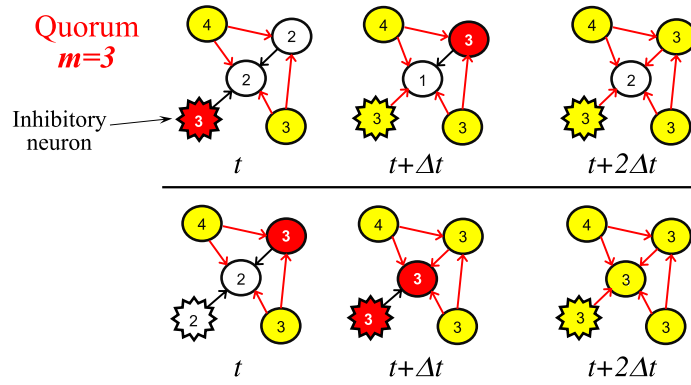


FIGURE 7.12 – **Modèle de percolation à quorum avec inhibition.** Les flèches représentent les liens dirigés entre les neurones. Les neurones représentés en jaune sont actifs (c'est-à-dire qu'ils ont déjà tiré), les blancs sont au repos et les rouges déclenchent qu'au moment indiqué. Un neurone inhibiteur est représenté par un cercle cabossé. La figure supérieure montre une séquence initiant au temps t , le neurone inhibiteur s'active (en raison d'entrées qui ne sont pas représentées ici) ; le potentiel de son voisin en sortie, au centre, passe de 2 à 1. Supposons qu'au temps $t + \Delta t$ le neurone supérieur droit soit aussi activé (aussi en raison d'entrée externes non représentées). À l'instant $t + 2\Delta t$ le potentiel du neurone en sortie augmente de 1. La figure inférieure montre qu'en présence d'inhibition l'ordre avec lequel les neurones sont activés compte. En effet, si on suppose que le neurone excitateur supérieur droit déclenche avant l'inhibiteur en bas à gauche (en raison d'un autre historique d'entrées externes) alors le neurone central est activé alors que ce n'est pas le cas avant la fin du processus dans l'exemple ci-dessus.

simulations Monte-Carlo explicites du processus en réalisant des statistiques sur plusieurs configurations de réseaux aléatoires et tirages de fractions initialement actives. Dans le modèle avec inhibiteurs il est cependant aussi possible d'écrire une équation auto-consistante pour la fraction stationnaire des neurones activés. De la même manière que discuté plus haut, mais en tenant compte de la combinatoire supplémentaire due aux liaisons inhibitrices on aboutit à l'équation auto-consistante suivante qui se calcule numériquement avec les mêmes techniques déjà commentées :

$$\Phi = f + (1 - f) \sum_{k=m}^{\infty} p_k \sum_{k_i=0}^{k-m} \binom{k}{k_i} \eta^{k_i} (1 - \eta)^{k-k_i} \sum_{i=0}^{k_i} \binom{k_i}{i} \Phi^i (1 - \Phi)^{k_i-i} \sum_{e=m+i}^{k-k_i} \binom{k-k_i}{e} \Phi^e (1 - \Phi)^{k-k_i-e} \quad (7.8)$$

Quelques comparaisons des résultats avec les deux méthodes, équation auto-consistante et Monte Carlo, avec $N = 100000$ neurones sont présentées dans la figure 7.13 montrant une très bonne correspondance. Cet accord est remarquable car l'équation auto-consistante représente la solution de champ moyen attendue à la limite d'un système de taille infinie qui ne tient pas compte des corrélations temporelles, alors que la percolation à quorum avec inhibiteur y est sensible (voir figure 7.12).

Notre résultat principal du modèle avec inhibiteur est d'avoir montré son équivalence avec un modèle purement excitateur sur un réseau avec une connectivité effective. Un examen attentif des règles de la percolation à quorum avec inhibition suggère en effet qu'un neurone avec k liens entrants dont k_i sont inhibiteurs, pourrait en moyenne se comporter comme un neurone avec $k - 2k_i$ liens entrants *purement* excitateurs en considérant que chaque neurone inhibiteur pourrait annuler un des excitateurs.

Pour tester cette idée nous avons réalisé des simulations Monte Carlo d'une part de modèles mixtes excitateurs et inhibiteurs avec une distribution Gaussienne du nombre de liens entrants inhibiteurs de moyenne \bar{k}_i et une distribution Gaussienne du nombre de liens entrants excitateurs de moyenne $\bar{k} - \bar{k}_i$, et d'autre part du modèle purement excitateur avec un nombre moyen de liens entrants $\bar{k} - 2\bar{k}_i$.

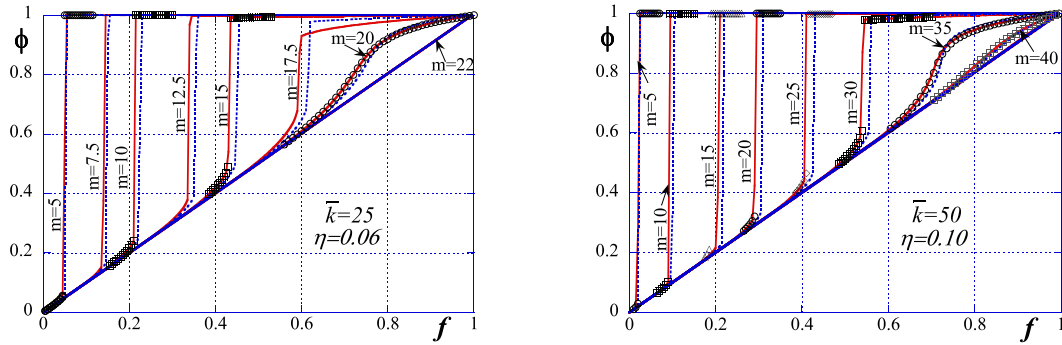


FIGURE 7.13 – **Modèle de percolation à quorum avec inhibition : comparaison des méthodes de résolution.** Comparaison entre simulation Monte Carlo explicites (lignes rouges solides) et solution de l'équation de champ moyen (symboles ouverts noirs) d'un réseau avec inhibiteurs et des simulations avec un réseau excitateur équivalent (ligne bleu pointillée) avec connectivité Gaussienne moyenne $\bar{k}_{eq} = \bar{k}(1 - 2\eta)$. $\bar{k} = 25$, $\sigma = 2.5$ et $\eta = 0.06$ (gauche) $\bar{k} = 50$, $\sigma = 5$ et $\eta = 0.10$ (droit).

Comme le montrent par exemple la figure 7.13, la réponse $\Phi(f)$ stationnaire du réseau mixte est très proche de celle du réseau purement excitateur avec le nombre équivalent réduit de liens entrants. Les différences entre les deux approches dépendent du quorum m et de la fraction initialement active f : elles augmentent avec m et en se rapprochant du saut où la fraction $\Phi - f$ de neurones activés en cascade subit une croissance soudaine. Notre article [FARDET et collab., 2018b] traite en détail les différences entre les modèles avec neurones inhibiteurs et celui excitateur équivalent.

Un autre résultat de cette étude de la percolation avec les neurones inhibiteurs est d'avoir déduit une relation sur la dépendance du point critique avec la fraction de neurones inhibiteurs plus précise que celle suggérée précédemment dans la littérature [SORIANO et collab., 2008]. En se référant à l'expression analytique du point critique obtenu dans le cadre de l'extension analytique du modèle de percolation à quorum [RENAULT et collab., 2014] discutée plus haut 7.2.2, on avait montré que $m_C = \bar{k} \left(1 - a \left(\frac{\sigma}{\bar{k}} \right) + b \left(\frac{\sigma}{\bar{k}} \right)^2 \right)$ et y insérant la valeur du nombre de liens du réseau équivalent on obtient aux ordres dominants en σ/k_{eq} :

$$m_C(\eta) = m_C(\eta = 0) - 2\bar{k}\eta \quad (7.9)$$

où $m_C(\eta = 0) = (\bar{k} - a\sigma)$. Cette relation est en très bon accord avec l'estimation de $m_c(\eta)$ sur des simulations Monte Carlo (figure 8 de [FARDET et collab., 2018a]).

7.2.6 Percolation à quorum dans des modèles dynamiques

Le modèle de percolation à quorum avait été introduit pour étudier la réponse d'une culture neuronale à des stimulations externes. On peut cependant se demander comment le processus de percolation qui a été caractérisé dans des populations neuronales vivantes se retrouve dans le cadre du comportement dynamique spontanée des réseaux. Est-ce que l'activation d'une bouffée de potentiels d'actions sur tout le réseau possédant un comportement dynamique au-delà d'une bouffée unique résulte bien d'un phénomène de percolation ? Ce scénario est vraisemblable : un certain nombre minimal de neurones individuels qui se retrouvent actifs quasi simultanément à cause de fluctuations pourraient représenter une fraction de la population dépassant le seuil f^* de percolation du cluster géant et entraîner la bouffée collective. Un schéma d'activation spontanée qui serait ici dominé par le bruit. Mais d'autres scénarios pour l'activation simultanée de la population sont possibles qui tiennent compte des propriétés dynamiques de chaque neurone. La question générale qui se pose, et nous occupe encore ainsi que bien d'autres groupes, ce sont les différents mécanismes de synchronisation et d'activité collective des populations de neurones.

Comme on l'a vu plus haut, le modèle de percolation semble pertinent pour décrire la réponse d'une population à une stimulation externe d'une certaine intensité. Ici nous avons voulu étudier le phénomène de percolation à quorum dans le cadre d'un modèle dynamique du potentiel neuronal appliqué à une population. Retrouve-t-on un processus de percolation lorsque les neurones sont décrits de manière dynamique? Nous avons choisi pour les simulations dynamiques le modèle de potentiel neuronal Adaptative Exponential Integrate & Fire (AEIF) [BRETTE et GERSTNER, 2005] en raison de sa relative simplicité et capacité à bien reproduire les comportements électrophysiologiques biologiques. Dans ce modèle l'évolution dynamique du potentiel d'un neurone est décrit par deux variables, le potentiel membranaire V et un courant adaptatif lent w , qui sont décrites par les équations :

$$\begin{cases} C_m \frac{dV}{dt} = -g_L(V - E_L) + g_L \Delta_T \exp\left(\frac{V - V_{th}}{\Delta_T}\right) - w + I \\ \tau_w \frac{dw}{dt} = a(V - E_L) - w \end{cases} \quad (7.10)$$

et la condition

$$\text{si } V > V_{peak} \quad \begin{cases} V \leftarrow V_r \\ w \leftarrow w + b \end{cases} \quad (7.11)$$

Les paramètres sont décrits en détail dans [BRETTE et GERSTNER, 2005], les principaux pertinents ici sont E_L , le potentiel de repos et V_{th} le seuil qui marque le début du déclenchement d'un potentiel d'action généré par l'exponentielle divergente. La différence $V_{th} - E_L$ se rapproche donc du quorum du modèle de percolation discuté plus haut.

La connexion d'un neurone A à un autre neurone B est implémentée par un courant post-synaptique « alpha » [ROTH et VAN ROSSUM, 2009], qui prend la forme, si un potentiel à lieu à $t = 0$, d'une injection dans le neurone receveur d'un courant de la forme :

$$I_s(t) = s_{AB} \frac{I_0}{\tau_s} t e^{-t/\tau_s} \quad (7.12)$$

où s_{AB} est un poids synaptique sans dimension et $I_0 = e \cdot 1 \text{ pA}$ est une constante de normalisation qui détermine la hauteur du pic au maximum du courant à la valeur $s_{AB} \text{ pA}$. Des entrées inhibitrices correspondent à des valeurs négatives de s_{AB} . On peut déclarer une fraction η des neurones comme inhibiteurs; toutes leurs synapses sortantes auront pour cela un poids négatif. Les simulations numériques ont été réalisées avec le simulateur neuronal NEST [GEWALTIG et DIEMANN, 2007].

Afin d'étudier la percolation à quorum dans le cadre du modèle dynamique nous avons défini le protocole suivant reproduisant avec le modèle AEIF des expériences numériques équivalentes aux simulations Monte Carlo du modèle discret : (i) une fraction aléatoire f de tous les neurones est activée via un courant post-synaptique externe suffisant qui les amène au-dessus de leur seuil et induit leur déclenchement simultané; (ii) en réponse à cette stimulation, on simule l'activité du réseau suivant le modèle AEIF tant que le nombre de neurones activés cesse d'augmenter.

Afin de reproduire au mieux les conditions du modèle de percolation à quorum dans lequel chaque neurone s'active une seule fois durant l'activation d'une salve, nous avons paramétré les neurones pour que leur période réfractaire après un potentiel d'action soit égale au temps de simulation (pour s'assurer qu'ils ne se déclenchent qu'une seule fois), et on a fixé le délai de transmission des axones à l'équivalent d'un pas de temps du modèle discret, soit $0,1 \text{ ms}$. Les autres paramètres implémentent un neurone de type « adaptive spiking », quoique ceci n'a pas d'impact ici. Il a fallu aussi ajuster le poids synaptique afin d'obtenir un parallèle entre le quorum q et l'excitabilité dans le modèle dynamique pour assurer le déclenchement d'un neurone au-dessus de m potentiels d'action entrants.

La figure 7.14 montre l'activité simulée de cette manière sur un réseau avec le modèle dynamique qui reproduit qualitativement le comportement de la transition de phase de percolation du modèle original, mais également quantitativement comme décrit plus en détails dans notre publication [FARDET et collab., 2018b]. Ce très bon accord confirme que le modèle de percolation à quorum, incorporant aussi des inhibiteurs, capture le comportement d'un modèle dynamique plus sophistiqué et est donc pertinent pour décrire l'initiation d'une bouffée d'activité sur un réseau.

Afin de rechercher des manifestations de cette percolation dans des conditions moins contrôlées nous avons simulé des réseaux de neurones excitateurs intrinsèquement oscillants dont le potentiel membranaire se dépolarise lentement sous l'influence de courants sodium persistants $I_{Na,p}$ jusqu'au déclenchement d'un potentiel d'action qui réinitialise un cycle. Par rapport au cas étudié ci-dessous de stimulation externe l'activité dans le réseau a pour origine l'activité propre de chaque unité. On permet aussi ici des bouffées de déclenchements en ayant relâchés la contrainte sur le temps réfractaire. Les neurones sont connectés via un réseau Gaussien de connectivité moyenne $\bar{k} = 100$ et déviation standard $\sigma = 5$, les potentiels d'action sont transmis entre neurones avec un retard constant de 1 ms.

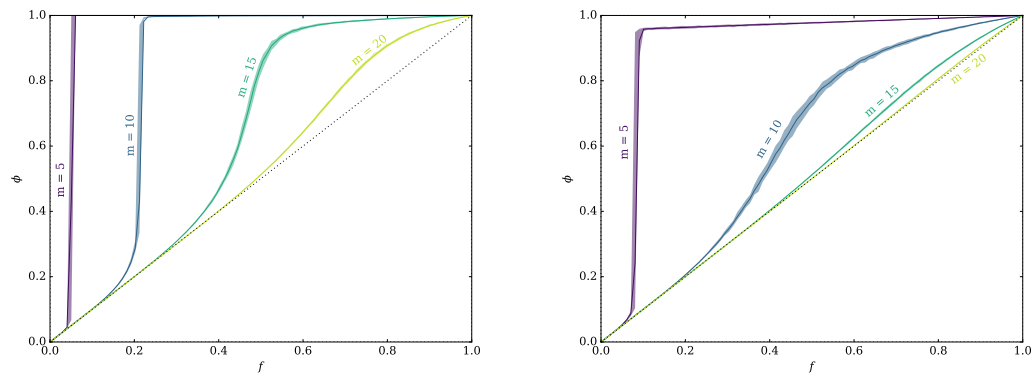


FIGURE 7.14 – **Modèle de percolation à quorum en fonction de la fraction d'inhibition.** Transition de phase simulée avec le modèle dynamique de percolation à quorum pour des fractions de neurones inhibiteurs $\eta = 5\%$ à gauche et 25% à droite. Courbes moyennes sur plus de 50 simulations pour quantifier les fluctuations. Simulations sur des réseaux de 10 000 neurones avec distribution Gaussienne des connectivités entrantes ($\bar{k} = 25$ et $\sigma_k = 5$). La courbe de transition moyenne est représentée par la ligne continue (pour des quorums croissants $\{5, 10, 15, 20\}$ de violet à vert). Les surfaces remplies délimitent les 5^e et 95^e pour-centiles. La ligne pointillée marque la courbe $\Phi = f$. Comme dans le modèle à champ-moyen, l'augmentation de la fraction d'inhibiteurs diminue la valeur critique du quorum $10 < m_c < 15$ pour $\eta = 0.05$ contre $5 < m_c < 10$ pour $\eta = 0.25$.

La figure 7.15 montre la trame d'une simulation faisant apparaître des bouffées de réseau périodiques, avec chaque bouffée dotée d'une structure particulière. Chaque bouffée est une succession d'activation de tranches de neurones synchronisés. Une première tranche de neurones synchronisés apparaît probablement résultant d'une combinaison suffisamment rapprochée de déclenchements spontanés et des effets de remise à zéro simultanés avec un renforcement positif [BOTTANI, 1995]. Cette première tranche d'activité synchronisée agit par la suite sur l'évolution du réseau comme s'il s'agissait d'une stimulation externe agissant sur la population elle-même et entraîne l'activation de neurones supplémentaires. Chaque tranche de déclenchements au sein d'une bouffée est initiée par la précédente et séparée temporellement par le délai de propagation ainsi que le temps pour chaque neurone pour intégrer une stimulation en entrée suivant le modèle AEIF. L'intervalle entre chaque tranche d'activité augmente avec chaque répétition en raison de l'adaptation, qui augmente de manière effective le quorum nécessaire pour l'activation globale car l'effet d'une tranche d'activité s'étale sur plusieurs milli-secondes tenant compte aussi des délais de transmission. À un certain point l'effet de l'activation d'une tranche précédente stimule un nombre insuffisant de neurones du réseau pour entraîner une nouvelle tranche d'activation et la

bouffée s'arrête. La propagation sur le réseau en lien avec les propriétés de connectivité devient apparente si on réordonne la trame de la simulation par ordre croissant de connectivité entrante des neurones dans la figure 7.15 C). L'activité se propage clairement le long de neurones avec des propriétés de connectivité différentes. Les neurones de plus haute connectivité atteignent le seuil plus facilement et s'activent avant le reste du réseau.

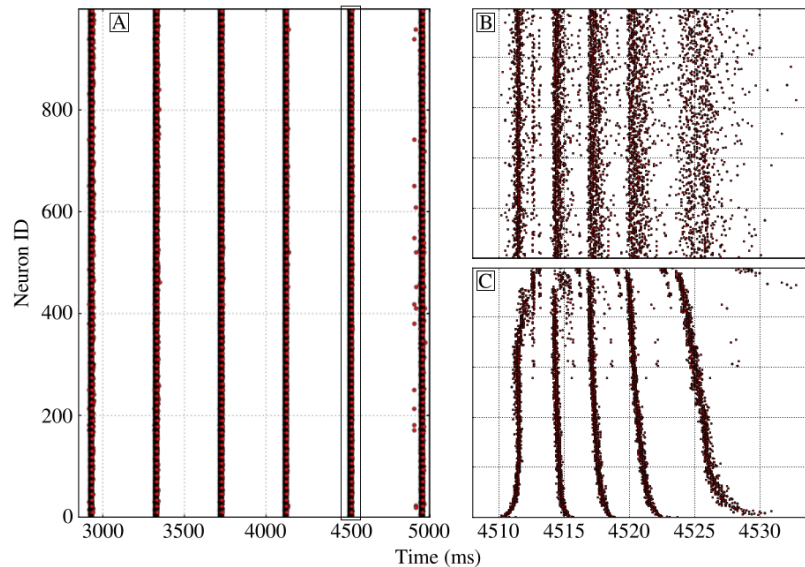


FIGURE 7.15 – **Modèle de percolation à quorum avec un modèle dynamique de neurone.** A) Trame d'un réseau de $N = 1000$ neurones oscillants avec distribution Gaussienne $\bar{k} = 100$ et $\sigma = 5$. B) Dynamique détaillée de la 5e bouffée (encadrée) avec les tranches successives de neurones synchronisées. C) La même activité qu'en B, mais avec les neurones ordonnés par ordre croissant de connectivité entrante.

Le paradigme de la percolation propose donc une interprétation des bouffées quasi synchronisées durant une bouffée. À part la première, chaque tranche synchronisée successive est la réponse de percolation à la précédente. La première tranche active démarre avec le déclenchement des neurones les plus faiblement connectés. Ces neurones ont reçu moins d'entrées lors de la bouffée précédente. Avec moins de signal en entrée, ils ont fait l'objet alors de moins d'adaptation et reprennent plus rapidement leur cycle oscillant. Ceci les conduit à devenir les initiateurs d'une nouvelle bouffée. Dans la première tranche de la nouvelle bouffée, ce sont ensuite les neurones les plus connectés qui recevant de ce fait davantage d'entrées venant des précurseurs, sont activés en premier et entraînent les tranches suivantes (à partir de la seconde) de la bouffée.

Ce scénario ne peut cependant être que partiel. Les processus de percolation en action durant une bouffée interfèrent en effet avec la dynamique complexe propre des neurones. L'analyse dynamique du modèle AEIF indique en effet que le comportement qualitatif des neurones peut changer en fonction des propriétés de courants injectés en entrée (intensité, taux de répétition), comme ceux résultant d'une somme d'un grand nombre de courants post-synaptiques dans une tranche d'une bouffée. Ainsi un neurone oscillant simple régulier soumis à un courant en entrée constant suffisant peut donner lieu à des bouffées de potentiels d'actions d'allure similaire aux bouffées de réseau. Le schéma de percolation que nous reconnaissons lors du déroulé d'une bouffée se combine avec les propriétés dynamiques des neurones. La synchronisation des neurones au sein d'une bouffée génère des courants postsynaptiques importants sur chaque unité, modulés par la connectivité de chacun et étalés dans le temps suivant l'étalement temporel de chaque tranche précédente et les délais de propagation. L'arrivée de tels courants sur un neurone contribue à l'exciter mais également à éventuellement le placer dans un mode dynamique différent de son fondamental. Cette richesse dynamique n'existe pas dans le cadre du modèle de percolation à quorum simple. Cette question nous a amené à étudier plus spécifiquement la dynamique des neurones au sein d'une bouffée de réseau et aboutir aux résultats sur le comportement synchronisé qui sera discuté dans la section 7.7.3.

7.2.7 Contribution au contrôle et à la mesure expérimentaux des cultures

La publication « Combining microfluidics, optogenetics and calcium imaging to study neuronal communication in vitro » [RENAULT et collab., 2015] est un résultat de la collaboration avec les équipes à l'Institut Curie et à l'Institut Weizmann dans le cadre de la thèse de Renaud Renault sur un sujet expérimental. La motivation initiale de ce travail était de mettre au point un système expérimental permettant d'explorer facilement les réponses de deux populations neuronales en contact à des stimulations externes afin de rechercher des phénomènes d'apprentissage dus à la plasticité neuronale. Le dispositif microfluidique développé dans ce but (figure 7.16) permet de mettre en contact deux chambres de culture connectées par des micro-canaux asymétriques qui privilégient un sens de connexion. En stimulant différenciellement les deux populations simultanément ou avec des écarts temporels, l'idée est de parvenir à renforcer ou réprimer les connexions d'une population vers l'autre. Ce changement de poids des connexions entre les chambres permettrait donc d'enregistrer une information dans le système.

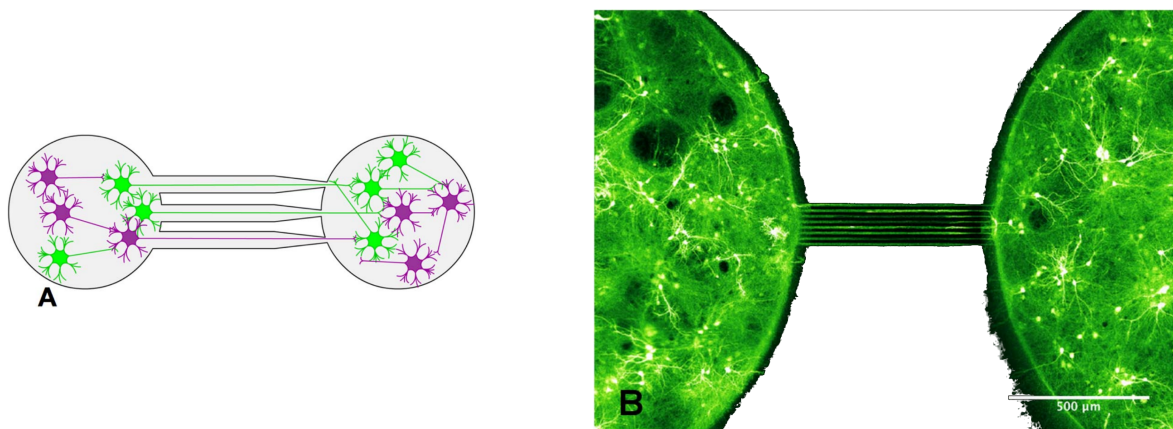


FIGURE 7.16 – **Dispositif neuronal simple à deux chambres.** A- Schéma montrant le dispositif neuronal comportant deux chambres d'ensemencement et connectées par des canaux rétrécis de gauche à droite à dessin de privilégier les connexions de gauche à droite et faire office de diodes axonales. B- Image par microscopie confocale d'un dispositif neuronal suivant ce schéma montrant des neurones exprimant la channelrodopsine Chr2-YFP (fausse couleur). Source : [RENAULT et collab., 2015]

Cette expérience requiert cependant de pouvoir techniquement effectuer ces stimulations séparées, tout en observant l'activité de l'ensemble du système. Le travail présenté dans la publication propose une solution pour permettre ce contrôle et ces mesures mettant en œuvre des dispositifs uniquement optiques qui évitent des systèmes de micro-électrodes délicats à concevoir et mettre en place.

Ceci a été réalisé en ajustant le choix de canaux optogénétiques à transférer dans les neurones pour que ceux-ci soient excitable optiquement à des longueurs d'onde n'interférant pas avec les signaux des rapporteurs fluorescents sensibles au Ca^{2+} utilisé pour observer l'activité électrique des neurones. Renaud Renault a ainsi déterminé durant son séjour à l'Institut Weizmann pendant sa thèse des choix compatibles de canaux et de marqueurs, qui ont pu être mis en application dans des expériences à l'Institut Curie, à savoir un indicateur Calcium orange excitable avec un laser blanc réglé à 549 nm pour la mesure de l'activité et une transfection avec la channelrodopsine Chr2-YFP pour l'excitation optogénétique avec une diode bleue à 470 nm.

La figure 7.17 illustre le fonctionnement de la construction avec des intensités de stimulation optique de plus en plus importantes. Cette expérience reproduit de manière optique l'expérience de percolation à quorum sur des neurones vivants. En absence d'inhibiteur CNQX des synapses excitatrices AMPA l'expérience du haut (figure 7.17A) montre une réponse constante indépendante de la durée d'excitation. Après addition de CNQX les réponses deviennent proportionnelles à la durée de la stimulation (figure 7.17B). Cette situation reproduit la percolation à quorum, l'ajout de CNQX correspondant à une augmentation du seuil m . En absence d'inhibiteur la stimulation op-

tique excite une fraction de neurones f au delà du seuil de percolation f^* et entraîne une réponse totale. Avec les inhibiteurs la situation de la figure 7.17B doit correspondre à un dépassement du seuil critique m_c avec une régime uniquement graduel.

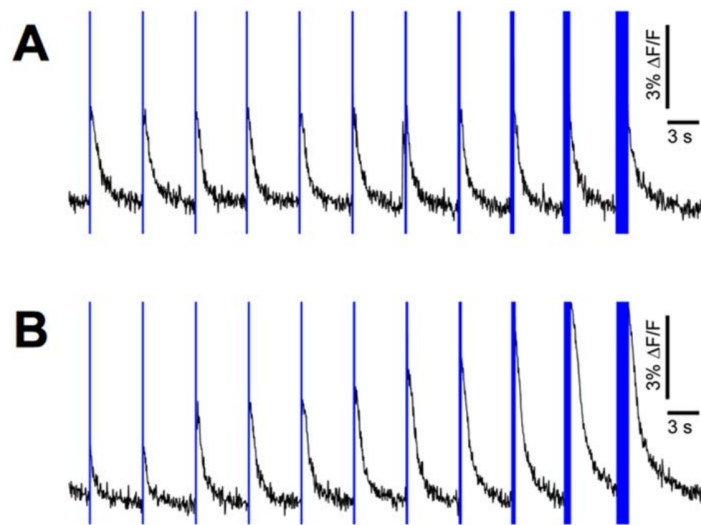


FIGURE 7.17 – **Induction et détection optique de bouffées d'activité.** Réponses typiques d'une culture (DIV 13) à des durées de stimulation différentes (1, 2, 4, 8, 16, 32, 64, 128, 256, 512 et 1024 ms). Les barres verticales bleues représentent les temps et durées de stimulation. B- Même expérience qu'en A après l'addition de 10 μ M de CNQX l'antagoniste des récepteurs AMPA. Source : [RENAULT et collab., 2015].

Afin de déterminer les performances de ces constructions dans un dispositif microfluidique, des expériences ont été réalisées avec un ensemencement asymétrique des chambres, l'une seule des deux étant transfectée avec la channelrhodopsine et excitable optiquement. Les expériences de la figure 7.18 permettent ainsi d'évaluer la transmission de signaux dans chaque direction. Cette expérience avec des canaux parallèle montre des délais de transmission toujours inférieurs à 5 ms.

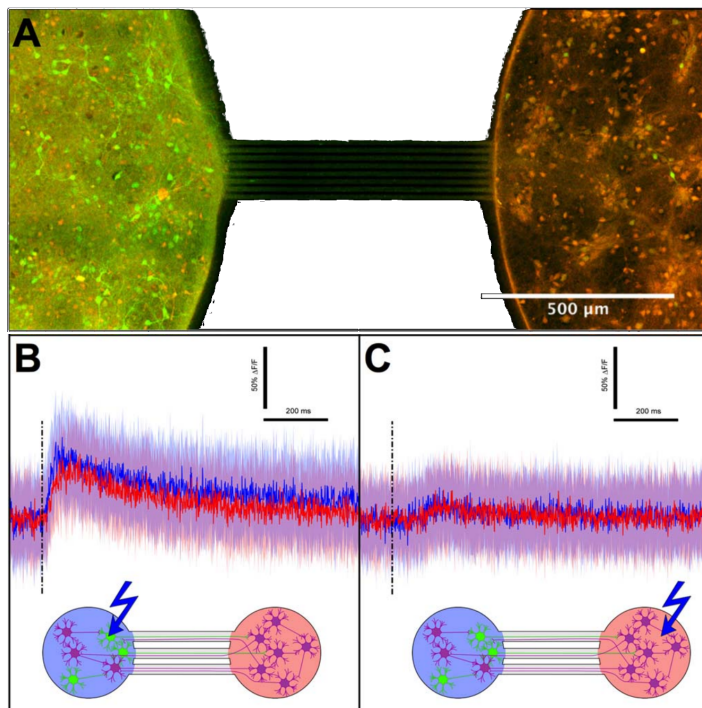


FIGURE 7.18 – **Stimulation locale et propagation de bouffée.** A : image du dispositif à deux chambres connectées par des canaux parallèles avec uniquement les neurones de gauche transfectés activable optogénétiquement; en vert les neurones portant ChR2-YFP, en orange le marqueur Ca-organge. 70% des cellules portant ChR2-YFP sont dans la chambre initiale. B et C : Comportements moyens après stimulation des neurones transfectés (B) et non transfectés (C). Les lignes colorées indiquent la fluorescence calcium moyenne sur 30 stimulations dans chaque chambre (bleu gauche), rouge (droite), les surfaces ombragée la zone d'une déviation standard. La ligne pointillée verticale l'instant de la stimulation. Source : [RENAULT et collab., 2015]

Une autre configuration test à été étudiée avec des canaux asymétriques servant à privilégier une direction de connections neuronales et faire office de diodes pour l'activité neuronale. Les comportements obtenus dont un exemple est visible dans la figure 7.19 reproduisent très bien les

attentes et le système a été utilisé pour rechercher des phénomènes de plasticités synaptiques qui renforceraient ou réprimeraient la propagation de signaux dans un sens ou dans l'autre après une phase d'apprentissage. Ces résultats négatifs quant à la présence de plasticité sont commentés dans la section 6.6.3 du chapitre 6 où est discutée la situation confuse de cette question.

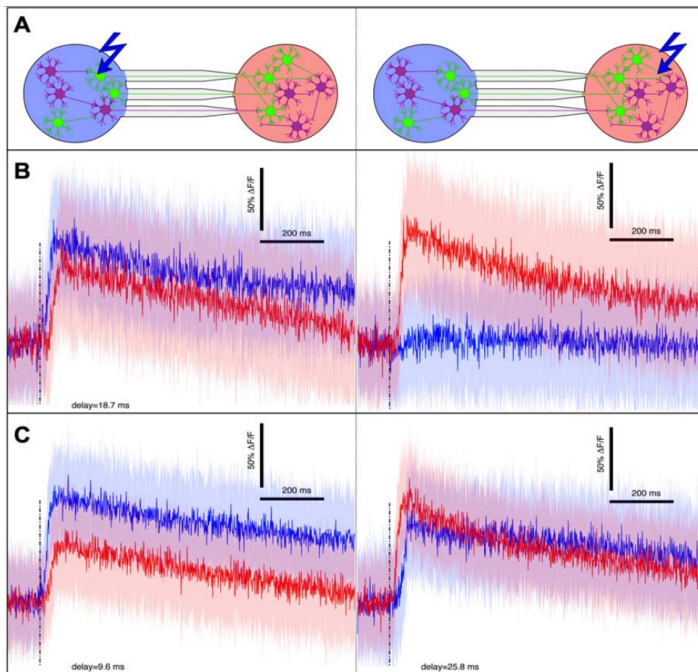


FIGURE 7.19 – **Transmission asymétrique.** A : diagramme explicatif du dispositif et des codes couleurs. B : exemple d'un dispositif montrant une transmission unidirectionnelle. C : Résultats d'un dispositif où la bouffée se propage aussi en sens inverse, quoique avec un retard plus grand. Source : [RENAULT et collab., 2015]

7.2.8 Résumé et perspective « Études de la stimulation et de la percolation dans les cultures »

Les articles présentés ci-dessus concernent des approches pour comprendre la stimulation d'activité dans les cultures neuronales ainsi que l'implication de l'équipe dans des développements expérimentaux pour le contrôle des cultures. Les résultats obtenus par nous et dans d'autres groupes ne sont encore qu'une étape dans le développement et l'utilisation des cultures pour des dispositifs neuronaux contrôlables à des fins d'explorations cognitives.

Les résultats des modèles de quorum percolation, soutenus par les expériences de nos collaborateurs, permettent d'envisager des dispositifs neuronaux mésoscopiques, où l'unité de base à la place d'un neurone est un groupe de neurones relativement homogène qui peut répondre d'une manière presque discrète à des stimulations. Des canaux sélectifs peuvent mettre en contacts de tels unités distinctes et former un réseau.

Un réseau minimal de cette nature formé de deux chambres de culture interconnectés a pu être conçu et les moyens de le contrôler et d'observer son activité mis au point. Cependant toutes les tentatives réalisées à ce jour dans le cadre de nos collaborations pour obtenir un effet de plasticité significatif et persistant entre les sous-populations en contact et réaliser un apprentissage ont été vaines. Ces efforts sont discutés en détails dans le dernier chapitre « Teaching neuronal devices : looking for Synaptic Plasticity » de la thèse de Renaud Renault.

7.3 Bouffées synchronisées de potentiels d'actions (Network bursting)

7.3.1 Motivation et contexte

Comme discuté plus haut et dans la section 6.5 du chapitre 6 de nombreuses populations neuronales en culture et *in vivo* affichent des comportements collectifs faits de répétitions espacées

dans le temps de bouffées de nombreux déclenchements quasi simultanés de tous les neurones. Ces séquences de forte activité quasiment synchronisée de toute la population se répètent périodiquement entre des périodes de quiescence et souvent ressemblent à des formes d'activité épileptique. Ce comportement périodique synchronisé particulier est un phénomène émergent observé dans des populations neuronales, aussi bien de zone cérébrales [MEISTER et collab., 1991] que dans des cultures neuronales [KAMIOKA et collab., 1996; LUCCIOLI et collab., 2014; STEGENGA et collab., 2008; WAGENAAR et collab., 2006]. C'est le comportement dynamique marquant observé dans les cultures adultes en absence de stimulation externe et dont l'origine est encore très discutée.

La source de toute activité dans la population doit d'abord se situer au niveau de l'activité de neurones individuels, ensuite suivie par les conséquences de la propagation de signaux synaptiques vers d'autres neurones. Plusieurs scénarios sont proposés dans la littérature pour justifier de l'activité individuelle, intrinsèque observée des neurones. Le premier rend compte avant tout de l'existence de différentes sources de bruit dans le système pouvant amener certains neurones au hasard à déclencher [ORLANDI et collab., 2013]. En fonction de la connectivité de ces neurones initiateurs, cette activité de bruit initiale peut se propager plus ou moins et éventuellement entraîner l'activation de l'ensemble du réseau. Selon le second scénario, les populations neuronales contiennent des neurones naturellement actifs, se déclenchant régulièrement par eux-mêmes [PENN et collab., 2016; RAMIREZ et collab., 2004] et pouvant entraîner la dynamique de tout le réseau. D'autres travaux ont aussi suggéré l'existence de certains neurones « hub » possédant des connexions à longue portée qui seraient responsables du déclenchement et de la synchronisation de bouffées de réseau [PICARDO et collab., 2011].

Pour les perspectives d'applications de recherche ou technologiques de cultures neuronales microfluidiques il serait intéressant d'apprendre à contrôler l'apparition de bouffées d'activités dans les dispositifs afin de maîtriser les comportements et pouvoir agir avec plus de finesse sur des facteurs comme la plasticité synaptique. L'étude de cette question est donc une continuation logique du travail sur les critères de propagation d'une vague de déclenchements à grande échelle avec les modèles de percolation discuté plus haut. Dans cette optique nous avons pu tirer parti dans l'équipe de résultats expérimentaux inédits et surprenants obtenus par nos collaborateurs Elisha Moses et Menahem Segal dans leur article de 2016 « Network synchronization in hippocampal neurons » [PENN et collab., 2016] qui illustrent des conditions inattendues.

Dans des expériences sur des cultures de neurones dissociées réalisées en contrôlant très précisément l'environnement chimique, ils ont ainsi montré qu'une majorité de près de 60% des neurones de l'hippocampe et du cortex se comportent comme des oscillateurs à relaxation qui se déclenchent périodiquement, aussi bien individuellement lorsqu'ils sont isolés, ou lorsqu'ils interagissent au sein d'un réseau. Couplés au sein d'un réseau les neurones d'une population synchronisent progressivement leurs activités en fonction de la force du couplage. Partant d'un couplage faible, les neurones commencent à converger vers une fréquence de périodicité commune, proche de la moyenne des fréquences propres. Lorsque le couplage augmente, le déphasage entre les neurones diminue et ils se synchronisent en fréquence et en phase produisant le comportement de bouffées de réseau synchronisées connu.

La synchronisation collective d'un grand nombre d'oscillateurs est un exemple classique de comportement complexe hors d'équilibre et d'auto-organisation étudié depuis longtemps dans de nombreux contextes y compris biologiques. L'article de nos collaborateurs est dans ce contexte particulièrement remarquable. Malgré le très grand nombre d'études sur la synchronisation collective, c'est peut-être un des premiers exemples d'étude expérimentale d'oscillateurs biologiques avec un contrôle du couplage. Cet article est donc une source d'inspiration très importante pour l'étude théorique des comportements de cultures.

Un travail qui que nous avons entamé dans l'équipe en se focalisant dans un premier temps sur les propriétés des bouffées de réseaux synchronisées elles-mêmes plutôt que sur les processus conduisant à cette synchronisation. Dans notre publication « Understanding the Generation of

Network Bursts by Adaptive Oscillatory Neurons, » [FARDET et collab., 2018a] nous avons donc pris les résultats de PENN et collab. [2016] comme point de départ pour étudier les particularités du comportement temporel particulier sous forme de bouffées. C'est-à-dire d'un comportement périodique caractérisé essentiellement par deux échelles de temps, celle de la durée de la phase active et celle de l'intervalle entre deux de ces phases.

Les propriétés des neurones et de leurs interactions font intervenir de nombreuses échelles temporelles résultant dans les variations hautement non linéaires du potentiel d'action durant un déclenchement, des courants post-synaptique et de délais de transmission lors d'une communication synaptique. On veut comprendre quels processus au niveau du neurone déterminent les deux échelles temporelles des bouffées synchronisées de réseau.

7.3.2 Résumé des résultats

Dans ce travail nous nous sommes concentrés sur la caractérisation des propriétés des bouffées de réseau comme attracteurs de la dynamique. L'étude de l'émergence de ce comportement, c'est-à-dire l'évolution de la population neuronale vers ce comportement asymptotique fait l'objet de projets actuellement en cours. Cependant, si une description précise de l'émergence du comportement collectif dans le réseau des neurones en culture fait défaut, on sait déjà avec confiance que la synchronisation de la population est un comportement générique robuste. De manière retrospectivement intéressante, j'avais abordé une telle question il y a très longtemps déjà dans ma thèse montrant comment la synchronisation est un comportement asymptotique très général pour des oscillateurs à relaxation couplés par des impulsions [BOTTANI, 1995]! Avec des modèles mathématiques de neurones et de synapses plus raffinés que dans mon ancienne publication, les simulations que nous avons réalisé sur des populations neuronales de quelques milliers à des centaines de milliers d'éléments ont toujours de manière consistante convergé vers ce comportement asymptotique pour tous les jeux de paramètres biologiquement significatifs et correspondant à des neurones périodiques adaptatifs. En effet le résultat principal de notre étude a été de montrer que la propriété d'adaptation est un critère suffisant pour donner des bouffées de réseau.

Se plaçant dans le cadre des résultats de Penn *et al.* nous avons pris dans notre étude une population de neurones oscillants adaptatifs. Le comportement périodique intrinsèque des neurones peut s'expliquer par la présence de courants de fuite empêchant un point fixe attracteur de leur dynamique. L'existence de tels canaux dits « pacemakers » à l'origine des rythmes neuronaux est caractérisée depuis une vingtaine d'années [BOIS et collab., 2007; GAUSS et SEIFERT, 2000; GRITSUN et collab., 2010]. La génération de la périodicité du potentiel membranaire est le résultat de l'action entre différents types de courants en particulier du courant I_h (hyperpolarization-activated) [LÜTHI et MCCORMICK, 1998] et du courant sodium persistant $I_{Na,p}$ [KOIZUMI et SMITH, 2008; LU et FENG, 2012].

La propriété d'adaptation des neurones, qui s'avère centrale dans nos résultats, est la capacité de ceux-ci d'évoluer leur comportement en fonction de l'excitation qu'ils reçoivent. En particulier l'excitabilité des neurones diminue en réponse à des activations précédentes. Ainsi stimulé par un courant continu, les neurones adaptatifs ont une fréquence de déclenchements périodiques qui diminue progressivement. Deux phénomènes se superposent alors : la génération de déclenchements rythmiques d'une part, intrinsèque ou stimulée par un courant d'entrée continu; impliquant plusieurs canaux avec des effets de dépolarisation, et d'autre par la modulation de ce rythme, c'est-à-dire l'adaptation, et qui implique l'activation d'autres types de canaux ioniques. Différents courants associés à ces canaux contribuent à cette adaptation notamment le courant potassium « muscarinique » I_M [FABER et SAH, 2002] et le courant potassium activé le calcium « slow after hyperpolarization » I_{AHP} [KOBAYASHI et KITANO, 2016].

Nous avons développé un modèle analytique sur une représentation effective du système que nous avons montré bien approximer des simulations numériques utilisant un modèle minimal classique de potentiel neuronal, le modèle Adaptive Exponential Integrate and Fire (AEIF) déjà

introduit plus haut. Le modèle analytique permet alors de calculer efficacement les comportements de populations en fonctions des paramètres du modèle de neurone et ainsi d'explorer les conséquences de chacun de ceux-ci sur la dynamique. Une exploration large impossible à réaliser avec le modèle numérique. Ceci nous a permis d'identifier les paramètres du modèle AEIF déterminant les bouffées de réseaux et de calculer leurs effets. En associant les paramètres du modèle avec des canaux ioniques réels on a pu ainsi conclure sur le rôle suffisant de courants d'adaptation pour les bouffées et anticiper des changements attendus dans des expériences sur les différents canaux impliqués.

L'étude est basée sur deux modèles de réseaux aléatoires non spatiaux. Un réseau entièrement homogène avec une connectivité fixe a été utile pour introduire la démarche et effectuer des calculs explicites. L'approche a pu ensuite être appliquée à un modèle plus hétérogène de distribution Gaussien des liens entrants comme utilisé dans le cadre de la percolation à quorum.

Description du comportement dynamique des bouffées synchronisée et attracteur

L'objectif consistait à construire un modèle « équivalent » analytique capable de décrire les propriétés de la dynamique synchronisée du système, c'est-à-dire, durée des bouffées, durée des intervalles entre bouffées, sans nécessiter de simulation Monte Carlo explicites du réseau.

La dynamique lors des bouffées de réseau est représentée dans la figure 7.21 pour un neurone représentatif, équivalent à tous les autres quand la synchronisation est parfaite. Plusieurs étapes constituent le déroulé d'un cycle des bouffées périodiques. Des phases successives qui vont faire l'objet dans la démarche de modélisation qui sera décrite ci-dessous d'approximations spécifiques pour chacune. Le comportement des deux variables du modèle AEIF durant un cycle de bouffé est décrit dans la figure qui permet de comprendre les principales caractéristiques de cette dynamique : intensité de la bouffée qui correspond au nombre de potentiels d'action durant celle-ci, ralentissement entre les potentiels d'actions durant la bouffée et durée de l'intervalle entre les bouffées.

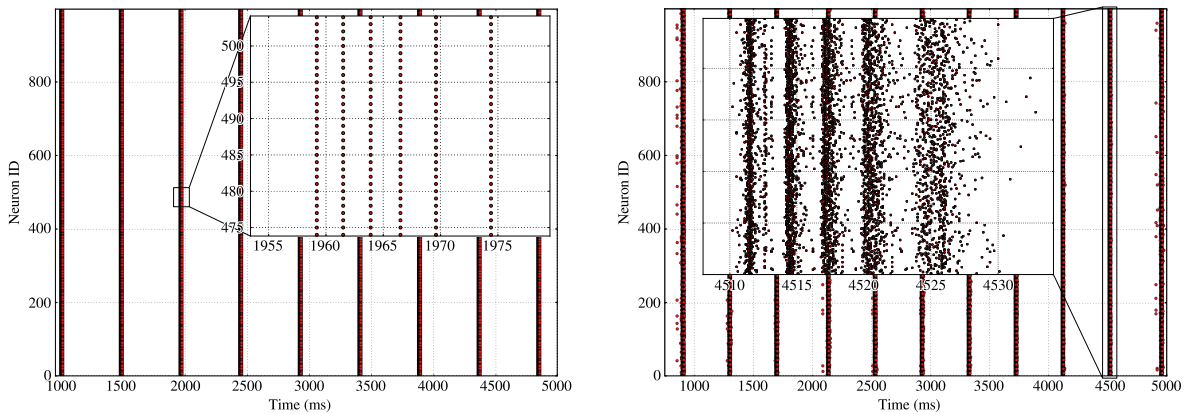


FIGURE 7.20 – **Trame des bouffées synchronisées de réseau.** L'encart montre le détail des potentiels d'action des neurones durant une bouffée unique, les intervalles temporels entre tranches d'activité synchronisée augmentent au fur et à mesure de l'augmentation de l'adaptation. Chaque ligne représente un neurone et les points indiquent les instants d'émission de potentiels d'action. A) Pour un graphe de degré fixe 100. B) Pour un réseau Gaussien de degré de connectivité entrante $\bar{k} = 100$ et $\sigma_k = 5$; les tranches synchronisées au sein d'une bouffée restent bien définies mais s'élargissent.

Le comportement du régime synchronisé est composé de bouffées se répétant périodiquement et contenant une rafale de potentiels d'actions est représenté sur la figure 7.20 pour deux types de réseaux de $N = 1000$, en A) très régulier de degré fixe et en B) avec une distribution Gaussienne du nombre de liens entrants sur chaque neurone. Les trames sont obtenues par des simulations de l'activité du réseau, c'est-à-dire l'intégration numérique du système des $2N$ équations différentielles du réseau avec le modèle AEIF. On y visualise directement les bouffées périodiques

durant lesquelles tous les neurones sont actifs simultanément et dans les encarts la structuration temporelle de l'activité des neurones au sein d'une bouffée. La figure 7.22 focalise sur un neurone et montre l'évolution temporelle des deux variables d'état V et w dans la sous figure A) et le diagramme des phases dans la sous figure B) et figure 7.22. Durant un cycle les variables d'état du neurones (V, w) ne suivent pas l'attracteur à vitesse constante : le neurone passe beaucoup plus de temps sur le chemin de récupération (V petit) par rapport à la phase active de la bouffée (succession de pics du potentiel V).

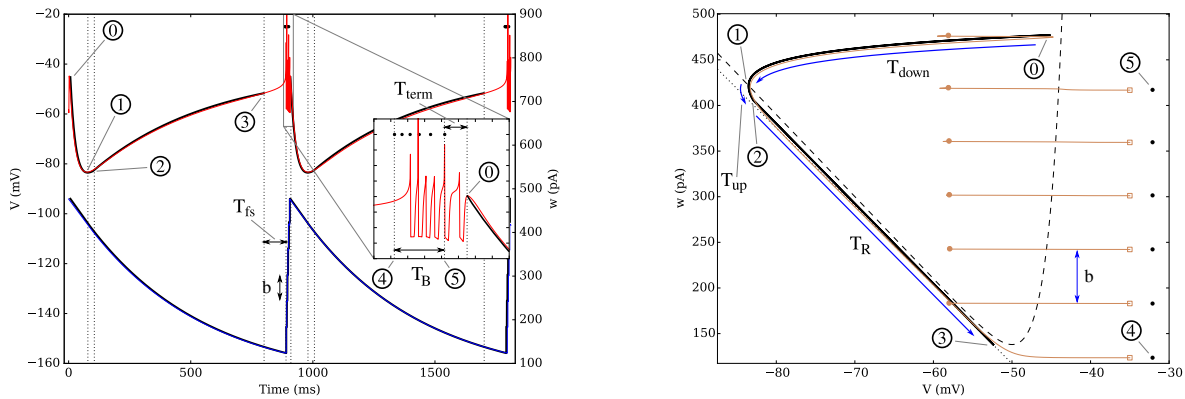


FIGURE 7.21 – **Évolution temporelle du modèle adaptatif exponentiel integrate & fire AEIFA** : Évolution temporelle des paramètres V et w pour notre modèle théorique (ligne noire épaisse et nombres entourés) et pour une simulation AEIF (ligne rouge fine pour V et bleu pour w) sur un graphe de degré fixe $k = 100$. Deux bouffées sont représentées et la dynamique entre les bouffées agrandie dans l'encart. Les instants des potentiels d'action prédits par notre modèle sont marqués par des points noirs. Les cercles numérotés indiquent les points principaux de la dynamique théorique où la nature de l'évolution temporelle change. **B** : Graphe de l'attracteur de la dynamique dans l'espace des phases pour le modèle théorique (ligne épaisse noire) et pour la simulation (ligne fine marron). Les trois premières phases de l'évolution tout de suite après une bouffée sont indiquées par des flèches bleues : le potentiel V décroît fortement jusqu'à une valeur minimale quand il croise la nullcline V (tirets); ensuite un court intervalle durant lequel le neurone évolue rapidement vers la courbe de récupération (ligne pointillée) suivie jusqu'au minimum de la nullcline V et le démarrage d'une nouvelle bouffée de potentiels d'action. La trajectoire des potentiels d'actions est coupée dans la figure (marqué par des carrés vides) et le point de remise à niveau quand le potentiel revient à V_r est marqué par un cercle plein, où la variable d'adaptation est incrémentée de b . Les paramètres neuro-naux et synaptiques correspondants à cette évolution peuvent être retrouvés dans la table S1 des parties supplémentaires [FARDET et collab., 2018b].

Description de l'approche de modélisation

Comme on s'intéresse uniquement aux propriétés du comportement synchronisé on se focalise sur cet état. Notre approche considère tous les neurones synchronisés et calcule de manière auto-consistante les propriétés de ce régime. Le schéma de la figure 7.23 dans le cas d'un réseau exactement homogène de connectivité k illustre la construction du modèle. Si toute la population est synchronisée, tous les neurones reçoivent la même entrée, c'est-à-dire la contribution simultanée de k neurones donnée par la somme des potentiels post synaptiques. Chaque neurone se comporte exactement comme tout autre.

Ceci signifie qu'un réseau de N neurones recevant k entrées de poids s est équivalent à N neurones isolés, chacun ayant une boucle de rétroaction sur lui-même de poids $k \times s$. Cette simplification est inexacte si les neurones n'ont pas le même nombre de connections entrantes, mais nous avons pu montrer qu'elle est encore très correcte pour des réseaux Gaussiens, voir d'autre types de réseaux aléatoires encore plus hétérogènes. Notre modèle de bouffées de réseau synchronisées est basé sur cette observation, exacte pour un réseau de degré fixe. La population entière est représentée par le comportement d'un neurone « équivalent » qui capture la dynamique moyenne. Ce neurone est sujet à la moyenne des entrées reçues par les neurones du réseau et est décrit par un

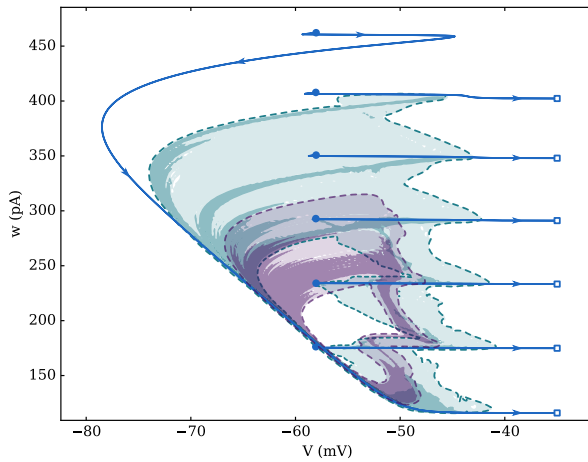


FIGURE 7.22 – **Attracteur de la dynamique simulée (intégration des équations différentielles) dans l'espace des phases pour 3 différents réseaux de 1000 neurones identiques.** L'attracteur pour le réseau à connectivité fixe est représenté par la ligne bleue (les potentiels d'action sont représentés par des carrés vides et les positions de remise au niveau de base par des cercles pleins). Pour les réseaux Gaussiens et afin de compenser pour la vitesse non constante, le niveau de couleur est le logarithme de la densité du passage des trajectoires dans un maillage de l'espace pour 200 simulations de 4 cycles chacune. Le grand attracteur en vert est associé à une faible variance, $\sigma_k = 4$, le plus petite attracteur en violet est pour une plus grande variance $\sigma_k = 20$. La taille du maillage est environ 0.05 mV le long de l'axe V et de 1 pA le long de l'axe w .

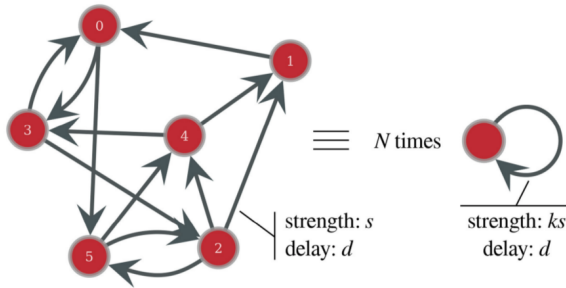


FIGURE 7.23 – **Schéma du modèle équivalent des bouffées synchronisées.** Représentation schématique de l'équivalence entre un réseau à degré fixe $k=2$ avec $N = 6$ neurones synchronisés de poids synaptique s avec un ensemble de N neurones indépendants possédant une boucle de rétroaction de poids $k \times s$.

seul système d'équations du modèle AEIF comme tous les neurones du système sont représentés par ce neurone « équivalent » unique :

$$\left\{ \begin{array}{l} \dot{V} = -(V - E_L) + e^V - w + I_e + \underbrace{\sum_j ksI_0 \cdot (t - t_j - d)\Theta(t - t_j - d)e^{-\frac{t-t_j-d}{\tau_s}}}_{I_{syn}(t)} \\ \tau_w \dot{w} = a(V - E_L) - w \end{array} \right. \quad (7.13)$$

Cette équation dynamique peut ensuite être traitée à l'aide de différentes approximations pour obtenir des expressions analytiques approchées de la dynamique. On contraint ensuite ces formules dans une équation auto-consistante exprimant la condition pour que les potentiels d'actions émis par le neurone durant une bouffée déterminent les bouffées suivantes (la manifestation de la boucle de rétroaction de la figure 7.23). La solution donne une description complète des propriétés dynamiques du neurone équivalent caractérisant la dynamique des bouffées synchronisées. On a pu appliquer cette approche à trois modèles différents de synapses (instantanées, continues et de forme alpha), aboutissant à trois équations auto-consistantes transcendantes qu'il est possible de résoudre numériquement pour obtenir les propriétés dynamiques des bouffées. Grâce à ces résolutions très rapides du modèle équivalent qui ne nécessitent pas la simulation de la dynamique sur un réseau de nombreux neurones on a pu calculer les propriétés dynamiques d'un grand nombre de jeux de paramètres et étudier la sensibilité du régime de bouffées par rapport aux 8 paramètres du modèle AEIF.

La caractéristique principale qui détermine la dynamique est la valeur maximale du courant d'adaptation w^* atteint à la fin d'une bouffée. En effet, dans tous ces cas le comportement a les caractéristiques d'un oscillateur à relaxation. Le courant w décroît lentement durant la phase de récupération puis augmente très rapidement durant la phase de bouffée jusqu'au seuil w^* qui détermine la fin de la bouffée et le démarrage d'un nouveau cycle. Suivant le modèle de synapse la valeur de w^* peut être déterminée via une équation auto-consistante et des approximations.

Synapses instantanées

Il est instructif de suivre la démarche dans le cas le plus simple qui peut être résolu avec un modèle des synapses instantanées représentées par des fonctions Dirac. L'évolution de la trajectoire sur l'attracteur dans l'espace des phases pour le cycle d'une bouffée est représentée dans la figure 7.24. Pour suivre l'évolution il faut prendre en compte qu'en dessous de la nullcline V les trajectoires sont orientées vers la droite (V croissant). Ici la nullcline w est en dessous en dehors de la figure, on se trouve au-dessus de celle-ci et les trajectoires sont verticalement vers le bas. On part d'un point sur la phase de récupération du neurone entre deux bouffées (image en haut à gauche). La trajectoire suit la nullcline V jusqu'à son minimum, où l'état n'est plus bloqué et V augmente, quand le potentiel atteint le seuil V_r le terme exponentiel fait diverger la variable produisant un potentiel d'action (la figure est coupée et ne montre pas le maximum du potentiel). Après l'émission d'un potentiel d'action, V est remise au niveau V_r (ligne verticale en tirets) et commence par décroître (partie gauche de la trajectoire) sous l'effet du courant de fuite $-g_L(V-E_L)$ et de l'adaptation $-w$, jusqu'au moment t^- de réception du potentiel post synaptique du réseau de l'émission de potentiel d'action qui vient d'avoir lieu (t^- est retardé du délai de transmission d).

À cet instant le potentiel est soudainement incrémenté par l'arrivée du courant synaptique à la valeur $V(t^+) = V(t^-) + kQ_s$, où Q_s est la charge correspondant à une entrée synaptique et k le degré ici fixe du neurone, position du point marron à droite de la trajectoire. Au même temps la variable d'adaptation w est incrémentée de b . Si $V(t^+)$ se trouve au-delà de la nullcline V alors le potentiel diverge à nouveau immédiatement, on a l'émission d'un autre potentiel d'action et la même phase va se répéter. La décroissance juste après l'émission et avant l'arrivée du courant synaptique (avec un retard d) va devenir de plus en plus grande à mesure que w augmente manifestant l'effet croissant des courants d'adaptation qui contribuent négativement \dot{V} . La bouffée continue avec la suite de l'émission de potentiels d'actions jusqu'à ce que w devienne plus grand qu'une limite w^* indiquée sur la figure par le point vert. À ce moment l'adaptation est tellement importante que la décroissance de l'état V du neurone après une remise au niveau V_r est telle, que le niveau atteint $V(t^-) = V_r - \Delta_{\text{leak}}$, incrémenté par l'arrivée du potentiel synaptique kQ_s , reste maintenant en deçà de la nullcline V . Il n'y a pas de nouvelle émission et le potentiel reprend à décroître pour atteindre et croiser la branche gauche de la nullcline V . Le potentiel V se retrouve alors sur la nullcline dans un état de quasi-équilibre et s'adapte instantanément à la diminution de w . L'état (V, w) glisse le long de la nullcline V , c'est la phase de récupération durant laquelle le courant d'adaptation diminue jusqu'au minimum de de la nullcline quand le même cycle pourra se répéter.

La valeur w^* quand l'état du neurone incrémenté de l'arrivée du potentiel post synaptique du réseau synchronisé ne dépasse plus la nullcline V sur la droite est donc la limite qui marque l'arrêt de la rafale de potentiels d'action qui compose la bouffée. Cette valeur limite se retrouve dans une équation auto-consistante pour w^* qui exprime d'un côté tout ce qui se passe après l'émission d'un potentiel : remise à niveau V_r , décroissance et réception du signal post synaptique, et de l'autre côté la valeur limite du potentiel le long de la nullcline $V_{\text{nullcline}}(w)$. La solution périodique d'un cycle d'activité doit donc satisfaire :

$$V_r + \Delta V_{\text{leak}}(w^*) + kQ_s = V_{\text{nullcline}}(w^*) \quad (7.14)$$

où $\Delta V_{\text{leak}}(w^*)$, négatif, représente la décroissance du potentiel entre l'instant juste après une émission de remise au niveau bas V_R et l'instant d (le retard de propagation) plus tard de réception du courant post synaptique par le reste du réseau (on rappelle que l'on suppose le réseau entièrement synchronisé et homogène, toutes les entrées arrivent au même temps). Ce terme est l'effet du courant de fuite et du courant d'adaptation, sa norme augmente avec w^* , g_L et d . Étant donné la forme de la branche de droite de la nullcline V dominée par l'exponentielle, on peut l'approximer comme verticale et considérer $V_{\text{nullcline}}(w^*) \sim \text{constante}$. L'équation (7.14) exprime donc une condition de balance entre V_r , kQ_s et la décroissance $\Delta V_{\text{leak}}(w^*)$. La taille d'une bouffées, c'est-

à-dire le nombre de potentiels d'actions émis, correspondra au nombre de potentiels d'actions émis pour que $\Delta V_{\text{leak}}(w^*)$ satisfasse l'équation (7.14). On peut donc utiliser cette relation pour déduire les propriétés des différents paramètres : plus V_r , Q_S et k sont élevés, plus d'émissions seront nécessaires pour faire en sorte que adaptation et fuite ($\Delta V_{\text{leak}}(w^*)$) compensent (c'est aussi le cas avec E_L qui a un effet antagoniste dans le courant de fuite). À contrario la bouffée sera plus petite avec moins d'émissions de potentiels si les contributions individuelles à la décroissance $\Delta V_{\text{leak}}(w)$ après chaque émission sont élevées permettant d'atteindre le seuil plus vite. Ceci est le cas notamment avec un délai d plus grand qui laisse plus de temps pour la décroissance, avec un courant d'adaptation b plus élevé, ainsi qu'une conductance de fuite g_L plus grande. Cet exemple reste simple et peut-être plus précisément développé, il figure dans l'annexe de [FARDET et collab., 2018a].

Ainsi l'équation (7.14) conduit à une expression pour w^* , d'où la dynamique complète du modèle peut-être déduite en considérant le comportement de relaxation de la variable d'adaptation w qui montre deux phases : une période de repos durant laquelle w diminue jusqu'à sa valeur minimale, et une période active durant laquelle w augmente rapidement jusqu'à son maximum. Ceci amène à des expressions pour les différentes phases de ce cycle conduisant aux principales valeurs caractéristiques de la dynamique telles que notamment la durée de la période de repos, le nombre de potentiels d'actions dans une bouffée, les intervalles de temps entre leurs émissions.

On peut remarquer que dans ce cas de synapses de type Dirac, instantanées, la nullcline V ne bouge pas, le courant synaptique arrive sur un intervalle de temps infinitésimal. Toute la dynamique peut-être suivie dans un diagramme des phases fixe. Ce n'est pas le cas général quand les entrées sur un neurone s'étalent dans le temps, d'une part parce que le signal d'une synapse possède une certaine durée (cas des synapses alpha), et d'autre part en raison de l'hétérogénéité du réseau et de la synchronisation imparfaite qui conduit à des entrées sur un neurone venant du reste du réseau étalées dans le temps. En conséquence le modèle AEIF de neurone possède un terme de courant en entrée I_e dû à ces entrées. Du fait de ce terme supplémentaire la nullcline V se trouve donc momentanément décalée verticalement tant que l'entrée est active et la dynamique du neurone à l'état (V, w) va dépendre de la position de ce point par rapport à la nullcline décalée.

7.3.3 Synapses continues

Le cas d'un modèle de synapse « continue » correspondant à une entrée I_S constante tout le long d'une bouffée correspond à cette situation. Le modèle de synapse « Dirac » précédent qui pouvait s'interpréter comme l'entrée parfaite d'un réseau exactement homogène et synchronisé. À l'opposé la synapse constante représente les contributions étalées sur toute la bouffée arrivant d'un réseau très hétérogène : les potentiels d'actions du réseau arrivent sur le neurone représentatif de manière uniforme tant que l'activité perdure. Je ne reproduirais pas ici le raisonnement pour cet autre cas extrême de synapse, il est traité dans l'article. Le principe de résolution est similaire au cas précédent à ceci près que la nullcline V se trouve décalée durant la bouffée (à titre indicatif voir seulement la figure 7.25).

Le modèle théorique développé fournit ainsi un cadre prédictif qui permet de déterminer comment les propriétés des bouffées collectives dépendent des différents paramètres des neurones. Comme les relations sont analytiques, donc très rapides à calculer il est possible de les utiliser pour prédire les propriétés des bouffées collectives pour un grand nombre de paramètres. Ceci a permis notamment de tester systématiquement l'influence sur la dynamique de chaque paramètre, figure 7.26.

7.3.4 Bouffées de réseau : conclusions et perspectives

La conclusion principale de ce travail sur la synchronisation des bouffées est d'avoir montré que dans une population de neurones intrinsèquement actifs l'adaptation suffit pour faire ap-

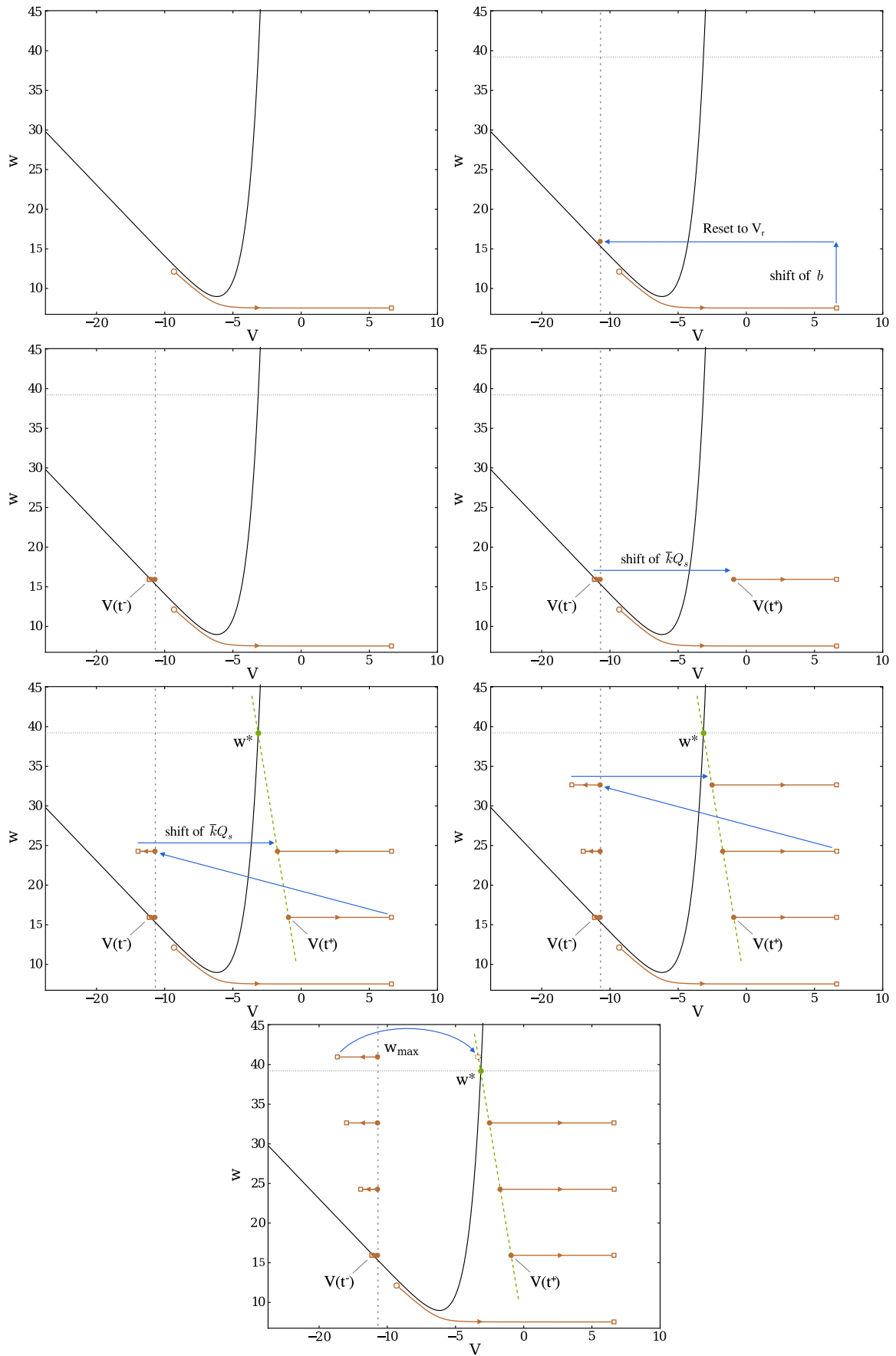


FIGURE 7.24 – Trajectoire dans l'espace des phases de l'attracteur des bouffées synchronisées. Description dans le texte.

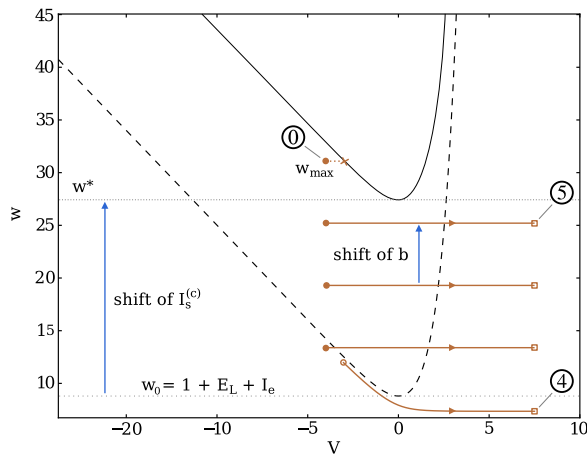


FIGURE 7.25 – **Trajectoire d'une bouffée pour des neurones synchronisés couplés par une synapse « continue »**. La bouffée commence quand le premier potentiel d'action est reçu (point 4). À ce moment un courant continu constant I_s est injecté dans le neurone ce qui décale la nullcline V (tirets) vers le haut (ligne continue noire). Le neurone émet jusqu'à ce que le dernier décalage b amène w au-dessus de w^* , à la valeur maximale w_{max} qui amène V à croiser la nullcline. Ceci marque la fin de la bouffée d'émissions et le début de la période de récupération (marquée 0).

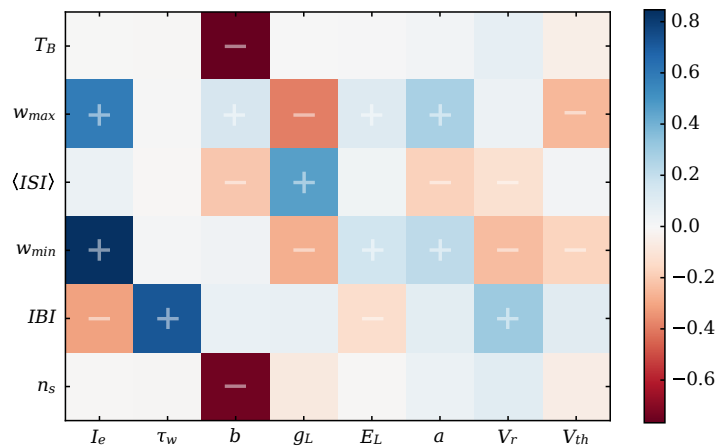


FIGURE 7.26 – **Matrice de corrélation pour les caractéristiques principales de la dynamique des bouffées en fonction des paramètres neuronaux du modèle AEIF**. Les paramètres de la dynamique sont T_B : la durée d'une bouffée, w_{max} : le courant d'adaptation maximal, $\langle ISI \rangle$ la valeur moyenne de l'intervalle de temps entre potentiels d'actions durant une bouffée, w_{min} : le courant d'adaptation minimal, IBI (Inter Burst Interval) : la durée entre deux bouffées, n_s : le nombre de potentiels d'actions durant une bouffée.

paraître deux échelles de temps dans la dynamique collective avec de brèves bouffées d'activité suivies de plus longues périodes. Les neurones individuels n'ont pas besoin d'avoir en propre un comportement par bouffée d'activité (type « intrinsically bursting » ou « chattering » [IZHIKEVICH, 2006]), ni la présence de neurones inhibiteurs, ni la plasticité synaptique. Des travaux récents ont montré comment un phénomène de plasticité à court terme comme la dépression des neurotransmetteurs dans les terminaux pré-synaptiques sont suffisants pour la génération de bouffées de réseau [COHEN et SEGAL, 2009; DAO DUC et collab., 2015]. L'adaptation étant une condition suffisante pour les bouffées, ces autres effets ne sont pas incompatibles.

L'effet de la taille des cultures est un facteur insuffisamment pris en compte dans les études des dynamiques de bouffées de réseau en culture. La création de cultures sur de grandes surfaces, et plus encore peut-être la difficulté à observer avec une résolution suffisante un échantillon au-delà de quelques millimètres carrés explique cette limitation. Or les dimensions physiques des populations en cultures affectent très probablement le comportement collectif. La communication entre neurones fait en effet intervenir des délais de transmissions liés à la propagation des potentiels d'actions dans les neurites. La durée de ces délais doit affecter la dynamique synchronisée car la perception de courants post-synaptiques fortement retardés par un neurone aura lieu quand celui-ci aura eu le temps de revenir à un état plus faiblement excitable. On peut donc s'attendre à une limite de l'extension spatiale pour l'existence de bouffées fortement synchronisées. Ceci pourrait expliquer des observations dans de grandes cultures de zones privilégiées d'initiation et de fronts de propagation d'activité [ORLANDI et collab., 2013]. À l'échelle de petits réseaux, d'autres

travaux ont montré, une déstabilisation de la synchronisation [YAMAMOTO et collab., 2016] qui est en cohérence avec la sensibilité à l'inhomogénéité des bouffées montrée dans notre analyse.

Une question centrale du traitement théorique des bouffées de réseau de notre publication est l'hypothèse de départ motivée par les expériences de PENN et collab. [2016] de la présence d'une proportion dominante de neurones intrinsèquement actifs périodiquement (dans notre étude pour nous avons considéré une population intégralement d'oscillateurs) et dont l'activité propre entraîne la dynamique collective. Or la source et la nature de l'activité spontanée des neurones dans les cultures est encore incertaine. Les expériences de PENN et collab. [2016] se distinguent par le contrôle très précis des conditions de cultures des neurones qui ont conduit à des conclusions inédites de leur nature oscillatoire. Des propriétés supposées auparavant [GRITSUN et collab., 2010] mais non observées dans d'autres conditions à cette échelle.

Le facteur expérimental clé de ces travaux a consisté dans l'utilisation durant les mesures d'un milieu chimiquement contrôlé permettant des dosages précis d'ions divalents connus pour intervenir dans les échanges ioniques durant les déclenchements neuronaux ainsi que pour le fonctionnement synaptique. Des conditions qui permettent le contrôle continu du couplage neuronal, mais qui peuvent affecter aussi affecter les propriétés naturelles des neurones, notamment dans la limite d'absence de couplage $[Ca^{2+}] = 0M$. Leurs conclusions biologiques peuvent donc encore être débattues, en particulier l'existence de la fraction très significative de neurones naturellement oscillants. Cette population est-elle encore oscillante dans les conditions avec couplage, ainsi que dans des conditions physiologiques ?

Ainsi, dans des expériences sur des cultures de neurones dissociées publiées la même année visant aussi à étudier les conséquences sur l'activité spontanée de changements sur le couplage des neurones SURESH et collab. n'ont pas obtenu les mêmes conclusions. Ces auteurs ont opéré dans des conditions usuelles avec un milieu de culture neuro-basal standard pour les mesure à l'aide de matrices d'électrodes, et des moyens pharmacologiques pour inhiber sélectivement les différents types de récepteurs synaptiques. Comme attendu ils ont aussi relevé les bouffées synchronisées caractéristiques, avec un comportement moins régulier que PENN et collab. [2016], mais surtout sans indication d'oscillations intrinsèques des neurones. Les conditions expérimentales employées par Suresh *et al.* sont fortement différentes de celles de PENN et collab. [2016]. Si elles sont effectivement plus usuelles, par utilisation de sérum comme milieu de cultures, elles n'en restent pas moins également artificielles.

Ces deux exemples illustrent la perplexité que l'on peut avoir au contact de la littérature sur les cultures neuronales. Les résultats montrent une variété de comportements déconcertante. Si les bouffées de réseau sont bien un comportement toujours observé, les propriétés de celles-ci et de l'activité globale sont considérablement différentes. Suivant les publications la dynamique des bouffées peut montrer des schémas très différents par exemple lors de la croissance et de l'explosion de l'activité d'une part, lors de la décroissance de l'activité d'autre part. Des interrogations aussi sur la présence de zones ou de neurones privilégiés trouvés associés à l'initiation des événements. Les comportements se distinguent aussi dans l'existence d'activité intermédiaire entre les bouffées collectives, des événements associés à des avalanches que plusieurs groupes associent aux manifestations de comportements critiques auto-organisés. Mais que d'autres groupes n'observent aucunement! (Voir la discussion à la section 6.5.2 du chapitre précédent.) Il est très probable que les détails des conditions expérimentales jouent alors un rôle déterminant majeur dans le comportement des cultures. Une situation qui met en évidence toute la prudence nécessaire pour extrapoler des conclusions obtenues en culture vers des interprétations pour des situations *in vivo*. Il manque un modèle expérimental établi pour les cultures reproduisant des conditions physiologiques vraisemblables. En vue des applications d'ingénierie de processeurs de calcul à base de dispositifs neuronaux la reproduction de conditions physiologiques n'est cependant pas une nécessité. Par contre une bonne connaissance du fonctionnement et des conditions d'utilisation des cultures serait précieuse pour contrôler les systèmes.

7.4 Développement neuronal et croissance de réseau

Dans cette section je présente une orientation initiée lors de la thèse de Renaud Renault et poursuivie avec celle de Tanguy Fardet concernant la modélisation et la simulation de la croissance de neurones dans des chambres microfluidiques. De tels dispositifs, parfois appelés puces neuronales, sont en effet des pistes prometteuses pour construire avec des cultures neuronales de modèles sophistiqués de pathologies cérébrales ou encore pour servir de plateformes d'études sur le contrôle d'activités neuronales.

La construction de puces neuronales repose actuellement en grande partie sur des démarches empiriques dans lesquels des designs divers sont expérimentalement testés pour valider les schémas de connectivité obtenus. Il n'existe pas à l'heure actuelle de modèles et d'outils permettant une ingénierie de ces systèmes par l'étude théorique préalable de plans de conceptions.

L'objet de l'axe de recherche présenté ici est bien d'établir les connaissances théoriques nécessaires à ces études, ainsi que des outils informatiques de simulation. Cet objectif implique de progresser dans la compréhension de la biologie et de la biophysique du développement des neurones en interaction avec leur environnement et la croissance des réseaux. En effet, la conception de dispositifs peut-être très contre-intuitive comme l'illustre par exemple le design infructueux de la figure 7.27 testé lors de la thèse de Renaud Renault. Une meilleure compréhension des interactions entre cônes de croissance et obstacles, ainsi qu'entre cônes de croissance et axones s'avère indispensables pour guider la conception des systèmes.

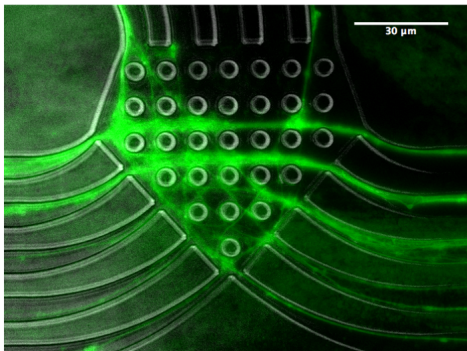


FIGURE 7.27 – **Tentative de design pour un mélangeur de connexions neuronales.** Les axones de deux populations distinctes arrivent via les canaux latéraux à partir de deux populations de neurones distinctes, les canaux sur le haut visaient à recueillir les prolongations d'axones d'origine mixée entre les deux chambres. Contrairement aux attentes les axones ne prennent que rarement cette orientation cible et plutôt connectent directement les neurones des chambres initiales. Source : [RENAULT, 2015, chapitre 1].

Durant la thèse de Renaud Renault nous avons étudié cette question en collaboration avec Jean Louis Viovy et Catherine Villard à l'Institut Curie où des expériences avec des captures vidéo de la croissance de neurones corticaux de rats dans des micro-environnements ont pu être réalisées. Hormis les résultats préliminaires dans cette thèse ces travaux n'ont pas encore fait l'objet de publication d'article. Les idées de bases ont en effet été reprises par Tanguy Fardet dans sa thèse en cours pour un raffinement des modèles de croissance et le développement d'un logiciel de simulation performant.

Comme discuté au chapitre précédent 6, section 6.7.1, il n'existe donc pas à l'heure actuelle d'un ensemble cohérent de modèles et de système informatique pour simuler la croissance de morphologie de neurones et de réseau directement exploitable et permettant l'exploration des effets de paramètres physiques ou biologiques sur les structures. Motivés pour ce qui nous concerne par la recherche d'une meilleure compréhension des réseaux dans les cultures neuronales microfluidique, et l'indisponibilité d'outils pour ces études nous avons démarré dans l'équipe Neurophysique à MSC un projet de modélisation et de code de simulation de ces processus de développement.

Ingrédients Il est certainement trop précoce encore de prétendre à un modèle biophysique pour tous les aspects du développement des neurones et de l'établissement du réseau. Il est plus raisonnable d'envisager alors une combinaison d'approches empirique et biophysique pour construire

une simulation intégrant les modèles disponibles de processus biologiques à des algorithmes de processus effectifs quand des représentations de mécanismes subjacents n'existent pas. Cette combinaison sera déjà un progrès utile par rapport à des algorithmes purs de génération de formes, notamment si le cadre informatique est modulaire et permet l'évolution des règles par type de processus.

L'enjeu est alors de construire un cadre flexible permettant de rassembler des modèles et de les faire fonctionner dans une simulation numériquement performante. Les principaux processus intervenant dans le développement des neurones sont ceux impliqués dans :

- l'initiation des neurites, avec la rupture de symétrie de la cellule *grosso modo* sphérique initiale. Il s'agit de rendre compte des mécanismes de polarisation du neurone avec l'apparition des premières branches dendritiques et l'établissement de l'axone par différenciation d'une dendrite.
- l'élongation des neurites et leur rétractation, qui dépendent de l'éloignement au soma, de la position des branches dans l'arbre dendritique, de leur diamètre. Donc l'ensemble des mécanismes déterminant la dynamique d'élongation des neurites, mais également l'arrêt de la croissance et éventuellement la rétractation des branches.
- le branchement de neurites. Les facteurs qui amènent les cônes de croissance à se dédoubler et ceux amenant à un nouvel embranchement apparaissant le long d'un segment de neurite déjà formé.
- l'orientation des cônes de croissance. Donc les mécanismes par lesquels les têtes des neurites naviguent en interagissant avec l'environnement physique et chimique. Les cônes de croissances sont sensibles à des marqueurs chimiques, éventuellement relâchés par les neurones eux-mêmes [SEGEV et BEN-JACOB, 2000], et doivent alors intégrer un système permettant la détection de gradients chimiques.
- l'établissement et la maturation des synapses. Des synapses sont possibles quand des sites présynaptiques et postsynaptiques sur des neurites différents sont suffisamment proches. Un certain pourcentage de ces sites de proximité formeront des synapses fonctionnelles, estimés *in vivo* à 25% des sites potentiels selon [STEPANYANTS et CHKLOVSKII, 2005]. Un pourcentage qui pourrait être significativement moindre dans des cultures en raison de l'organisation à 2 dimensions, et de la plus faible densité de neurones, mais dont les mécanismes d'établissement restent mal connus.

Un modèle général de croissance d'un réseau doit rendre compte de ces différents phénomènes biologiques et biophysiques et comporter des règles empiriques ou des modèles des mécanismes impliqués. Il s'agit donc d'un projet de modélisation multi-échelle et multi-physique que nous avons envisagé sur le long terme pour combiner les différents phénomènes impliqués dans la morphogénèse des neurones dans des milieux contraint géométriquement et en interaction.

Ce projet a démarré dans la dernière partie de la thèse de Renaud Renault qui en a fait l'exploration préliminaire et une première approche empirique. Il s'est poursuivi durant la thèse de Tanguy Fardet pour donner des fondements solides à ces développements aussi bien sur le plan informatique d'un code de simulation efficace que pour l'implémentation de modèle biophysiques. Dans notre cadre d'intérêt actuel portant sur les cultures on focalise les modèles en développement et le simulateur à la croissance de neurones répartis sur un plan dans une couche plane d'épaisseur réduite. Une géométrie quasiment bidimensionnelle mais dotées d'obstacles matérialisant les structures microfluidiques des dispositifs complexes de cultures qui nous intéressent. Aussi aux processus listés ci-dessus nous faudra t'il inclure les interactions des cônes de croissance avec des obstacles physiques. Dans la section ci-dessous je présente les principales étapes accomplies dans cette thématiques et l'avancement actuel.

7.4.1 Modèle empirique de Renaud Renault

Pour disposer rapidement de réseaux réalistes en support de nos études de dynamique d'activité d'une part, et d'autre part pour offrir un outil pratique à nos collaborateurs de l'Institut Curie développant de nouveaux designs de puces neurones, notre premier objectif visait davantage la construction d'un modèle empirique utilisable plutôt que l'implémentation de modèles biophysiques microscopiquement justifiés mais de développement à plus long terme.

Le modèle construit durant la thèse de Renaud Renault [RENAULT, 2015] s'est ainsi focalisé sur l'élongation d'un axone, sans embranchements et consisté fondamentalement dans une marche aléatoire dirigée avec des règles d'interaction. Une démarche empirique rendue aussi possible par la réalisation d'expériences de croissance dédiées permettant une paramétrisation à partir d'ajustements des paramètres du modèle avec des observations [RENAULT, 2015] chapitre 2. Le modèle inclut les interactions des cônes de croissance avec des obstacles ainsi qu'avec d'autres axones, permettant d'approcher les interactions entre neurones dans la croissance des réseaux. Renaud Renault a également réalisé à l'Institut Curie dans l'équipe de Jean-Louis Viovy des séries d'expériences qui ont servi à déterminer les paramètres d'interactions du modèle décrit ci-dessous. Ces mesures ont exploité des dispositifs microfluidiques conçus spécifiquement pour isoler des axones en élongation et pouvoir les suivre dans des enregistrements vidéos.

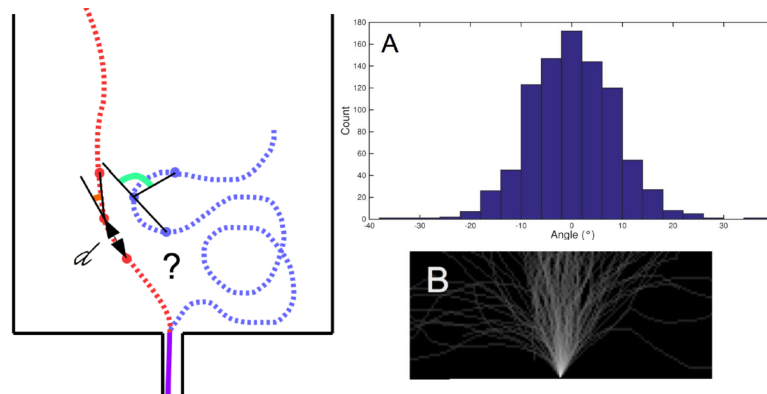


FIGURE 7.28 – **Croissance d'axones non contraints.** *Gauche* : Schéma indiquant comment un segment initial d'axone (en violet) se comporte différemment (segments en rouge ou en bleu) suivant la distribution angulaire des angles entre segments successifs. *En haut à droite* : A- Distribution angulaire normalisée des changements de direction entre segments successifs déduits à partir d'axones libres en croissance. *En bas à droite* : B- Superpositions des images des traces de croissance des expériences de croissance libre utilisées pour calculer la distribution angulaire en A (traces de 105 axones). Source : [RENAULT et collab., 2015]

Le modèle d'élongation d'un segment après l'autre consiste ici dans une marche aléatoire avec une longueur de persistance. Cette longueur de persistance est liée à la rigidité de l'axone venant essentiellement de son cytosquelette. En absence d'obstacles dans un rayon donné l'orientation du nouveau segment est tirée aléatoirement d'une distribution gaussienne de probabilité centrée sur l'orientation du segment précédent. La largeur de cette distribution définit la marche aléatoire orientée et reflète la longueur de persistance de la trajectoire d'un axone en croissance sans obstacles. La croissance libre d'un axone est ainsi modélisée itérativement par l'ajout d'un nouveau segment suivant une direction résultant d'un tirage aléatoire. Chaque axone correspond à une suite de segments de $3 \mu\text{m}$, plus petits que la longueur de persistance typique des axones.

La manière avec laquelle un cône de croissance sonde l'environnement est biologiquement très complexe [ROTH et collab., 2014] et difficile à reproduire dans une simulation générale. On choisit pour l'instant de condenser tous les processus biologiques (adhésion focale, traction, ...) et simplifier les interactions avec l'environnement comme un processus à deux modes : les obstacles peuvent attirer ou repousser les cônes de croissance. Un cône actif a une probabilité plus élevée de s'orienter vers l'obstacle attractif le plus proche dans une zone d'exploration, que de choisir une autre orientation. Le contraire pour des obstacles répulsifs. La force de l'attraction ou de la

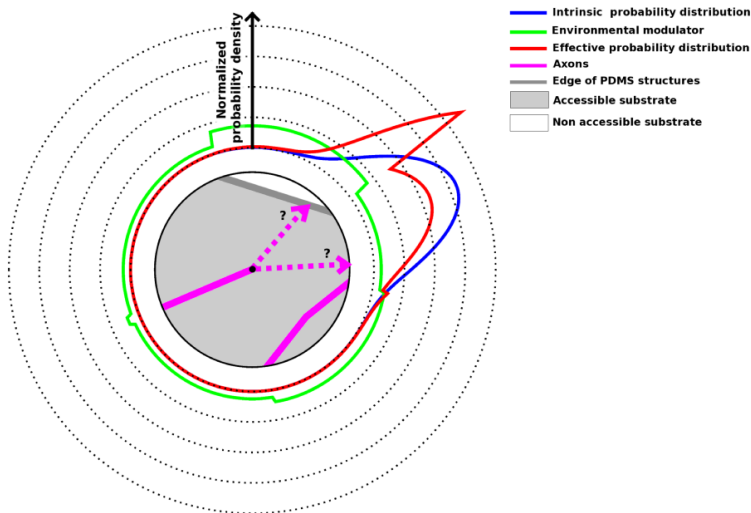


FIGURE 7.29 – **Modélisation de la croissance d'un axone dans un milieu hétérogène.** La direction d'un nouveau segment d'axone est tirée au hasard à partir d'une distribution de probabilités qui tient compte de l'affinité du cône de croissance avec les conditions dans chaque direction. Notamment dans le modèle implémenté l'affinité avec les murs des chambres de cultures et d'autres axones. Source : [RENAULT et collab., 2015]

répulsion peut-être paramétrée pour exprimer des interactions avec différents types d'objets.

La distribution des angles entre segments successifs a été déterminée à partir d'observations dans des cultures dans des expériences avec des enregistrements dans le temps des cheminements de neurones libres. R. Renault a ainsi déterminé dans sa thèse qu'en absence d'obstacles les angles d'orientation suivent une distribution angulaire Gaussienne (figure 7.28) avec une moyenne nulle et un écart-type de $8,2^\circ$.

Si l'élongation de l'axone s'opère dans un milieu hétérogène, les interactions entre la tête de croissance et le milieu doivent être prises en compte. Des obstacles, un changement d'environnement chimique, ainsi que de la présence éventuelle d'autres neurones vient moduler la probabilité d'orientation des nouveaux segments d'élongation. Ces influences sont capturées dans une modulation de la distribution angulaire des nouvelles directions suivant l'environnement le long de chaque orientation (figure 7.29). Ce modèle rappelle la biologie de navigation d'un cône de croissance avec l'excroissance de filopodes qui sondent le voisinage de l'extrémité et déterminent la direction d'avancement par rapport à leur accroche avec le substrat ou des obstacles.

La distribution des orientations est mise à jour à chaque étape d'avancement en fonction de l'environnement dans le voisinage de la position terminale. Cette distribution capture ainsi les perquisitions qu'effectuent les protrusions des cônes de croissance de neurones avant de consolider la direction de leur avancement. Elle sera modulée par l'environnement un facteur en fonction de l'accessibilité et des objets dans les différentes directions. Cette modulation représente les interactions et est paramétrée par des facteurs représentant les affinités avec les différents types d'obstacles, parois des structures d'une part et axones d'autre part. En multipliant la distribution angulaire de l'élongation libre par les facteurs de modulation environnementaux à chaque étape on obtient la distribution de probabilité effective selon laquelle est choisie aléatoirement la direction d'élongation du nouveau segment.

Les observations de croissance dans des chambres microfluidiques gravées montrent que les axones ont une tendance à suivre les contours des obstacles, ainsi qu'à se regrouper en filaments. ce dernier phénomène est connu sous le nom de fasciculation et est observé dans de nombreuses situation également *in vivo*. Ces deux manifestations sont les facteurs environnementaux affectant l'élongation pris en compte à ce stade. Ils sont paramétrés par respectivement une constante d'affinité des bords ϵ , et par une constante de fasciculation φ (Fig. 7.30).

La constante d'affinité de la bordure a été calculée en laissant les neurones croître en présence d'obstacles avec des murs en dièdre de différents angles. Il apparait que les axones atteignant une bordure la suivent très préférentiellement y compris le long de coins quand ceux-ci ne sont pas trop aigus, dont l'angle est supérieur à un angle critique estimé à 135° . Cet angle limite permet la

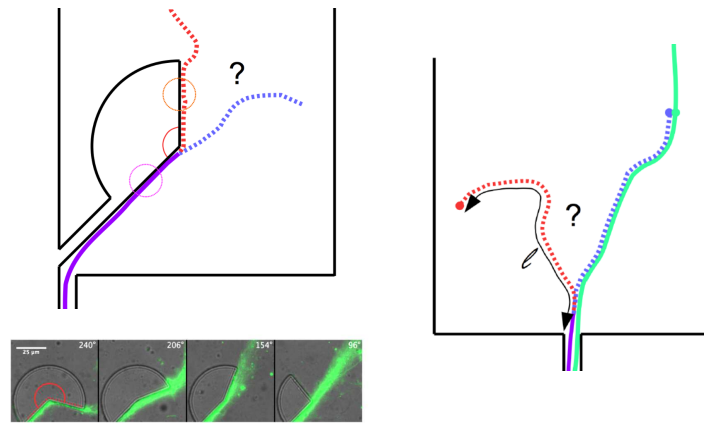


FIGURE 7.30 – **Affinité des cônes de croissance aux bordures d'obstacles.** *En haut à gauche* : Schéma montrant comment un segment d'axone initiale (en violet) se comporterait différemment (segments rouges ou bleus) suivant la valeur de la constante d'affinité des bords ϵ . Le coin angulaire d'un obstacle est entouré d'un cercle rouge. Les deux régions avant et après ce coin d'obstacle sont indiquées avec des cercles rose et orange respectivement. *En bas à gauche* : résultats expérimentaux montrant le comportement d'axones en croissance autour des coins d'obstacles avec différents angles. Au delà d'un angle critique, ici entre 54° et 96° , la grande majorité des axone cesse de suivre la bordure. *Droite* : Fasciculation des axones. Schéma représentant comme le segment initiale (en violet) d'un second axone arrivant dans la chambre à partir du bas se comporterait différemment suivant la force de la constante de fasciculation.

détermination du paramètre d'affinité avec les bordures ϵ . La constante de fasciculation a été déterminée quant-à elle à partir des observations de 30 paires de neurones initialement en contact : dans $66.7 \pm 8.6\%$ des cas le contact s'est poursuivi au-delà d'une distance de $20 \mu\text{m}$. Le paramètre de fasciculation a été ajusté à l'aide de simulations pour obtenir une probabilité équivalente dans le modèle. Enfin, en modulant chaque direction possible pour la progression d'un cône de croissance par les facteurs d'affinité des bordures et de fasciculation suivant le principe indiqué dans la figure 7.29 on obtient ainsi la probabilité effective pour l'élongation des axones.

L'implémentation de ce modèle dans un code simple Matlab a été validée par des comparaisons avec des situations expérimentales, comme celle de la figure 7.31. Dans cette expérience un obstacle asymétrique oriente la croissance des axones dans deux demi plans différents. Les trajectoires de 10 axones ayant pénétré la chambre de croissance par le bas sont relevées par fluorescence pour quantifier la localisation des trajectoires de part et d'autre de l'obstacle en fonction de son inclinaison. La simulation numérique de ces expériences conduit à la même répartition angulaire asymétrique des trajectoires en fonction de l'inclinaison de l'obstacle qu'observée dans les expériences en culture (figure 7.31B et C). On note cependant une différence d'apparence entre l'image expérimentale en A) et la simulation des trajectoires en C). Dans l'expérience montrant des cultures de 2 semaines, les trajectoires sont plus rectilignes que ce qui ressort du modèle. Dans le vieillissement de la culture on observe en effet une rectification des neurones, sous l'effet probable de forces de tensions, qui n'est pas prise en compte à ce stade dans le modèle. L'aplatissement des trajectoires des axones n'affecte cependant pas la répartition asymétrique de leurs trajectoire étudiée dans l'expérience.

Une première exploitation de ce simulateur a donné lieu par la suite à la conception non intuitive d'un nouveau design de « diodes » à neurones permettant de discriminer plus précisément les sens de connexion entre deux populations neuronales [RENAULT et collab., 2016] tirant parti de l'angle critique de suivi des bordures le long de la croissance introduit plus haut.

En dépit de ce succès et de la similitude de comportements de croissance simulées et observés, cette première modélisation est cependant restée limitée et n'a pas fait l'objet d'un article en l'état. Il s'agissait en effet encore d'un modèle très simple, sans implémentation d'arbres dendritiques ni de branchements des neurites et sans établissement de connexions synaptiques. Outre l'exigence

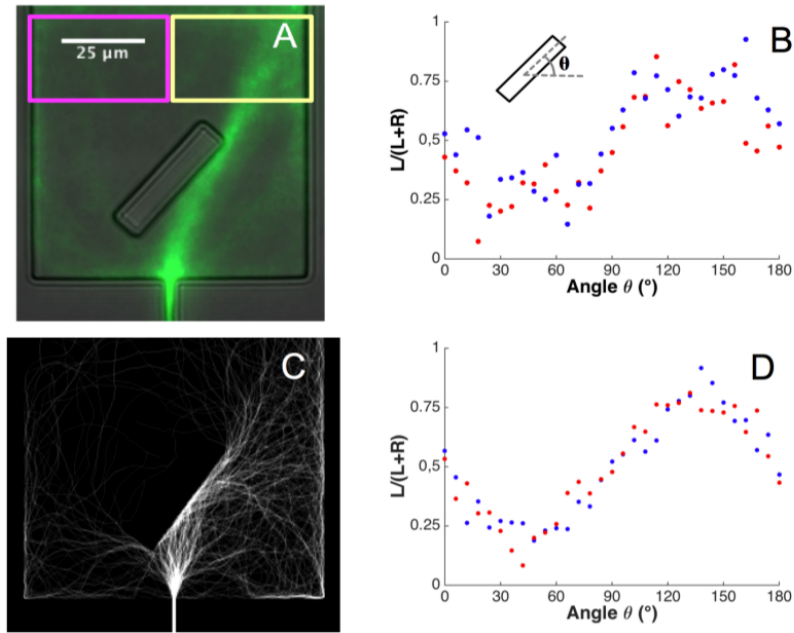


FIGURE 7.31 – **Test du modèle de croissance : séparation asymétrique.** A : En vert, fluorescence moyenne sur 10 expériences marquant le chemin des axones en interaction avec un obstacle incliné de 50° après 2 semaines de croissance. B : Fluorescence relative dans le demi plan gauche en fonction de l'angle θ d'inclinaison de l'obstacle. En bleu les points du rapport $\frac{L}{L+R}$ pour les 31 inclinaisons et en rouge le complémentaire symétrique horizontal ($1 - \frac{R}{L+R}$). C : Image résultat de la simulation du modèle de croissance montrant les chemins des axones. D : Profil de répartition dans les deux demi plans des trajectoires extrait des simulations (10 simulations de 30 neurones). Source : [RENAULT, 2015]

supplémentaire de modélisation, l'inclusion de ces éléments dans une approche plus complète pose également des exigences fortes sur l'efficacité numérique des codes de calcul en raison de l'augmentation des objets actifs et des interactions durant le développement du réseau. L'extension du modèle dans ce sens et le développement d'un code de simulation à haute performance ont fait ainsi partie du projet de thèse de Tanguy Fardet et est actuellement en voie de conclusion.

7.4.2 Modèle de croissance étendu et simulateur

La modélisation et la simulation de la croissance d'un réseau neuronal, même dans la situation particulière d'une culture plane, fait intervenir de nombreux processus biologiques et physiques. Plusieurs modèles doivent être combinés et mis en œuvre dans un code de calcul efficace permettant de réaliser des simulations de dimension suffisante avec les ordinateurs disponibles. Étant donné le nombre de cellules et d'éléments en interaction dans les cultures neuronales, on imagine aisément que leur simulation est numériquement très exigeante. Il est donc indispensable de trouver des implémentations algorithmiques des modèles biophysiques très performantes pour permettre une modélisation et des simulations réalistes. Se basant sur les résultats encourageants de la thèse de Renaud Renault on a donc entrepris un travail d'une part de nature d'informatique scientifique et d'autre part de modélisation des processus avec un souci d'optimisation calculatoire. Dans le cadre de sa thèse Tanguy Fardet a ainsi mis en place un framework de codage du simulateur selon des méthodes adaptées pour du calcul haute performance sur des superordinateurs et évolutif pour pouvoir modifier ou adjoindre de nouveaux modèles des différents processus pris en compte. D'autre part, avec notamment le travail de master « Tesi di Laurea » d'un étudiant de l'université de Rome Alessio Quaresima, plusieurs modèles de processus impliqués dans la croissance ont été repris, réanalysés et optimisés pour leur mise en œuvre efficace dans le simulateur. Certains modèles de la littérature ont ainsi été repris mais optimisés d'un point de vue algorithmiques. C'est notamment le cas de tous les processus dépendant d'un taux pro-

babiliste par unité de temps comme, par exemple, le modèle de branchements de cônes de croissance (voir plus bas). L'implémentation naïve de tels processus est d'évaluer à chaque intervalle de temps successif la réalisation ou non de chaque processus probabiliste. De cette façon cependant le nombre d'opérations augmente avec le nombre d'agents actifs dans la simulation et dépend également de la résolution temporelle. L'optimisation numérique mise en place à chaque fois que possible a consisté à inverser la probabilité et à déterminer la distribution des intervalles de temps avant un prochain événement. Le tirage à partir de cette distribution des temps d'attente avant une nouvelle action réduit considérablement les opérations nécessaires et les paramètres de simulation devient également indépendants de la résolution temporelle. Je décrirais ici seulement brièvement certains de ces développements très récents de la modélisation pour lesquels des publications sont encore en préparation.

Les processus pris en compte à ce stade sont d'une part la navigation des cônes de croissance dans l'environnement, et d'autre part le développement des neurites avec la création et la dynamique de branchements. Les neurones modélisés ici aspirent en effet à reproduire des formes aussi réalistes que possibles avec le développement de plusieurs neurites, axone et dendrites, pouvant se ramifier.

Modèle de navigation libre

On maintient comme modèle de navigation libre le principe d'une marche aléatoire avec une longueur de persistance que l'on avait déjà utilisé. Cependant plusieurs processus de ce type sont possibles, avec différentes manières de mener une déambulation au hasard qui peuvent amener à des trajets différents, ou bien à des implémentations algorithmiques plus ou moins efficaces pour la même allure. Alessio Quaresima dans sa thèse de laurea a consacré beaucoup d'efforts à évaluer différentes variantes et leur pertinence pour la modélisation des neurites. Il s'agit de processus aléatoires non Markoviens car chaque étape doit tenir compte de l'histoire précédente.

Notre première implémentation pour la navigation des cônes de croissance correspond essentiellement à une marche aléatoire corrélée (correlated random walk). Ce processus simple conduit à des formes de trajectoires plausibles, mais le modèle souffre de plusieurs difficultés. Il ne tient pas compte en premier lieu de variations de vitesse d'élongation qui sont observées expérimentalement et ne décrit pas une évolution réaliste dans le temps. Aussi, le lien entre les paramètres du modèle de croissance et une valeur de longueur de persistance ne peut être ajusté que par simulation numérique. L'ajustement des paramètres s'avère en outre être très délicat : la longueur de persistance varie exponentiellement avec certains paramètres rendant difficile leur détermination. Outre cela le modèle de marche aléatoire corrélée possède une dépendance non linéaire avec la résolution temporelle qui le rend difficile d'utilisation.

Il a fallu donc repenser le processus stochastique pour l'élongation et chercher à utiliser la longueur de persistance l_p directement comme paramètre. La solution plus convaincante à ce stade a consisté dans l'implémentation d'un processus « run and tumble » comme ceux caractérisant la chemotaxie de certaines bactéries [ALLENA et AUBRY, 2012; PAOLUZZI et collab., 2013] où la cellule progresse selon des sections rectilignes (« run ») interrompues par des événements de réorientation appelés « tumble » (culbute). Dans ce modèle le cône de croissance possède une certaine probabilité de tourner dépendant uniquement de la distance l parcourue : $P_{\text{turn}}(l) = \frac{l}{\lambda}$.

Ce modèle est appelé « uniform turning rate » (UTR) et produit des segments droits de longueur distribués exponentiellement avec une moyenne λ :

$$P_{\text{straight}}(l) = \frac{1}{\lambda} \exp(-l/\lambda) \quad E[l] = \lambda \quad \text{Var}[l] = \lambda^2 l_p = \frac{2\lambda}{\sigma}$$

Dans ce modèle la longueur de persistance est un paramètre et est indépendante d'autres éléments de la simulation comme la vitesse de croissance ou la résolution temporelle. Quelques exemples (incluant le branchement décrit ci-dessous) sont présentés dans la figure 7.32. Les formes

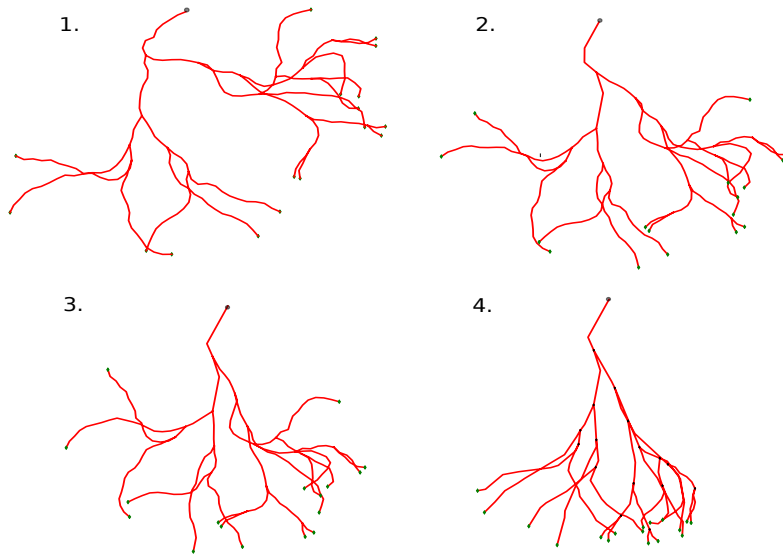


FIGURE 7.32 – Neurites générées par le modèle « run and tumble ». Les 4 images présentent un neurone généré en augmentant la longueur de persistance.

générées, à deux dimensions, semblent pouvoir reproduire des morphologies de neurones réels. On reviendra sur les validation effectuées après la discussion de la modélisation du branchement des neurites qui détermine également la morphologie.

Interaction avec l'environnement

L'interaction avec l'environnement reprend le modèle de Renaud Renault discuté précédemment. L'implémentation dans le nouveau simulateur par Alessio Quaresima est un algorithme optimisé notamment par rapport au nombre de directions (filopodes) et la largeur angulaire de la zone d'exploration. Contrairement à l'élongation dans un milieu libre déterminée uniquement par la longueur de persistance, en présence d'obstacles les cônes peuvent se retrouver dans des géométries confinées d'où il peut être difficile de sortir. La performance numérique a également été améliorée par l'utilisation de bibliothèques de géométrie performantes afin de déterminer à moindre coût la présence d'obstacles en dessous d'un rayon donné de sondage. À ce stade le nouveau simulateur n'implémente cependant pas la fasciculation. La parallélisation du code de calcul qui permet de répartir la simulation des différents cônes de croissances sur des processeurs différents complique très sérieusement la prise en compte d'un environnement pouvant changer dynamiquement, comme l'apparition de nouveaux obstacles venant de la poussée d'autres neurites. Pour des situations moins sensibles à ce phénomène la nouvelle implémentation reproduit cependant bien, avec une efficacité numérique considérablement supérieure, les observations et expériences déjà testées durant la thèse de Renault Renaud (figure 7.33).

Modèles de branchement des neurites

Le progrès principal de la nouvelle modélisation avec les travaux de thèse de Tanguy Fardet et la « laurea » de Alessio Quaresima est la prise en compte essentielle du branchement des neurites. L'arborescence des axones et neurites est en effet leur caractéristique morphologique principale déterminant leurs capacités opérationnelles. Le branchement des neurites est un phénomène dynamique qui contribue aussi à la plasticité du réseau neuronal avec la possibilité d'apparition de nouvelles branches ou de disparition. Les mécanismes biologiques et physiques qui déterminent ou contribuent à la création et l'évolution des branchements sont très complexes et encore insuffisamment connus. En conséquence les approches de modélisation restent en grande partie très phénoménologiques, certains mécanismes mieux caractérisés pouvant être explicitement pris en

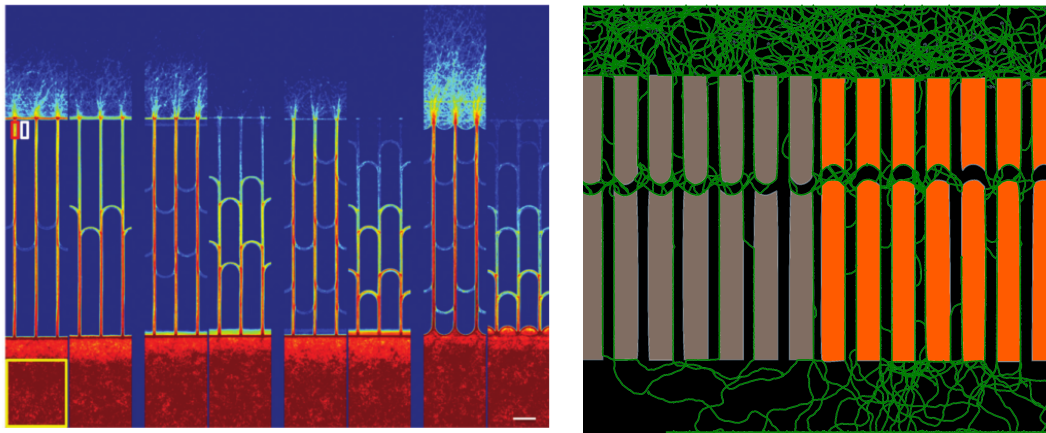


FIGURE 7.33 – **Schéma de construction d'une diode neuronale en « U » et simulation.** Gauche : dispositif à diode implémenté par nos collaborateurs à l'Institut Pierre Gilles de Gennes mis au point avec le modèle de croissance de Renaud Renault [RENAULT et collab., 2016]. Des canaux parallèles reliant les chambres de culture du haut et du bas sont connectées par des canaux transversaux en forme de « U » selon 2 orientations possibles. En raison de l'adhésion préférentielle aux parois les axones ont tendance à suivre la courbure progressive d'un canal en « U » s'il se présente. En rouge (fausses couleurs) les neurones marqués par fluorescence qui sont ensemencés uniquement dans la chambre inférieure. Suivant l'orientation de canaux transversaux les axones en croissance à partir du bas parviennent à la chambre supérieure ou bien retournent vers le bas. De gauche à droite le nombre de connections en « U » transversales augmente ainsi que la sélectivité du dispositif. Source : [RENAULT et collab., 2016]. Droite : Simulation des diodes neuronales en « U » avec le nouveau simulateur. En orange diodes haut vers bas. En gris : diodes bas vers haut. Sous l'effet de l'attraction avec les parois la plupart des axones dans la chambre du bas proviennent des dispositifs orange en correspondance avec les observations expérimentales de gauche.

compte.

Dans l'approche explorée par Alessio Quaresima le branchement est un processus dynamique qui intervient durant le développement pendant la navigation des cônes de croissance. Deux manières sont possibles correspondant aux observations expérimentales. Ainsi un branchement peut intervenir soit au niveau d'un cône de croissance qui se divise en deux, soit au niveau d'une branche latérale d'où un nouveau cône de croissance peut émerger. Les différents cônes qui croissent en parallèle sont en compétition pour des ressources, conduisant à des variations de vitesses d'élongation pouvant causer un arrêt de celle-ci, voir une rétractation des branches.

Trois phases de modélisation interviennent donc pour reproduire la croissance des arbres : i) le mécanisme de branchement lui même, pour la division soit d'un cône soit d'une branche ; ii) les caractéristiques physiques des branches, à savoir leur section et leurs orientations initiales et enfin iii) la détermination des vitesses d'élongation respectives en fonction de ces caractéristiques.

Branchements par division de cônes de croissance Les chercheurs hollandais Van Pelt and Van Oyen ont un long historique de publications visant à décrire efficacement la morphologie de neurones [VAN OUYEN et collab., 2014; VAN PELT et collab., 1997; VAN PELT et UYLINGS, 2003]. Leur approche est phénoménologique cherchant un modèle mathématique effectif pour ces morphologies. L'essentiel de leur modélisation y compris pour les travaux ultérieurs est introduite en [VAN PELT et UYLINGS, 2003]. Le mécanisme de branchement est uniquement par division des cônes de croissance et les paramètres uniquement phénoménologiques, mais comme étudié dans les différentes publications de ces auteurs le modèle est capable de reproduire la diversité des morphologies neuronales identifiées. Ce modèle est basé sur une relation pour la probabilité de création de branches dépendant du nombre de branches $n(t)$ déjà existantes, d'un taux de base de branchement, du nombre de nœuds maximal entre une extrémité de croissance et le soma et d'un facteur d'asymétrie. Ce modèle a été implémenté dans notre simulateur par Alessio Quaresima dans une

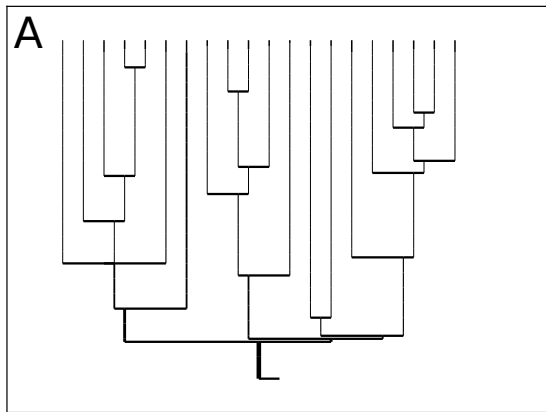


FIGURE 7.34 – **Arbre dendritique produit avec l'implémentation du modèle BEST optimisé.** La figure représente un exemple de structure dendritique possible.

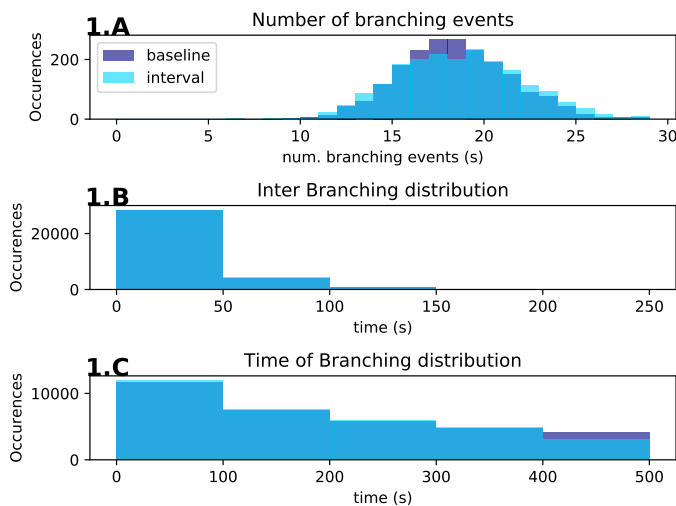


FIGURE 7.35 – **Comparaison du modèle BEST original et optimisé.** Modèle BEST noté « baseline » en bleu foncé; modèle BEST optimisé avec calcul des temps d'attente pour les branchements, noté « interval » et en bleu clair. Les trois histogrammes montrent la comparaison entre les deux implémentations de l'algorithme pour trois propriétés sur 1000 exécutions. Les distributions des nombres de branchements, des instants de branchements et aussi des intervalles de temps entre branches correspondent très bien.

version numériquement très optimisée basée non pas sur une probabilité par unité de temps mais la durée d'attente avant un tel événement. Un exemple d'arborescence est donné dans la figure 7.34 et la figure 7.35 indique le très bon accord de la version optimisée avec le modèle original.

Branchements latéraux Le processus de création de branches latérales le long de branches déjà établies suit des résultats publiés dans [SZE BENYI et collab., 1998]. Cette création locale d'un nouveau cône de croissance, donc d'une nouvelle branche, semble être liée à une fluctuation locale d'actine, causée par exemple par une pause du cône de croissance sur la branche d'origine. Les observations détaillées par SZE BENYI et collab. [1998] rapportent des données sur la dépendance de la localisation de la création de nouvelles branches latérales par rapport à la distance à l'extrémité de la branche mère ainsi qu'une dépendance explicite dans le temps. Ces relations phénoménologiques sont incorporées dans le modèle de notre simulateur avec des paramètres ajustés aux données de SZE BENYI et collab. [1998].

Angles et diamètres des branches La détermination des orientations après la création d'une nouvelle branche, que ce soit par division d'un cône de croissance terminal ou latéralement, est essentielle pour la reproduction d'une morphologie réaliste. La considération des diamètres des branches est également importante car ces facteurs seront critiques pour déterminer les vitesses d'élongation et le devenir à long terme des branches. La question des angles et des diamètres a été traitée avec précision dans [SHEFI et collab., 2004] et notre simulateur se limite à reprendre

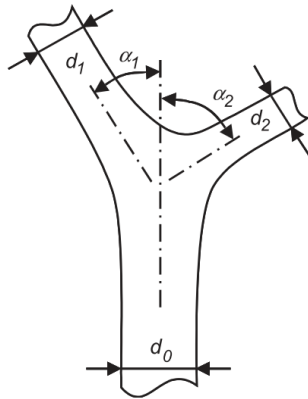


FIGURE 7.36 – **Angles et diamètres des nouvelles branches dendritiques d'après [SHEFI et collab., 2004]**. En absence de facteurs externes d'attraction les diamètres des nouvelles branches $d_{1,2}$ sont reliées au diamètre d_{parent} de la branche parent par une relation de conservation généralisée $d_{parent} = d_1^\eta + d_2^\eta$, avec η un exposant. Les nouveaux angles sont déterminés à partir de l'équilibre mécanique. Dans ce modèle les branches tirent sur la bifurcation avec une force proportionnelle à leur diamètre. Une hypothèse cohérente avec le modèle de propagation des cônes de croissance d'adhésion focale où l'action de traction des cônes de croissance est exercées sur le substrat par des moteurs moléculaires, donc proportionnellement à la surface.

les modèles et données qui y sont données pour les branchements par division de cônes. Ce modèle suppose un équilibre des forces de tension lors de la division, avec des tensions dépendantes d'une puissance des diamètres. Ceux-ci pour les branches filles tirées aléatoirement avec une règle de conservation par rapport au diamètre de la branche parent (figure 7.36). Il semble manquer des études concernant le branchement latéral, aussi une distribution ad hoc en loi de puissance des diamètres des nouvelles branches a été utilisée dans l'implémentation.

Croissance avec compétition S'inspirant des travaux de VAN OOYEN [2004] sur le développement neuronal on a repris le principe d'une compétition des cônes de croissances dans la dynamique du développement. En effet les cônes de croissance nécessitent l'apport de plusieurs molécules spécifiques pour opérer l'élongation des neurites, notamment des molécules d'actine et de tubuline. Une ressource comme la tubuline est en effet nécessaire pour la synthèse des microtubules, voir seulement de leur persistance [HELY, 2001]. Leur quantité disponible est donc critique à la dynamique. Ces molécules sont véhiculées à partir du soma par les neurites. Leur arrivée aux cônes de croissance va déterminer la dynamique de l'extrémité, à savoir si celle-ci va pouvoir s'allonger, fera une pause, ou bien connaîtra une rétractation. En dépit de leurs conséquences significatives sur la dynamique de développement les interactions via la compétition entre neurites d'un neurone ont été assez peu étudiée hors la première approche de Van Ooyen [VAN OOYEN, 2004; VAN OOYEN et collab., 2001]. Ces idées ont été développées par même équipe dans une étude plus récente [HJORTH et collab., 2014] dans lequel les éléments nécessaires au cône de croissance (tubuline dans l'étude) sont produits dans le soma et transportés vers le soma soit de manière diffusive soit active. Ce modèle a été implémenté dans un modèle à plusieurs compartiments où chaque segment d'une neurite utilise une partie de la ressource au même temps de son transfert. Malgré les résultats intéressants de ce modèle, il n'est pas utilisable dans le cadre de notre simulateur qui ne permet pas à ce jour une dynamique multi-compartiments. Ceci a amené à développer un modèle plus simplifié reprenant les principes de la compétition dont il n'y a pas lieu de donner plus qu'un résumé ici.

Pour chaque cône de croissance i on représente ainsi la quantité de la ressource qui s'y trouve disponible par une variable a_i dont la variation est décrite est une équation différentielle simple, bilan d'une part entre l'utilisation (pour l'élongation) et la perte (dégradation, fuite) de la ressource et d'autre part l'arrivée de la ressource à l'extrémité. La quantité de ressource qui arrive à un cône de croissance i est une fraction de la quantité totale de ressource dans la neurite connectée au soma que tous les cônes qui y sont rattachés doivent se partager. La compétition se manifeste dans la répartition entre chaque cône de cette quantité totale suivant un principe de rétroaction positive : les cônes les plus actifs sont les plus demandeurs et seront les mieux servis. Au même temps que cette dynamique de transport et de répartition pour chaque cône de croissance, une autre équation différentielle décrit la production à taux constant de la ressource dans le soma vers une quantité d'équilibre moins l'écoulement d'une certaine fraction vers les neurites. Toutes ces

équations différentielles sont stochastiques avec l'ajout de termes aléatoires pour rendre compte de fluctuations d'utilisation de la ressource indépendamment dans chaque neurite, ainsi que de fluctuations de sa production dans le soma.

L'intérêt de quantifier cette dynamique sur une ressource critique est de pouvoir ensuite l'associer aux régimes de comportement dont elle est limitante. On connaît trois types de comportements du cône de croissance. En premier lieu sa progression avec élongation de la neurite, qui sera possible si la quantité a_i disponible un instant donné dans le cône i est suffisante, supérieure à un certain seuil. Si au contraire la quantité de ressource est trop faible, inférieure à un autre seuil a priori différent du précédent, alors le cône de croissance se rétracte. Vitesses d'élongation et de rétraction dépendent de a_i . Entre les deux seuils de ressource, le cône de croissance est en pause sans déplacement. À ces deux valeurs seuils on peut en rajouter une troisième, supérieure : quand la quantité de ressource est très abondante elle induira une division du cône de croissance avec la création de deux branches.

Ce modèle d'équations différentielles stochastiques a été développé et implémenté par Alessio Quaresima. Un exemple de la dynamique de compétition à trois branches est donné dans la figure 7.37 et un exemple de morphologie générée à la figure 7.38. L'inclusion de la compétition est un mécanisme qui peut expliquer l'apparition d'asymétries dans les arborescences dendritiques dont les modèles phénoménologiques de Van Pelt et Van Ooyen rendaient compte avec un paramètre ad hoc.

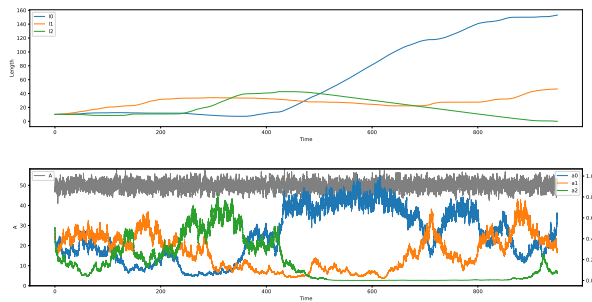


FIGURE 7.37 – Étude de la dynamique du modèle de compétition pour 3 cônes de croissance en interaction. B.1 : distance curvilinéaire en fonction du temps depuis le soma jusqu'à chacun des 3 cônes en compétition. Les branches pour s'allonger ou se rétracter. On observe la rétraction complète de la branche l_2 . B.2 : quantité de la ressource critique reçue au cours du temps par chaque cône et quantité totale disponible (en gris).

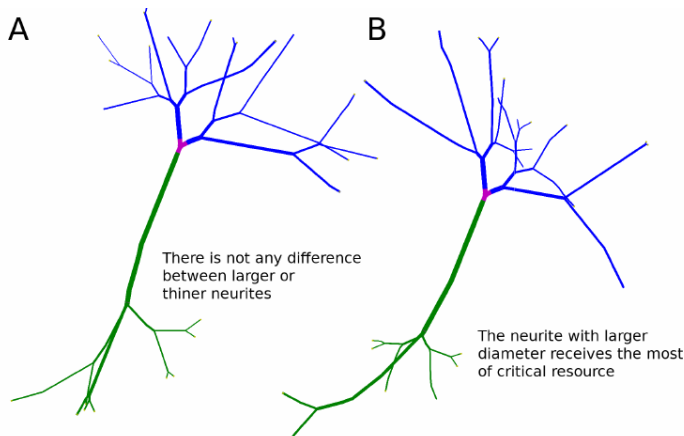


FIGURE 7.38 – Compétition dépendante du diamètre. Neurones générés par le modèle de compétition. A : toutes les branches reçoivent à peu près la même quantité de ressource. B : la quantité reçue par les cônes de croissance dépend du diamètre de leur branche, tous les autres paramètres identiques à A).

Clairement la croissance des neurites est plus complexe que le modèle implémenté. Ici une seule ressource limitante était prise en compte, représentant a priori la tubuline. D'autres ressources ont aussi une dynamique dans les neurones et peuvent intervenir de manière analogue comme l'actine, les vésicules acheminant les éléments de membrane nécessaires à l'élongation, et le transport de mitochondries nécessaires aux processus énergétiques [HJORTH et collab., 2014]. Néanmoins ce nouveau modèle est une première étape pour inclure un tel processus dans un simulateur de morphologie neuronale rapide et semble pouvoir rendre compte de morphologies observés et des phénomènes de maturation du réseau (arrêt et rétractation).

Le simulateur peut réaliser un grand nombre de structures neuronales en fonction des différents paramètres pour les processus incorporés comme le taux de branchement, l'asymétrie des branches, la longueur de persistance la vitesse de base d'élongation. Ces paramètres doivent être ajustés par l'utilisateur s'il veut reproduire un type donné de neurone. La manière de réaliser ces ajustements est un projet en soi. Une attention néanmoins déjà été portée à cette question en recherchant explicitement un modèle de navigation libre des cônes de croissance pouvant être facilement paramétrée à partir d'observation.

Afin de valider les morphologies générées, et afin de réaliser les comparaisons avec des morphologies caractérisées à partir d'observation nécessaires à la recherche de paramètres il faut disposer de descripteurs statistiques des formes complexes permettant de dépasser la similitude visuelle.

Des descripteurs statistiques de la forme comme la tortuosité qui caractérise la courbature des trajectoires, la distance à l'origine de l'extrémité, la longueur de persistance, peuvent être utilisés pour comparer les formes simulées à des données extraites à partir d'observations. Ces comparaisons ont pu être réalisées uniquement sur quelques exemples de neurones du cortex visuel dont l'organisation naturelle est planaire et dont la morphologie est disponible dans la base de données Netmorph.

7.4.3 Évaluation et extensions futures

L'évaluation du simulateur est un sujet difficile car elle dépend avant tout de la disponibilité de données expérimentales de comparaison. Or, on l'a déjà évoqué pour des cultures et des neurones, à 2 dimensions ces données sont largement manquantes. Ces évaluations sont encore en cours de développement ainsi que les modèles et les codes du simulateur. Plusieurs exemples montrant la versatilité des morphologies possibles ont cependant pu être générés (figure 7.39). La particularité majeure du simulateur reste la possibilité de générer des morphologies dans des géométries confinées avec des obstacles. Son potentiel a été illustré dans figure 7.40 avec une zone de croissance circulaire délimitée par des parois. Des géométries plus complexes sont en préparations et illustreront la thèse de Tanguy Fardet.

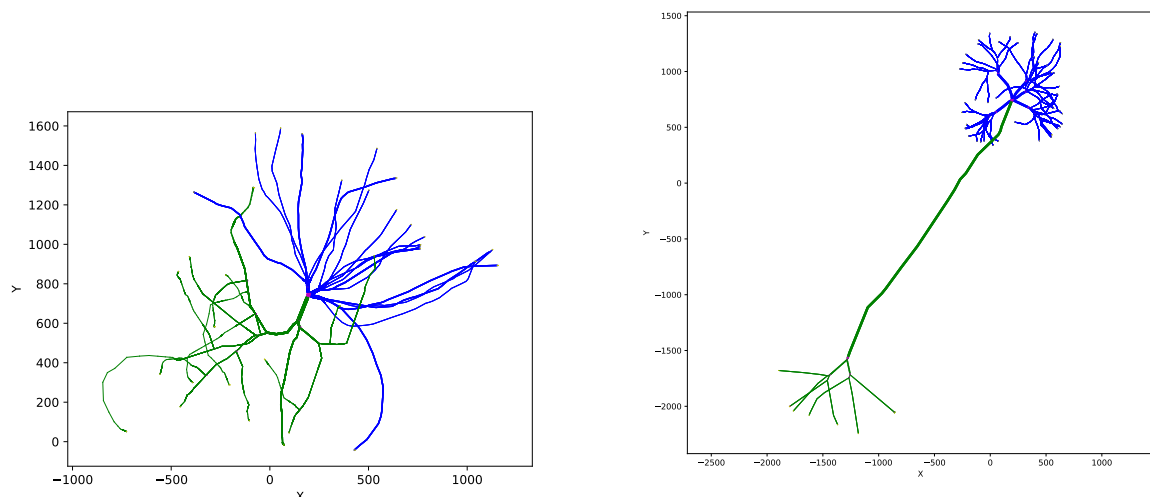


FIGURE 7.39 – **Reconstitutions de différents types de neurones.** Représentations préliminaires de neurones générés de morphologies différentes de similarité apparente avec des neurones Chandelier (à gauche) et Multipolaire (à droite) obtenues avec des paramètres ad hoc ajustés manuellement. Des analyses statistiques de descripteurs de l'arborescence (analyse de Scholl) confirment cette ressemblance/

Une extension du modèle non encore implémentée concerne le guidage des cônes de croissance par des traceurs chimiques diffusant dans l'environnement [SEGEV et BEN-JACOB, 2000].

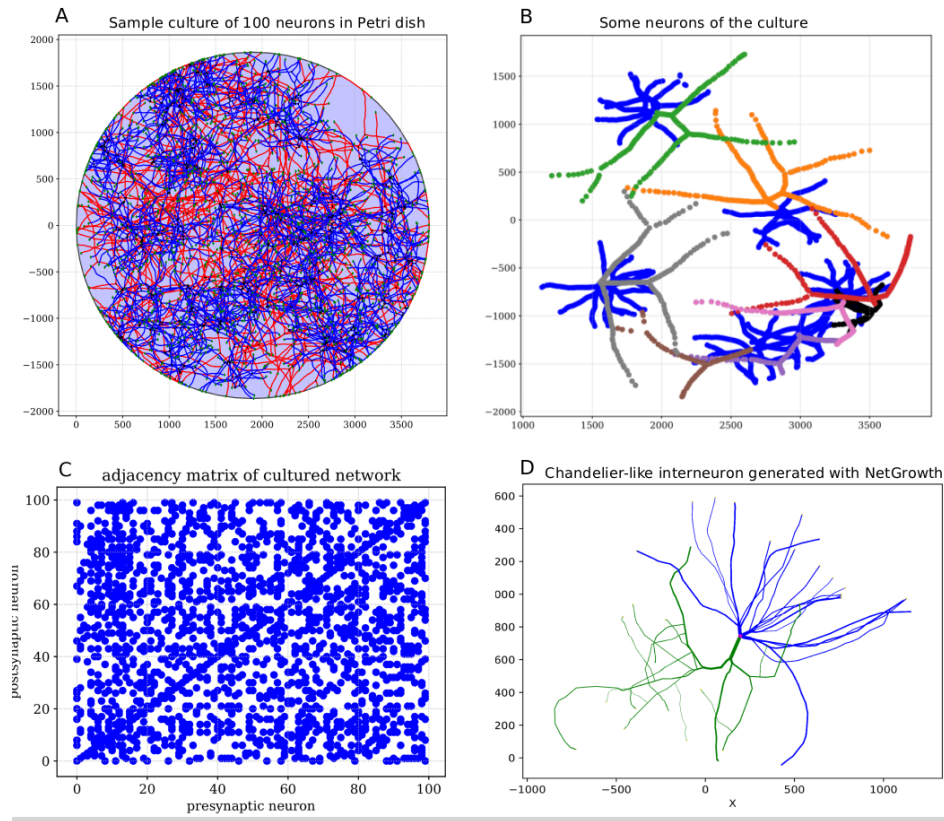


FIGURE 7.40 – **Exemples de génération de morphologies dans des cultures confinées.** Croissance de neurones dans un milieu circulaire. A e B : les neurones interagissent avec la bordure circulaire de la chambre. C : à chaque croisement d'un axones avec les neurites d'un autre est associé une probabilité de connexion synaptique, le réseau de connexions ainsi établi peut-être représenté sous forme d'une matrice d'adjacence. D : sans contrainte les neurones ont une morphologie suivant des types caractéristiques contrôlables par les paramètres du simulateur, ici exemple d'une morphologie de type neurone Chandelier.

Des substances pouvant également être relâchées par les neurones eux-mêmes et conduisant à une situation complexe dans laquelle à la fois interactions directes de contact et signalisation à distance entre des neurones en croissance déterminent le développement du réseau. L'inclusion de modèles chemotactiques dans le système de simulation demandera de prendre en compte aussi les phénomènes de diffusion moléculaires.

Enfin à terme on souhaiterait aussi pouvoir explorer les changements morphologiques dans les cultures impliquant des réarrangements physiques de la localisation des corps cellulaires et des neurites sous l'actions des tensions s'exerçant entre les cellules et les changements au niveau de l'adhésion avec le substrat. Ces développements demanderont de prendre en compte forces et déplacements avec une nouvelle approche mécanique.

L'amélioration des modèles et du simulateur de croissance va nous permettre de l'utiliser pour appuyer les études théoriques sur l'activité dans les cultures afin de les baser sur des connectivités plus réalistes que les représentations très empiriques disponibles actuellement. En outre, en raison des dystrophies observées sur les axones et dendrites de neurones dans le cas de maladies neurodégénératives comme Alzheimer [CAI et collab., 2006; GRACE et collab., 2002] et Parkinson [GALVIN et collab., 1999; SKAPER et collab., 2018; TAGLIAFERRO et BURKE, 2016], nous souhaitons appliquer cet outil afin d'aider à caractériser les mécanismes impliqués dans ces troubles.

7.5 Conclusion du chapitre

J'ai présenté dans ce chapitre mes travaux concernant les comportements et structures de réseaux neuronaux en culture. Une orientation que j'ai initié en 2012 convainquant mes collègues Pascal Monceau puis Stéphane Métens du laboratoire Matières et Systèmes complexes de former équipe pour explorer les possibilités curieuses de ces systèmes excitables vivants. Deux docteurs formidables Renaud Renault puis Tanguy Fardet nous ont permis dans l'équipe rapidement progresser sur les différents sujets discutés plus haut et développer les collaborations avec des expérimentateurs de pointe sur les cultures, Elisha Moses, Jean-Louis Viovy, Catherine Villard et Jordi Soriano.

Les premiers travaux qui nous ont occupés concernaient le développement de la percolation à quorum initialement introduite par l'équipe d'Elisha Moses afin de chercher à caractériser la connectivité dans les cultures. Les publications qui en ont résulté discutées dans la section 7.7.2 sont de nature très théoriques. Elles contribuent davantage certainement à la théorie de certains types de percolation qu'à la compréhension réelle de cultures neuronales. Ces travaux indiquent néanmoins des limitations des conclusions sur la connectivité des analyses par les premiers modèles naïfs de percolation à quorum des auteurs qui nous ont inspirés? Si le mécanisme de percolation à quorum semble bien confirmé pour la réponse d'une population à une stimulation les taux de décroissance du potentiel membranaire, la variabilité de l'excitabilité des neurones que nous avons pris en compte dans nos modèles affectent significativement les prédictions numériques. D'une manière que nous ne pouvons quantifier malheureusement.

L'activité spontanée est un thème central pour les populations neuronales y compris en culture. Elle intéresse naturellement les physiciens qui y trouvent les manifestations de systèmes complexes collectifs. Des comportements qui m'ont à titre personnel rappelé des anciens sujets durant mon doctorat : la criticalité auto-organisée et la synchronisation collective d'oscillateurs. Sujets lointains de ma thèse. Quelle surprise de les retrouver vingt ans après basés, enfin, sur un terrain expérimental! Seule la synchronisation m'a occupé pour l'instant; les comportements éventuellement critiques dans les cultures restant une question controversée qui demanderait une recherche spécialisée approfondie. Elisha Moses, Yaron Penn et Menahem Segal venaient justement de conclure leur étude remarquable sur les bouffées spontanées de réseau émergeant par synchronisation de neurones intrinsèquement oscillants en augmentant le couplage synaptique. Ceci a conduit à notre modèle des bouffées de réseaux présenté à la section 7.7.3. Il s'agit d'un scénario de dynamique collective des neurones en cultures dominé par une fraction importante de cellules oscillantes. Une situation qui n'est pas établie en toute généralité, ce qui appelle à des études complémentaires aussi bien théoriques qu'expérimentales. Certaines expériences pour confirmer la validité de notre modèle théorique ont été récemment entamées par Tanguy Fardet au laboratoire de Elisha Moses.

Une de nos motivations de l'étude des cultures neuronales concerne leur utilisation comme plateforme d'exploration des conditions nécessaires à des fonctions cognitives. Nous l'avons vu, les techniques de microfabrication et de microfluidique permettent la fabrication de dispositifs sophistiqués structurant les cultures en de multiples compartiments pouvant être reliés de manière sélective. La conception de tels dispositifs reste technologiquement exigeante et les propriétés des réseaux qui s'y établissent mal connues. Ceci motive l'axe de recherche discuté dans la section 7.7.4 visant à modéliser le développement de la morphologie des neurones et des réseaux dans des environnements contraints. Se basant sur la probable importance de la forme des neurones pour leurs fonctions physiologiques dans les cerveaux, j'ai été très surpris de constater à quel point les mécanismes de croissance de ceux-ci et d'établissement des réseaux neuronaux sont mal caractérisés et peu modélisés. Nous avons donc entamé des efforts importants non seulement de modélisation de processus sous un angle biophysique mais aussi pour établir les bases suivant des principes modernes de programmations d'une plateforme de simulation flexible de morphogenèse des neurones et réseaux.

La revue de ces différents travaux me les fait vraiment percevoir avant tout comme des préliminaires pour ce qui deviendra vraiment intéressant! Une perspective fondamentale de futures orientation est l'exploration du lien entre schémas de connectivité et complexité dans l'activité neuronale. Dans cette perspective avec Catherine Villard et Jordi Soriano nous avons pour projet l'utilisation de dispositifs microfluidiques avec différentes formes de géométries sophistiquées pour manipuler la connectivité spontanée des réseaux au sein de cultures et chercher à faire émerger des comportements plus complexes que les bouffées synchronisées. Avec ces outils nous voulons aussi suivre en parallèle la croissance du réseau et l'évolution des comportements au cours du développement du réseau, ce qui n'est pas possible dans des modèles animaux.

Pour la théorie il s'agira d'une part d'établir des méthodes de caractérisation de la complexité de l'activité neuronale pour analyser les dynamiques collectives expérimentales ou simulées. D'autre part de poursuivre la recherche sur les modèles biophysiques du développement neuronal et de la croissance des réseaux pour explorer l'organisation qui s'établit et aider à la conception de dispositifs microfluidiques. Outre l'intérêt pour la question fondamentale ci-dessous, ces développements peuvent jouer pour la conception de modèles de cultures neuronales microfluidiques de pathologies neurologiques.

Enfin, une question majeure concernant les dynamiques possibles et fonctions cognitives éventuelles de cultures a trait à la plasticité synaptique qui s'y manifeste et l'apprentissage possible. Des propriétés comme nous l'avons vu mal contrôlées qui nécessitent encore aussi bien des efforts expérimentaux précis que des études théoriques. Une première approche d'ors et déjà envisagée est d'étudier les conséquences des modèles courants de plasticité sur les dynamiques spontanées ou induites dans des cultures de degré de désordre de connectivité contrôlé.

De nombreuses autres questions et sujet d'études sont d'actualité et peuvent faire de futurs projets, j'en indique une liste, non exhaustive, sans développer dans l'encadré 7.1. Des perspectives variées dont la poursuite dépendra naturellement de nombreux facteurs, dont les financements, la participation de collègues et étudiants et l'élément crucial du temps!

Liste non détaillée de sujets ouverts sur les cultures de neurones dans la suite de mes recherches :

- Caractériser la dynamique spatio-temporelle de la croissance d'une bouffée de potentiels d'actions. Comment se construit la synchronisation du réseau lors de la croissance d'une bouffée et comment est-elle reliée à la topologie sous-jacente.
- Établir une mesure de la complexité de l'activité spontanée d'une culture.
- Étudier les effets de l'hétérogénéité de types neuronaux dans l'activité d'une population. Mieux préciser les conséquences des effets du désordre et du bruit.
- Explorer d'un point de vue théorique des scénarios d'apprentissage dans des cultures en fonction de modèles de synapses.
- Étudier le rapport entre excitation et inhibition dans les réseaux pour mieux en comprendre les raisons. Quel est l'effet des neurones inhibiteurs durant la synchronisation?
- Quels sont les effets de différentes fractions de neurones oscillateurs dans une population? Clarifier si les oscillations de neurones individuels observées dans des cultures sont induites par l'environnement (bruit synaptique) ou par des courants persistants (sodium)
- Déterminer les paramètres biologiques qui affectent la variabilité des caractéristiques des bouffées de réseau.
- Quels sont les rôles relatifs des différents processus synaptiques, lents (récepteurs AMPA) et rapides (récepteurs NMDA) pour la synchronisation?
- Comprendre les interactions entre les cônes de croissance et des obstacles (physiques, chimiques, autres cellules). Comment les contraintes physiques affectent l'organisation du réseau et ses propriétés fonctionnelles?
- En dépit du désordre apparent est-ce que une complexité morphologique émerge durant la croissance du réseau en culture?
- Décrire et comprendre le vieillissement physique d'une culture avec les réarrangements des somas et neurites.

Encadré 7.1

7.6 Références

- 2017, «Model-based cognitive neuroscience», *Journal of Mathematical Psychology*, vol. 76, doi : 10.1016/J.JMP.2016.10.010, p. 59–64, ISSN 0022-2496. [292](#)
- ALLEN, R. et D. AUBRY. 2012, «'Run-and-tumble' or 'look-and-run'? A mechanical model to explore the behavior of a migrating amoeboid cell», *Journal of Theoretical Biology*, vol. 306, doi : 10.1016/j.jtbi.2012.03.041, p. 15–31, ISSN 00225193. [325](#)
- BOIS, P., R. GUINAMARD, A. E. L. CHEMALY, J.-F. FAIVRE et J. BESCOND. 2007, «Molecular regulation and pharmacology of pacemaker channels.», *Current pharmaceutical design*, vol. 13, n° 23, doi :10.2174/138161207781368729, p. 2338–49, ISSN 1873-4286. [310](#)
- BOTTANI, S. 1995, «Pulse-Coupled Relaxation Oscillators : From Biological Synchronization to Self-Organized Criticality», *Physical Review Letters*, vol. 74, n° 21, doi :10.1103/PhysRevLett.74.4189, p. 4189–4192, ISSN 0031-9007. [304](#), [310](#)
- BRESKIN, I., J. SORIANO, E. MOSES et T. TLUSTY. 2006, «Percolation in living neural networks.», *Physical Review Letters*, vol. 97, n° 18, doi :10.1103/PhysRevLett.97.188102, p. 1–4, ISSN 00319007. [290](#), [291](#), [292](#)
- BRETTE, R. et W. GERSTNER. 2005, «Adaptive exponential integrate-and-fire model as an effective description of neuronal activity.», *Journal of neurophysiology*, vol. 94, n° 5, p. 3637–3642. [303](#)

- CAI, D., M. ZHONG, R. WANG et collaborateurs. 2006, «Phospholipase D1 corrects impaired betaAPP trafficking and neurite outgrowth in familial Alzheimer's disease-linked presenilin-1 mutant neurons.», *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, n° 6, doi :10.1073/pnas.0510710103, p. 1936–40, ISSN 0027-8424. [332](#)
- COHEN, D. et M. SEGAL. 2009, «Homeostatic Presynaptic Suppression of Neuronal Network Bursts», *Journal of Neurophysiology*, vol. 101, n° 4, doi :10.1152/jn.91085.2008, p. 2077–2088, ISSN 0022-3077. [317](#)
- COHEN, O., A. KESELMAN, E. MOSES et collaborateurs. 2010, «Quorum percolation in living neural networks», *EPL*, vol. 89, n° 1, doi :10.1209/0295-5075/89/18008, p. 1–6, ISSN 02955075. [291](#), [292](#)
- DAO DUC, K., C. Y. LEE, P. PARUTTO et collaborateurs. 2015, «Bursting reverberation as a multiscale neuronal network process driven by synaptic depression-facilitation», *PLoS ONE*, vol. 10, n° 5, doi :10.1371/journal.pone.0124694, p. e0124694, ISSN 19326203. [317](#)
- FABER, E. S. L. et P. SAH. 2002, «Physiological role of calcium-activated potassium currents in the rat lateral amygdala.», *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 22, n° 5, doi :22/5/1618[pil], p. 1618–1628, ISSN 1529-2401. [310](#)
- FARDET, T., M. BALLANDRAS, S. BOTTANI, S. MÉTENS et P. MONCEAU. 2018a, «Understanding the generation of network bursts by adaptive oscillatory neurons», *Frontiers in Neuroscience*, vol. 12, n° FEB, doi :10.3389/fnins.2018.00041, ISSN 1662453X. [302](#), [310](#), [315](#)
- FARDET, T., S. BOTTANI, S. MÉTENS et P. MONCEAU. 2018b, «Effects of inhibitory neurons on the quorum percolation model and dynamical extension with the Brette–Gerstner model», *Physica A : Statistical Mechanics and its Applications*, vol. 499, doi :10.1016/j.physa.2018.02.002, p. 98–109, ISSN 03784371. [300](#), [302](#), [304](#), [312](#)
- GALVIN, J. E., K. URYU, V. M. LEE et J. Q. TROJANOWSKI. 1999, «Axon pathology in Parkinson's disease and Lewy body dementia hippocampus contains alpha-, beta-, and gamma-synuclein.», *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, n° 23, p. 13450–5, ISSN 0027-8424. [332](#)
- GAUSS, R. et R. SEIFERT. 2000, «Pacemaker oscillations in heart and brain : a key role for hyperpolarization-activated cation channels», *Chronobiol Int*, vol. 17, n° 4, p. 453–469, ISSN 0742-0528. [310](#)
- GEWALTIG, M. et M. DIEMANN. 2007, «NEST :NEural Simulation Tool», *Scholarpedia*, vol. 2, p. 1430. [303](#)
- GRACE, E. A., C. A. RABINER et J. BUSCIGLIO. 2002, «Characterization of neuronal dystrophy induced by fibrillar amyloid beta : implications for Alzheimer's disease.», *Neuroscience*, vol. 114, n° 1, p. 265–73, ISSN 0306-4522. [332](#)
- GRITSUN, T. A., J. LE FEBER, J. STEGENGA et W. L. C. RUTTEN. 2010, «Network bursts in cortical cultures are best simulated using pacemaker neurons and adaptive synapses», *Biological Cybernetics*, vol. 102, n° 4, doi :10.1007/s00422-010-0366-x, p. 293–310, ISSN 03401200. [310](#), [318](#)
- HELY, T. G. B. 2001, «A Computational Model of Dendrite Elongation and Branching Based on MAP2 Phosphorylation», *Journal of Theoretical Biology*, vol. 210, n° 3, doi :10.1006/jtbi.2001.2314, p. 375–384, ISSN 00225193. [329](#)
- HJORTH, J. J. J., J. VAN PELT, H. D. MANSVELDER et A. VAN OYEN. 2014, «Competitive dynamics during resource-driven neurite outgrowth», *PLoS ONE*, vol. 9, n° 2, doi :10.1371/journal.pone.0086741, p. e86741, ISSN 19326203. [329](#), [330](#)

- ISOKAWA, M. 1997, «Membrane time constant as a tool to assess cell degeneration», *Brain Research Protocols*, vol. 1, n° 2, doi :10.1016/S1385-299X(96)00016-5, p. 114–116, ISSN 1385299X. 296
- IZHIKEVICH, E. 2006, «Bursting», doi :10.4249/scholarpedia.1300. URL <http://www.scholarpedia.org/article/Bursting>. 317
- KAMIOKA, H., E. MAEDA, Y. JIMBO, H. P. ROBINSON et A. KAWANA. 1996, «Spontaneous periodic synchronized bursting during formation of mature patterns of connections in cortical cultures.», *Neuroscience letters*, vol. 206, n° 2-3, p. 109–112, ISSN 0304-3940. 309
- KOBAYASHI, R. et K. KITANO. 2016, «Impact of slow K⁺ currents on spike generation can be described by an adaptive threshold model», *Journal of Computational Neuroscience*, vol. 40, n° 3, doi :10.1007/s10827-016-0601-0, p. 347–362, ISSN 15736873. 310
- KOIZUMI, H. et J. C. SMITH. 2008, «Persistent Na⁺ and K⁺-Dominated Leak Currents Contribute to Respiratory Rhythm Generation in the Pre-Botzinger Complex In Vitro», *Journal of Neuroscience*, vol. 28, n° 7, doi :10.1523/JNEUROSCI.3916-07.2008, p. 1773–1785, ISSN 0270-6474. 310
- LU, T. Z. et Z. P. FENG. 2012, «NALCN : A regulator of pacemaker activity», *Molecular Neurobiology*, vol. 45, n° 3, doi :10.1007/s12035-012-8260-2, p. 415–423, ISSN 08937648. 310
- LUCCIOLI, S., E. BEN-JACOB, A. BARZILAI, P. BONIFAZI et A. TORCINI. 2014, «Clique of functional hubs orchestrates population bursts in developmentally regulated neural networks.», *PLoS computational biology*, vol. 10, n° 9, doi :10.1371/journal.pcbi.1003823, p. e1003823, ISSN 1553-7358. 309
- LÜTHI, A. et D. A. MCCORMICK. 1998, «H-current : Properties of a neuronal and network pacemaker», *Neuron*, vol. 21, n° 1, doi :10.1016/S0896-6273(00)80509-7, p. 9–12, ISSN 08966273. 310
- MEISTER, M., R. WONG, D. BAYLOR et C. SHATZ. 1991, «Synchronous bursts of action potentials in ganglion cells of the developing mammalian retina», *Science*, vol. 252, n° 5008, doi :10.1126/science.2035024, p. 939–943, ISSN 0036-8075. 309
- MÉTENS, S., P. MONCEAU, R. RENAULT et S. BOTTANI. 2016, «Finite-size effects and dynamics of giant transition of a continuum quorum percolation model on random networks», *Physical review. E*, vol. 93, n° 3, doi :10.1103/PhysRevE.93.032112, p. 032112, ISSN 2470-0045. 292, 293, 294, 295, 300
- MONCEAU, P., R. RENAULT, S. MÉTENS et S. BOTTANI. 2016, «Effect of threshold disorder on the quorum percolation model.», *Physical review. E*, vol. 94, n° 1-1, doi :10.1103/PhysRevE.94.012316, p. 012316, ISSN 2470-0053. 298
- VAN OOYEN, A. 2004, «Competition in neurite outgrowth and the development of nerve connections», *Progress in Brain Research*, vol. 147, n° SPEC. ISS., doi :10.1016/S0079-6123(04)47007-1, p. 81–99, ISSN 00796123. 329
- VAN OOYEN, A., A. CARNELL, S. DE RIDDER et collaborateurs. 2014, «Independently outgrowing neurons and geometry-based synapse formation produce networks with realistic synaptic connectivity.», *PloS one*, vol. 9, n° 1, doi :10.1371/journal.pone.0085858, p. e85858, ISSN 1932-6203. 327
- VAN OOYEN, A., B. P. GRAHAM et G. J.A. RAMAKERS. 2001, «Competition for tubulin between growing neurites during development», *Neurocomputing*, vol. 38-40, doi :10.1016/S0925-2312(01)00487-8, p. 73–78, ISSN 0925-2312. 329
- ORLANDI, J. G., J. SORIANO, E. ALVAREZ-LACALLE, S. TELLER et J. CASADEMUNT. 2013, «Noise focusing : the emergence of coherent activity in neuronal cultures supplementary information», *Nature Physics*. 309, 317

- PAOLUZZI, M., R. DI LEONARDO et L. ANGELANI. 2013, «Effective run-and-tumble dynamics of bacteria baths», *Journal of Physics : Condensed Matter*, vol. 25, n° 41, doi :10.1088/0953-8984/25/41/415102, p. 415 102, ISSN 0953-8984. [325](#)
- PENN, Y., M. SEGAL et E. MOSES. 2016, «Network synchronization in hippocampal neurons», *Proceedings of the National Academy of Sciences*, doi :10.1073/pnas.1515105113, p. 201515 105, ISSN 0027-8424. [309](#), [310](#), [318](#)
- PICARDO, M. A., P. GUIGUE, P. BONIFAZI et collaborateurs. 2011, «Pioneer GABA cells comprise a subpopulation of hub neurons in the developing hippocampus», *Neuron*, vol. 71, n° 4, doi : 10.1016/j.neuron.2011.06.018, p. 695–709, ISSN 08966273. [309](#)
- RAMIREZ, J.-M., A. K. TRYBA et F. PEN. 2004, «Pacemaker neurons and neuronal networks : an integrative view ~», doi :10.1016/j.conb.2004.10.011. [309](#)
- RENAULT, R. 2015, *Emergent design of neuronal devices*, thèse de doctorat, Université Paris-diderot. URL <http://www.theses.fr/2015USPCC161>. [319](#), [321](#), [324](#)
- RENAULT, R., J.-B. DURAND, J.-L. VIOVY et C. VILLARD. 2016, «Asymmetric axonal edge guidance : a new paradigm for building oriented neuronal networks», *Lab Chip*, vol. 16, n° 12, doi :10.1039/C6LC00479B, p. 2188–2191, ISSN 1473-0197. [323](#), [327](#)
- RENAULT, R., P. MONCEAU et S. BOTTANI. 2013, «Memory decay and loss of criticality in quorum percolation.», *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 88, n° 6, p. 062 134, ISSN 1550-2376. [295](#), [297](#)
- RENAULT, R., P. MONCEAU, S. BOTTANI et S. MÉTENS. 2014, «Effective non-universality of the quorum percolation model on directed graphs with Gaussian in-degree», *Physica A : Statistical Mechanics and its Applications*, vol. 414, doi :10.1016/j.physa.2014.07.028, p. 352–359, ISSN 03784371. [292](#), [293](#), [302](#)
- RENAULT, R., N. SUKENIK, S. DESCROIX et collaborateurs. 2015, «Combining microfluidics, optogenetics and calcium imaging to study neuronal communication in vitro», *PLoS ONE*, vol. 10, n° 4, doi :10.1371/journal.pone.0120680, ISSN 19326203. [306](#), [307](#), [308](#), [321](#), [322](#)
- ROTH, A. et M. C. W. VAN ROSSUM. 2009, «Modeling Synapses», dans *Computational Modeling Methods for Neuroscientists*, The MIT Press, ISBN 9780262013277, p. 139–160, doi :10.7551/mitpress/9780262013277.003.0007. [303](#)
- ROTH, S., M. BISBAL, J. BROCARD et collaborateurs. 2014, «Shaping neurons : how morphological constraints affect axonal polarity», *Neuroscience*. [321](#)
- SEGEV, R. et E. BEN-JACOB. 2000, «Generic modeling of chemotactic based self-wiring of neural networks.», *Neural networks : the official journal of the International Neural Network Society*, vol. 13, n° 2, p. 185–99, ISSN 0893-6080. [320](#), [331](#)
- SHEFI, O., A. HAREL, D. B. CHKLOVSKII, E. BEN-JACOB et A. AYALI. 2004, «Biophysical constraints on neuronal branching», *Neurocomputing*, vol. 58-60, doi :10.1016/j.neucom.2004.01.085, p. 487–495, ISSN 09252312. [328](#), [329](#)
- SKAPER, S. D., M. BARBIERATO, V. FERRARI, M. ZUSSO et L. FACCI. 2018, «Culture of Rat Mesencephalic Dopaminergic Neurons and Application to Neurotoxic and Neuroprotective Agents», dans *Methods in molecular biology (Clifton, N.J.)*, vol. 1727, p. 107–118, doi :10.1007/978-1-4939-7571-6_8. [332](#)
- SORIANO, J., I. BRESKIN, E. MOSES et T. TLUSTY. 2007, «Percolation approach to study connectivity in living neural networks», dans *AIP Conference Proceedings*, vol. 887, ISBN 0735403902, ISSN 0094243X, p. 96–106, doi :10.1063/1.2709591. [292](#)

- SORIANO, J., M. RODRÍGUEZ MARTÍNEZ, T. TLUSTY et E. MOSES. 2008, «Development of input connections in neural cultures.», *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, n° 37, doi :10.1073/pnas.0707492105, p. 13 758–63, ISSN 1091-6490. [292](#), [302](#)
- STEGENGA, J., J. LE FEBER, E. MARANI et W. RUTTEN. 2008, «Analysis of Cultured Neuronal Networks Using Intraburst Firing Characteristics», *IEEE Transactions on Biomedical Engineering*, vol. 55, n° 4, doi :10.1109/TBME.2007.913987, p. 1382–1390, ISSN 0018-9294. [309](#)
- STEPANYANTS, A. et D. B. CHKLOVSKII. 2005, «Neurogeometry and potential synaptic connectivity», *Trends in Neurosciences*, vol. 28, n° 7, doi :10.1016/j.tins.2005.05.006, p. 387–394, ISSN 01662236. [320](#)
- SURESH, J., M. RADOJICIC, L. L. PESCE et collaborateurs. 2016, «Network burst activity in hippocampal neuronal cultures : the role of synaptic and intrinsic currents», *Journal of Neurophysiology*, vol. 115, n° 6, doi :10.1152/jn.00995.2015, p. 3073–3089, ISSN 0022-3077. [318](#)
- SZEBENYI, G., J. L. CALLAWAY, E. W. DENT et K. KALIL. 1998, «Interstitial branches develop from active regions of the axon demarcated by the primary growth cone during pausing behaviors.», *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 18, n° 19, p. 7930–40, ISSN 0270-6474. [328](#)
- TAGLIAFERRO, P. et R. E. BURKE. 2016, «Retrograde Axonal Degeneration in Parkinson Disease», *Journal of Parkinson's Disease*, vol. 6, n° 1, doi :10.3233/JPD-150769, p. 1–15, ISSN 18777171. [332](#)
- VAN PELT, J., A. E. DITYATEV et H. B. M. UYLINGS. 1997, «Natural variability in the number of dendritic segments : Model-based inferences about branching during neurite outgrowth», *Journal of Comparative Neurology*, vol. 387, n° 3, doi :10.1002/(SICI)1096-9861(19971027)387:3<325::AID-CNE1>3.0.CO;2-2, p. 325–340, ISSN 00219967. [327](#)
- VAN PELT, J. et H. B. M. UYLINGS. 2003, «Growth functions in dendritic outgrowth», *Brain and Mind*, vol. 4, n° 1, doi :10.1023/A:1024160131897, p. 51–65, ISSN 13891987. [327](#)
- WAGENAAR, D. A., J. PINE et S. M. POTTER. 2006, «Searching for plasticity in dissociated cortical cultures on multi-electrode arrays.», *Journal of negative results in biomedicine*, vol. 5, doi :10.1186/1477-5751-5-16, p. 16, ISSN 1477-5751. [309](#)
- YAMAMOTO, H., S. KUBOTA, Y. CHIDA et collaborateurs. 2016, «Size-dependent regulation of synchronized activity in living neuronal networks.», *Physical review. E*, vol. 94, n° 1-1, doi : 10.1103/PhysRevE.94.012407, p. 012 407, ISSN 2470-0053. [318](#)

Chapitre 8

Conclusion générale

Ce mémoire a parcouru une large partie de mes activités de recherche sur une période assez longue démarrant vers les tournant du millénaire. J'y ai longuement parlé de sujets de biologie des systèmes et de neurosciences, deux domaines très actifs de la recherche contemporaine des Sciences du Vivant ayant bénéficié – comme bien d'autres secteurs – d'avancées technologiques majeures d'instrumentation scientifique et de capacités d'analyse. Ces deux champs de recherche s'occupent d'objets du vivant différents, molécules et gènes dans un cas, neurones et autres cellules cérébrales dans l'autre. Pourtant, ils partagent pour le physicien que je suis, de nombreux points communs qui ont conduit à mon implication sur ces deux terrains. Ce parallèle a été aussi relevé en 2008 dans une intéressante analyse par le neurobiologiste théorique [SCHUTTER et DE SCHUTTER](#).

Biologie des systèmes et neurosciences des affinités. Biologie des systèmes et neurosciences se sont retrouvés dans mes champs d'activité car j'y ai perçu le thème commun de l'émergence de systèmes vivant autonomes traitant l'information de l'environnement et réalisant des choix comportementaux en conséquence. Des systèmes cognitifs dans les deux cas. Le premier mettant en œuvre des ensembles interconnectés de gènes et de molécules déterminant les états biochimiques des cellules et intervenant dans l'héritage et l'évolution naturelle. Le second basé essentiellement sur des ensembles de neurones interconnectés et déterminant les états cérébraux d'activité électrique.

Dans les deux systèmes des réseaux d'unité actives possèdent une organisation complexe permettant les comportements de leurs fonctions physiologiques sophistiquées respectives. Les deux domaines manifestent des processus de traitement de l'information et des opérations de calcul pour produire en résultat l'adaptation des systèmes aux situations environnementales. De l'information sur les états moléculaires, intra- et extra-cellulaires au niveau de l'activité des gènes, et de l'information venant des différents récepteurs sensoriels sous forme de différents codages avec des potentiels d'actions pour les animaux. On a donc à faire avec des circuits distribués de communication et d'opérations en contact avec des environnements divers et fluctuants appelés à déterminer de façon fiable le devenir du système lui-même.

Au delà des différences radicales entre leurs composants, cellules, molécules et gènes qui en changent fondamentalement la nature, les réseaux génétiques et les réseaux de neurones partagent en Tout cas des méthodes et outils d'étude, si ce n'est peut-être aussi certaines propriétés générales. Des structures et phénomènes qui s'étudient par une panoplie commune d'outils théoriques relevant de la physique statistique, des systèmes dynamiques, de la théorie des graphes et des réseaux complexes et qui impliquent des modélisations mathématiques et des méthodes de résolution et simulation similaires.

Biologie des systèmes et neurosciences théoriques : challenges de modélisation. Un des défis majeurs, aussi bien dans l'étude des systèmes cognitifs génétiques (biologie des systèmes) et neuronaux est bien de déterminer à la fois quelles sont les propriétés logiques de ces systèmes et comment matériellement elles prennent naissance. Ce programme implique de faire le lien entre toutes les échelles des phénomènes et relier la modélisation des mécanismes à celles des propriétés fonctionnelles. La modélisation est alors confrontée à différentes limitations : la nécessité de connaissance suffisantes et pertinentes sur les systèmes, la difficulté de réaliser des abstractions pertinentes entre les différents éléments et les différentes échelles des systèmes impliqués, la complexité numérique des problèmes qui empêche de les calculer ou de les simuler correctement.

La production massive de données biologiques ne pâlie que partiellement au manque de connaissances. Elle nécessite un filtrage éclairé des informations pertinentes et utiles. Un travail lié à l'abstraction nécessaire à l'établissement de modèles informatifs. Une bonne connaissance de la modélisation et de la biologie semble alors indispensable. Une autre limitation des connaissances préliminaires aux modèles tient à leurs inhomogénéité : précises et abondantes sur cer-

tains aspects, incertaines et manquantes sur d'autres. La modélisation doit apprendre à travailler avec cette irrégularité des connaissances. Les méthodes classiques de modélisation ne savent le faire que très partiellement. C'est une des raisons du développement des approches Bayésiennes qui peut-être observé. Une meilleure prise en compte dans les modèles de notre confiance dans chacun de leurs éléments. Enfin, il n'est pas toujours possible d'exploiter les modèles. En effet, même si une représentation formalisée de nos savoirs sur le système étudié a pu être construite, la complexité calculatoire du modèle peut le rendre impossible à résoudre ou simuler même sur des ordinateurs modernes.

Vers des nouvelles approches de modélisation. Les modèles sous forme mathématiques sont souvent construits sur des sous-systèmes délimités précis permettant d'éclairer un ou quelques phénomènes qui s'y déroulent à une échelle donnée. Le succès de nombreuses modélisations de circuits biochimiques, neuronaux, de processus de morphogenèse ont reposé sur la délimitation possible d'un système suffisamment cohérent permettant de faire l'abstraction d'autres éléments et échelles naturelles de son environnement. Ceci est ce qui permet la construction de modèles réduits, avec un petit nombre d'éléments manipulables par des méthodes mathématiques classiques. Ces approches, faisant appel à des systèmes d'équations différentielles ordinaires ou partielles, à des relations algébriques, aux méthodes de bifurcation des systèmes dynamiques seront toujours très pertinentes dans toutes les situations suffisamment claires (même si leur application peut-être techniquement bien difficile).

Cependant, il est clair que la science ne s'intéresse plus de nos jours uniquement à la recherche de systèmes modèles suffisamment délimités pour être solvables avec un ensemble donné de techniques établies. Les ressources technologiques actuelles conduisent à l'ambition d'étudier des systèmes de grande taille, hétérogènes et combinant des processus d'échelles temporelles et spatiales multiples. On a bien la sensation que l'étude des systèmes biologiques nécessite de prendre en compte une intégration forte entre composants divers de petite et grande taille, lents et rapides et possédant des rôles multiples suivant les contextes : des pièces des différents types de réseaux vivants, gènes, biomolécules, cellules, neurones intervenant dans une multiplicité de fonctions différentes. C'est ce que mets en évidence par exemple mon étude sur les motifs génétiques à la section 5.5.2.3 : les éléments de circuits génétiques sont partagés entre nombreux processus. Les systèmes biologiques n'ont peut-être pas la modularité que l'on voudrait qu'ils aient pour en faciliter la modélisation et l'analyse. Il est bien difficile de décider de la pertinence d'une abstraction d'un modèle. La réduction au petit circuit des gènes p53-mdm2 qui m'a tellement occupé (5.5.3.1) est-elle, par exemple, vraiment pertinente pour la situation biologique ? Elle l'était nécessaire pour en permettre un traitement, déjà par trivial mathématique. Mais hors de l'accord, très partiel avec quelques résultats expérimentaux particuliers, qu'apprend t-on réellement sur l'organisation biologique ? Les mêmes approches mathématiques qu'utilisés dans ce cas pourront encore être utilisées pour des modèles un peu plus sophistiqués, mais guère vraiment plus complets.

Pour devenir utiles les modèles particuliers devront pouvoir s'intégrer dans des modèles plus larges, combinant processus de natures différentes et différentes échelles. Dans cette optique les approches traditionnelles de la modélisation mathématique et de physique sont très probablement insuffisantes. Les formalismes d'équations mathématiques ne sont plus adéquats seuls. De nouveaux formalismes de modélisation sont nécessaires pour permettre une abstraction suffisamment expressive et flexibles des systèmes étudiés. Ces formalismes sont couplés à des outils numériques d'une part pour leur manipulation et la conception sur ordinateurs des modèles et d'autre part pour la simulation numériquement efficace de ceux-ci. L'informatique est ici centrale. En tant que science de l'information pour la conception de formalisme et en tant que science du calcul automatique pour la conception des algorithmes de calcul nécessaires.

Biologie des systèmes et neurosciences vers une convergence méthodologique Un questionnement fondamental commun, un outillage théorique similaire, une conjonction d'apports tech-

nologiques dans les deux domaines on peut s'attendre à une communauté importante travaillant entre biologie des systèmes et neurosciences. Mais comme aussi relevé par [SCHUTTER et DE SCHUTTER](#), en dépit des convergences les communautés scientifiques respectives sont bien distinctes. Peut-être en raison de l'historique différents entre les disciplines. S'étant surtout développées depuis le début des années 2000, à l'âge de l'internet et des réseaux sociaux, la biologie des systèmes a intégré de nombreuses méthodes modernes de programmation et des usages d'ouverture et de partage. En comparaison, les neurosciences théoriques sont restées plus longtemps dans un schéma plus classique avec les deux principaux simulateurs neuronaux encore de nos jours NEURON [[CARNEVALE et HINES, 2005](#); [HINES, 1989](#)] et GENESIS [[BOWER et BEEMAN, 1998](#)] développés à partir des années 1980 avec des codes plus fermés et une compatibilité des modèles problématique entre simulateurs différents. Des initiatives d'ouverture et de standardisation pour le partage de données et de modèle se développent aussi dans les neurosciences théoriques et numériques, mais sont plus récentes [[DAVISON, 2008](#); [HINES et collab., 2004](#); [RICHMOND et collab., 2014](#)].

La biologie des systèmes a donné lieu depuis longtemps déjà à de nombreuses initiatives et recherches pour des outils généraux et flexibles. Différents langages comme Kappa et Biocham, discutés à la section 4.4.3 explorent des moyens d'écriture et de représentation adaptés pour la complexité des systèmes biochimiques et génétiques. Des réflexions et projets dans le même esprit semblent en cours également dans les neurosciences visant à développer des formalisations pour représenter et tester des hypothèses sur les mécanismes cérébraux intégrant tous les niveaux des processus neuronaux aux fonctions cognitives [[PALMERI et collab., 2017](#)].

Vers plus de mécanique! Réseaux génétiques et réseaux neuronaux m'apparaissent donc comme deux situations biologiques de systèmes cognitifs, agissant sur des formes d'information différentes. Leurs logiques respectives peuvent partager ou non des principes constitutifs et opérationnels qu'il nous appartient de caractériser. Cependant aux yeux de physiciens il importe aussi outre les propriétés mathématiques de caractériser l'action des facteurs mécaniques et matériels, les conséquences des lois physiques de la nature qui se manifestent sur ces systèmes et en guident développement et propriétés. Une problématique très insuffisamment prise en compte dans l'essentiel de la littérature biologique aussi bien en biologie des systèmes qu'en neurosciences. Dans ces deux domaines, si on fait abstraction de la caractérisation des structures particulières de chacun, l'attention des chercheurs se focalise essentiellement sur les mécanismes et la logique des processus.

On s'intéresse aux principes de codage, de câblage et de traitement des informations. En perdant facilement de vue les contraintes physico-chimiques qui interviennent. C'est à ce niveau que les physiciens peuvent utilement intervenir. Ils peuvent compléter les visions centrées sur l'information des généticiens, neurobiologistes et nombreux informaticiens théoriques participant à ces études. La considération de contraintes physiques s'imposant globalement pour l'organisation du génome est ainsi l'idée directrice des travaux que j'ai mené avec Reiner Veitia sur les effets de nombre de copie de gènes et de ploïdie présentés dans la section 5.5 du chapitre 5 qui en font leur originalité. Des effets de nombre, de volume, de concentration déterminent des conditions prises en compte par l'évolution naturelle pour le fonctionnement des génomes. Un autre cas de contraintes globales qui s'imposent à toutes les constantes cinétiques des voies de signalisation cellulaires est l'exemple que j'ai seulement mentionné dans ce manuscrit, au chapitre 5, section 5.7 de la réponse au stress hyper-osmotique chez la levure : tant que la cellule ne s'est pas adaptée la pression osmotique induit une diminution du volume cellulaire qui provoque un encombrement moléculaire affectant de manière globale les paramètres cinétiques du métabolisme. Dans le cadre des cultures neuronales une autre situation de contraintes physico-chimiques s'appliquant au développement du réseau est celle s'imposant à l'organisation de la culture en développement. Ces situations sont quelques exemples seulement, qui doivent être encore développés. La génomique, la biologie des systèmes et les neurosciences théoriques sont en effet des disciplines

traitant fondamentalement d'information, mais possèdent une implémentation physique. L'illustration de principes d'organisation physique y agissant est difficile et je n'ai fait que de l'effleurer.

Je conclurai ainsi ce mémoire avec cette remarque et la motivation à chercher davantage encore dans mes travaux futurs les déterminants matériels, physiques qui guident le développement des systèmes logiques du vivant, qu'ils soient biochimiques ou neuronaux.

8.1 Références

- BOWER, J. M. et D. BEEMAN. 1998, *The Book of GENESIS : Exploring Realistic Neural Models with the GEneral NEural Simulation System*, Springer New York, ISBN 9781461216346, 458 p.. [344](#)
- CARNEVALE, N. T. et M. L. HINES. 2005, *The NEURON book*, Cambridge University Press, ISBN 9780511541612, 457 p.. [344](#)
- DAVISON, A. P. 2008, «PyNN : a common interface for neuronal network simulators», *Frontiers in Neuroinformatics*, vol. 2, doi :10.3389/neuro.11.011.2008, ISSN 16625196. [344](#)
- HINES, M. 1989, «A program for simulation of nerve equations with branching geometries», *International journal of bio-medical computing*, vol. 24. [344](#)
- HINES, M. L., T. MORSE, M. MIGLIORE, N. T. CARNEVALE et M. L. HINES. 2004, «ModelDB : A Database to Support Computational Neuroscience», *Journal of Computational Neuroscience*, vol. 17, n° 1, doi :10.1023/B:JCNS.0000023869.22017.2e, p. 7–11, ISSN 0929-5313. [344](#)
- PALMERI, T. J., B. C. LOVE et B. M. TURNER. 2017, «Model-based cognitive neuroscience», *Journal of Mathematical Psychology*, vol. 76, n° Pt B, doi :10.1016/j.jmp.2016.10.010, p. 59–64, ISSN 00222496. [344](#)
- RICHMOND, P., A. COPE, K. GURNEY et D. J. ALLERTON. 2014, «From Model Specification to Simulation of Biologically Constrained Networks of Spiking Neurons», *Neuroinformatics*, vol. 12, n° 2, doi :10.1007/s12021-013-9208-z, p. 307–323, ISSN 1539-2791. [344](#)
- SCHUTTER, E. D. et E. DE SCHUTTER. 2008, «Why Are Computational Neuroscience and Systems Biology So Separate?», *PLoS Computational Biology*, vol. 4, n° 5, doi :10.1371/journal.pcbi.1000078, p. e1000078, ISSN 1553-7358. [342](#), [344](#)

Quatrième partie

Annexe

Annexe A

Glossaire

ADN poubelle (junk DNA) Qualificatif donné aux fractions de séquence d'ADN dans les génomes constitués en grande partie d'éléments mobiles répétés et sans rôle fonctionnel apparent. [I](#)

algorithme génétique algorithme d'optimisation (recherche d'un extrémum) basé sur un principe évolutif avec l'itération d'opérations d'évolution stochastique et de sélection sur une population de jeux de paramètres. L'exploitation de l'espace des phases par ce processus d'évolution sélection permet de se rapprocher des optima. [I](#)

allèle variante génétique d'un gène. Dans un organisme diploïde le génotype est formé de deux gènes pouvant être de variantes alléliques différentes. [I](#)

allèle récessif allèle tel que le phénotype de l'hétérozygote est déterminé par l'autre allèle. Le phénotype associé à l'allèle récessif n'est exprimé que si l'individu porte deux copies de celui-ci, donc de génotype aa. [I](#)

allopolyploidization, Processus de formation d'un organisme allopolyploïde qui possède des copies multiples d'ensembles de chromosomes venant d'espèces différentes. [I](#)

altération épigénétique Mécanismes sur l'ADN qui affectent l'expression des gènes à proximité. L'épigénétique agit principalement en affectant l'accessibilité des gènes par le biais des protéines, notamment histones liés à l'ADN. [I](#)

analyse de stabilité Étude des points fixes d'un système dynamique, de leurs domaines d'existence et propriétés de comportement du système dans leur voisinage (point attractif, répulsif, périodique). [I](#)

aneuploïde se dit d'un organisme qui contient un nombre de chromosomes différents considéré comme « anormal » par rapport à sa ploïdie standard. [I](#)

apoptose mort cellulaire contrôlée (suicide cellulaire) en cas de dommages excessifs à une cellule. [I](#)

automate cellulaire Système de simulation basé sur un ensemble de d'unités réparties sur un réseau, les cellules, interagissant les unes avec les autres suivant leur voisinage avec des règles simples d'évolution. [I](#)

autopolyploïdie si les jeux de chromosomes proviennent de deux ou plusieurs génomes différents. [I](#)

bifurcation changement de nature du comportement d'un système dynamique. [I](#)

catalyse processus de destruction d'une macromolécule biologique. [I](#)

chaîne de Markov Description probabiliste de l'évolution avec un temps discret de processus stochastiques dépendant uniquement du dernier état. La prédiction probabiliste du futur dépend uniquement de l'état présent, sans mémoire des états antérieurs. [I](#)

- chemotaxis** Processus de navigation dans l'espace d'organismes vivants guidé par la détection dans l'environnement de composés chimiques. [I](#)
- chromatine** nom donné à l'ADN chromosomique enroulé autour des protéines d'histones et qui peut se trouver dans différents états de conformation, ouverte (euchromatine) ou fermée (hétérochromatine) sous l'effet d'altération épigénétiques. [I](#)
- chromatine (remodelage)** processus de modification dynamique de la chromatine qui permet une régulation de l'expression génétique sur de longues durées et transmissible durant la division cellulaire. [I](#)
- cis-régulation** Mécanismes de régulation impliquant plusieurs gènes et séquences génétiques portées par le même chromosome. À la différence de la trans-régulation qui fait intervenir des interactions entre gènes et séquences de chromosomes différents. [I](#)
- clade** groupe d'organismes vivants formé par un ancêtre et l'ensemble de ses descendants, il représente donc une totalité de descendance. [I](#)
- Coopérativité** Dans le contexte de la régulation génétique, se dit de différents régulateurs agissant avec des effets plus qu'additifs. Ainsi la présence initiale d'un premier régulateur accroît la probabilité de présence du second par rapport à une liaison indépendante. [I](#)
- cycle cellulaire** Cycle de vie d'une cellule entre sur une séquence de reproduction incorporant une mitose. [I](#)
- cycle-circadien** Cycle de vie quotidien de cellules et organismes vivants sur à peu près 24 h en phase avec le jour. [I](#)
- délétion** mutation génétique caractérisée par la perte de matériel génétique sur un chromosome, ou d'un chromosome en entier. [I](#)
- déphosphorylation** processus de libération d'un groupe chimique phosphate depuis une protéine phosphorylée. [I](#)
- diauxique (transition)** changement dans le métabolisme cellulaire quand une culture a consommé le glucose disponible. [I](#)
- dimère** molécule comportant deux sous-unité. [I](#)
- dominance génétique** caractéristique d'un allèle dans le cadre de l'héritage Mendélien de déterminer seul un phénotype. [I](#)
- double hybride (technique)** technique de biologie moléculaire permettant la détection d'interactions protéines-protéine. [I](#)
- élément transposable** Séquence d'ADN capable de se déplacer de manière autonome dans un génome, aussi appelé transposon. [I](#)
- enzyme de restriction** Protéine capable de couper une molécule d'ADN au niveau d'une séquence spécifique particulière de nucléotides caractéristique. Les enzymes de restriction sont des mécanismes de défense génétiques à des infections virales et sont des outils majeurs de génie génétique. [I](#)
- épigénétique** correspond à des mécanismes de modification de manière éventuellement réversible de l'expression des gènes en contrôlant leur accessibilité dans le génome sans changer la séquence de l'ADN. Les modifications épigénétiques peuvent être adaptatives à des conditions environnementales, transmissible lors de la division cellulaire et dans certains cas dans les cellules germinales. [I](#)
- espace des phases** espace multi-dimensionnel formé de variation de l'ensemble des variables d'un système dynamique, chaque variable correspondant à l'une des dimensions. [I](#)
- espérance maximisation (algorithme)** L'algorithme d'optimisation d'espérance-maximisation (expectation-maximization algorithm) permet une recherche itérative d'optima qui comporte la répétition de cycles de calcul d'une fonction de l'espérance de la vraisemblance des paramètres

issus de l'itération précédente et suivie de la recherche des variations de ces paramètres permettant de la maximiser. [I](#)

état quasi-stationnaire pour un système dynamique possédant des échelles de temps très différentes, c'est un état pour lequel les variables rapides (caractérisées par l'échelle de temps la plus courte) sont quasiment constantes, alors que les variables lentes connaissent encore une évolution. [I](#)

état stationnaire états invariants dans le temps d'un système dynamique. Les états stationnaires s'ils existent sont au croisement des nullclines. [I](#)

euchromatine Forme décondensée, ouverte de la chromatine permettant l'accès de la machinerie de transcription aux gènes. [I](#)

euthérien Classe de mammifères possédant une gestation avec un placenta .XXXXXX CHECK. [I](#)

FACS Fluorescence Activated Cytometry Sorting : technique de tri de cellules uniques à partir de l'analyse de signaux optiques et de fluorescence. Cette technique permet l'acquisition de statistiques de caractéristiques mesurées individuelles sur les cellules d'une population. [I](#)

Flux Balance Analysis méthode mathématique de simulation du métabolisme. Elle calcule les flux stationnaires de transformation de la biomasse le long des voies de réactions métaboliques. La méthode est numériquement très efficace car elle suppose le réseau métabolique opérant à l'état stationnaire, ce qui réduit le modèle à un ensemble d'équations linéaires. [I](#)

génomique Étude du génome, c'est-à-dire de l'ensemble de l'information de séquence d'ADN d'un organisme. [I](#)

génotype La composition allélique de tous les gènes d'un individu, qui peut aussi inclure l'ensemble de l'ADN non codant. [I](#)

germinale (cellule) cellules d'un animal ou d'un végétal susceptibles de former les gamètes pour la transmission de l'information génétique servant à la génération de la descendance. [I](#)

GFP Green Fluorescent Protéine, protéine fluorescente utilisé comme rapporteur d'expression de protéine. La séquence de la GFP peut-être fusionnée à la séquence codant la protéine d'intérêt dont l'expression produit une protéine hybride marquée par fluorescence et détectable. [I](#)

haploïde se dit d'une cellule ou d'un organisme si son jeu complet de chromosomes est d'une seule copie chacun. [I](#)

haploinsuffisance phénotype anormal pouvant survenir quand un ou plusieurs gènes ne sont plus présents dans le nombre de copie de l'haplotype normal, c'est-à-dire du nombre de copies des chromosomes et de chaque gène considéré comme normal pour l'espèce. [I](#)

Hemiascomycete classe dans le règne des champignons qui inclut les différents types de levures. [I](#)

hétérochromatine Forme condensée, fermée de la chromatine empêchant l'accès de la machinerie de transcription aux gènes. [I](#)

hétérotrophe Désigne un organisme vivant nécessitant de se nourrir de constituants organiques préexistants. [I](#)

hétérozygote Individu qui est porteur pour un gène donné (un locus génétique) de deux allèles différents. [I](#)

histone Protéines étroitement associées à l'ADN dans les chromosomes et qui déterminent l'organisation de celui-ci. L'ADN est enroulé autour des histones sous une forme compacte ou empêchant, ou non l'accès de parties de celui-ci à la machinerie d'expression génétique. [I](#)

holoenzyme Enzyme composée de l'assemblage complexe de plusieurs protéines différentes. [I](#)

homozygote Individu qui est porteur pour un gène donné (un locus génétique) de deux allèles identiques. [I](#)

immunoprécipitation technique de biologie moléculaire pour isoler et concentrer (précipiter) une protéine particulière d'un échantillon contenant de nombreux types en utilisant une protéine antigène capable de se lier spécifiquement à la cible.. [I](#)

kinase toute enzyme permettant la phosphorylation. [I](#)

lien peptidique interaction chimique qui relie les acides aminés au sein d'une protéines. Les liens peptidiques sont de différentes natures mettant en œuvre des interactions chimiques et forces différentes (lien hydrogène, van der Waals, covalent, hydrophobe). [I](#)

lyse processus de destruction ou de mort de tissus biologiques ou de cellules par des processus chimiques, physiques ou biologiques. [I](#)

lysogénie Processus dans les bactéries d'incorporation dans le génome de la séquence d'un virus bactérien (phage) qui va y résider de manière latente jusqu'à son activation éventuelle. [I](#)

Metabolic Control Analysis Technique d'analyse de réseaux métaboliques, de signalisation et génétiques pour déterminer comment les variables, flux de biomasse, concentrations, dépendent des paramètres. S'apparente à l'analyse de sensibilité des paramètres. [I](#)

métazoaire Nom de clade désignant les animaux. Les organismes ainsi qualifiés sont des eucaryotes (ont des cellules avec un noyau), multicellulaires, hétérotrophes. Ces hétérotrophes ne se nourrissent pas par absorption, contrairement aux champignons. [I](#)

méthylation processus d'addition d'un groupe méthyl (CH₃-) à l'ADN qui conduit à la fixation préférentielle d'histone et une fermeture de la chromatine. [I](#)

méthylation (schéma de) Distribution de l'ADN méthylé dans le génome qui détermine les régions non accessibles à l'expression génétique. [I](#)

méthylation de l'ADN Processus d'addition d'un groupe méthyl (CH₃-) à l'ADN agissant sur la liaison d'histones et favorisant une forme compacte de la chromatine. [I](#)

microarray où puce à ARN, technique à haut débit permettant la mesure simultanée de niveau d'expression de milliers d'ARN message à partir d'échantillons. [I](#)

microfabrication Ensemble de techniques permettant la réalisation de structures à l'échelle sub millimétrique appliquées notamment en microfluidique. Il s'agit essentiellement de méthodes basées sur la photolithographie à partir de plans de conception sur des surfaces sensibles qui après traitement chimique forment un tampon gravé des schémas souhaités, qui peut-être utilisé soit pour l'impression de motifs sur un substrat, soit comme moule pour le moulage des structures à l'aide de gels. [I](#)

microfluidique Ensemble de techniques pour la conception et la fabrication de réacteurs chimiques et d'environnements de culture cellulaires structurés à l'échelle sub-millimétriques ainsi qu'un contrôle chimique précis de ces milieux. [I](#)

mitose Les processus de duplication et division cellulaire. [I](#)

morphogénèse Processus du développement des organismes de formation des organes et des formes. [I](#)

multimère molécule comportant plusieurs sous-unités. [I](#)

nullcline courbe dans l'espace des phases pour chaque variable d'un système dynamique sur laquelle la variable est stationnaire (elle est constante). Les nullclines associées à chaque variable sont définies par la condition d'une dérivée nulle de la variable correspondante. [I](#)

Ohnologue se dit d'un gène dupliqué venant d'un événement de duplication du génome entier. Le terme Ohnologue fait référence à Susumu Ohno (1928-2000) qui a proposé l'existence de deux événements de duplications génomiques dans l'arbre ancestral des vertébrés. [I](#)

- oncogène** Facteurs chimiques favorisant la survenue de cancers. Catégorie de gènes dont l'activité ou le dysfonctionnement est lié la survenue de cancers. [I](#)
- PCR** réaction en chaîne par polymérase (PCR : polymerase chain reaction), technique de biologie moléculaire permettant l'amplification (multiplication) de faibles quantités de séquences d'ADN *in vitro* avec des facteurs de multiplication très élevés (ordre du milliard). La PCR est sélective et est ajustées pour l'amplification de séquences spécifiques, permettant de l'utiliser à des fins de détection très sensibles de matériel génétique.. [I](#)
- péptide** polymère d'acides aminés reliés entre eux par des liaisons peptidiques. [I](#)
- perte de fonction (mutation)** mutation de nucléotides sur la séquence d'ADN codante d'un gène portant à la perte complète de la fonction de celui-ci. [I](#)
- phénotype** L'ensemble des traits observables d'un organisme. Parfois utilisé restreint au contexte d'un ou quelques traits. [I](#)
- phosphatase** toute enzyme permettant la déphosphorylation. [I](#)
- phosphorylation** processus de transfert d'un groupe chimique phosphate à partir d'une molécule donneuse riche, notamment l'ATP, vers une molécule cible, en particulier protéine. La phosphorylation est un processus majeur d'activation de protéines et de signalisation cellulaire. [I](#)
- plasmide** Molécule d'ADN distincte de l'ADN chromosomique et capable de répllication autonome. Les plasmides, présents en plusieurs copies quasi exclusivement au sein de cellules bactériennes, sont porteurs de gènes conférant des avantages sélectifs et interviennent transferts d'information génétiques entre cellules de la même espèce ou différentes. Ces molécules dont des outils majeurs du génie génétique. [I](#)
- pléiotropie** propriété pour un gène ou une protéine d'être associé à plusieurs fonctions différentes. [I](#)
- pléidie** nombre de copies de l'ensemble des chromosomes, donc du nombre de loci (positions) pour chaque gène pouvant porter chacun un forme allélique différente. [I](#)
- polymérase** unité moléculaire enzymatique intervenant dans la lecture de l'information génétique d'ADN d'un gène et la synthèse d'une molécule d'ARN messenger. [I](#)
- Potentiel d'action** Signal électrique émis par un neurone sous forme d'un pulse de dépolarisation de la membrane cellulaire. En anglais: spike. [I](#)
- protéase** nom génériques des enzymes qui réalisent la protéolyse, c'est-à-dire la digestion, on dit aussi la catalyse, en petits fragments d'une protéine par hydrolyse des liens peptidiques. Il existe plus de 80 familles d'enzymes protéases. [I](#)
- proteomique** Étude du protéome, c'est-à-dire de l'ensemble des protéines exprimées à un instant donné, dans des conditions données, dans un organisme. [I](#)
- pyllotaxie** Désigne les propriétés de l'arrangement des feuilles sur une branche de plante qui en détermine l'aspect global. [I](#)
- Réseau Bayésien** Graphe orienté de propagation de probabilités conditionnelles entre variables aléatoires. Représente l'évolution probabiliste de différentes variables état d'un système. [I](#)
- réseau phloémien** Réseau des veines chez les plantes de circulation de la sève. [I](#)
- réseau trophique** Réseau écologie des chaînes alimentaires, représentant la consommation alimentaire entre espèces. [I](#)
- résidu** Partie d'un acide aminé composant une protéine qui ne participe pas aux liaisons faisant la structure d'une protéine. [I](#)
- ribosome** unité moléculaire enzymatique intervenant dans la lecture de l'information génétique de molécules d'ARN messenger et la synthèse de protéines. [I](#)

RT-PCR (Reverse Transcriptase PCR) technique de biologie moléculaire permettant de mesurer l'activité transcriptionnelle. La RT-PCR associe une transcription inverse (RT) suivie d'une PCR. Elle permet de synthétiser le brin complémentaire d'ADN d'un ARN messager qui est ensuite amplifié par PCR. [I](#)

sénescence cellulaire processus par lequel une cellule cesse de se diviser. [I](#)

somatique (cellule) sont toutes les cellules formant le corps d'un organisme multicellulaire à l'exclusion des cellules germinales qui transmettent l'information génétique à la descendance. C'est-à-dire toutes les cellules n'appartenant pas à la lignée germinale : gamètes, cellules germinales. [I](#)

spéciation Processus évolutif par lequel de nouvelles espèces vivantes se forment à partir d'ancêtres communs. [I](#)

spike Terme en anglais pour potentiel d'action. [I](#)

Synergie Dans le contexte de la régulation génétique, se dit de différents régulateurs agissant simultanément avec des effets simplement additifs. [I](#)

synthèse processus de production d'une macromolécule biologique. [I](#)

système dynamique en mathématiques et physique un système dynamique est une relation qui décrit la dépendance temporelle d'un point dans l'espace des phases. [I](#)

TAP tag (technique) Tandem Affinity Purification: technique de biologie moléculaire de purification pour étudier les interactions protéines-protéines. Elle implique la fusion des protéines avec une molécule spécifique, le marqueur TAP. [I](#)

taux métabolique vitesse de transformation de l'énergie dans un organisme vivant. [I](#)

taxon groupe de tous les organismes vivants possédant en commun certains caractères taxinomiques ou diagnostiques bien définis. L'espèce constitue le taxon de base de la classification systématique. Plus le rang du taxon est élevé et plus le degré de ressemblance des individus concernés est faible, c'est-à-dire plus le nombre de caractères que des plantes, animaux, champignons, bactéries ou chromistes ont en commun entre eux est faible, et inversement. [I](#)

tetraploïde se dit d'un organisme qui contient quatre jeux de ses chromosomes différents. [I](#)

titrage Méthode d'analyse chimique consistant à utiliser une solution de concentration volumique connue (appelée titrant) afin de déterminer la concentration d'une espèce (appelée analyte ou espèce titrée) contenue dans une solution de concentration inconnue. La quantité mesurée du réactif titrant nécessaire pour neutraliser l'analyte permettra de déterminer la concentration recherchée. [I](#)

traduction Processus de synthèse d'une protéine à partir de l'information sur une molécule d'ARN par un ribosome. [I](#)

transcription Processus de synthèse de l'ARN messager à partir de la lecture de l'information sur la séquence codante d'ADN d'un gène par le « complexe de transcription » formé essentiellement par une polymérase. [I](#)

transcriptomique Étude du transcriptome, c'est-à-dire de l'ensemble de l'activité de transcription en cours à un instant donné, dans des conditions données, dans un organisme. [I](#)

transduction de signal processus de signalisation moléculaire entre la perception de conditions environnementales moléculaires captées par des récepteurs sur la membrane cellulaire. [I](#)

transposon Élément transposable dans une séquence d'ADN capable de se déplacer de manière autonome dans un génome, par le mécanisme de transposition sous l'effet d'enzymes « transposases » sur certaines séquences spécifiques délimitant la séquence transposable. [I](#)

trans-régulation Mécanismes de régulation impliquant plusieurs gènes et séquences génétiques sur des chromosomes différents. À la différence de la cis-régulation qui fait intervenir des interactions entre gènes et séquences sur un même chromosome. [I](#)

ubiquitine ligase enzyme liant un marqueur, l'ubiquitine, sur des protéines favorisant leur reconnaissance par des protéases et leur dégradation. [I](#)

Valeur C On appelle valeur C (C-value) la quantité d'ADN dans le génome haploïde d'un organisme, c'est-à-dire dans un jeu de copie seul de ses chromosomes. [I](#)

whole Genome Duplication Duplication de génome entier, processus durant la méiose de non dissociation des chromosomes qui conduit à la polyploïdie, c'est-à-dire une cellule possédant des copies additionnelles du génome entier. [I](#)

zygote Cellule chez les eucaryotes, résultat de la reproduction sexuée provenant de la fusion des deux gamètes, un ovule et un spermatozoïde. [I](#)