

Numéro d'ordre 2008-006

Année 2008

Apprivoiser la vie : Modélisation individu-centrée de systèmes biologiques complexes

Habilitation à Diriger des Recherches présentée par
Guillaume Beslon, Docteur en Informatique

Devant

L'Institut National des Sciences Appliquées de Lyon

Pour obtenir

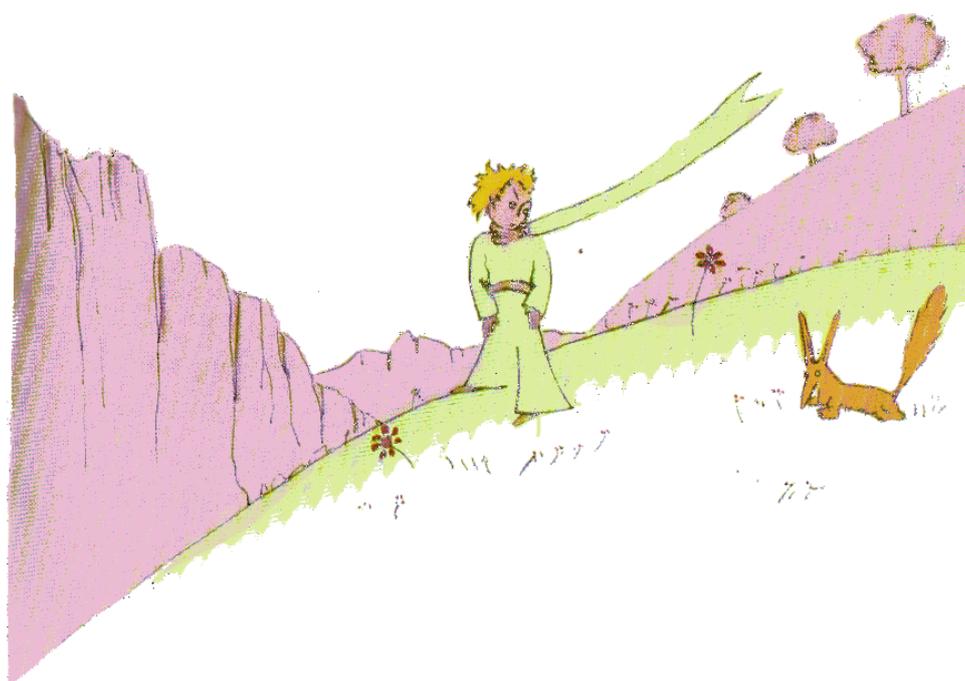
Le grade d'HDR

Formation doctorale

Informatique et Information pour la Société (EDIIS)

Soutenue le 14 novembre 2008 devant le jury composé de :

Gilles Bernot	Professeur, Université de Nice (rapporteur)
Paul Bourguin	Professeur, Ecole Polytechnique de Paris
Joël Favrel	Professeur émérite, INSA Lyon
Jean-Michel Fayard	Professeur, INSA Lyon
François Képès	Directeur de recherche, Université d'Évry
Alain Mille	Professeur, Université Claude Bernard - Lyon 1
Michel Morange	Professeur, Ecole Normale Supérieure (rapporteur)
Michel Morvan	Professeur, Ecole Normale Supérieure de Lyon (rapporteur)



- Qu'est-ce que signifie "apprivoiser" ? [demanda le petit prince]
 - C'est une chose trop oubliée, dit le renard. Ça signifie "créer des liens ...".
 - Créer des liens ?
 - Bien sûr, dit le renard. Tu n'es encore pour moi qu'un petit garçon tout semblable à cent mille petits garçons. Et je n'ai pas besoin de toi. Et tu n'as pas besoin de moi non plus. Je ne suis pour toi qu'un renard semblable à cent mille renards. Mais, si tu m'apprivoises, nous aurons besoin l'un de l'autre. Tu seras pour moi unique au monde. Je serai pour toi unique au monde [...]
- Le renard se tut et regarda longuement le petit prince :
- S'il te plaît... apprivoise-moi ! dit-il.
 - Je veux bien, répondit le petit prince, mais je n'ai pas beaucoup de temps. J'ai des amis à découvrir et beaucoup de choses à connaître.
 - On ne connaît que les choses que l'on apprivoise [...]

Remerciements

Il est d'usage, dans un mémoire de thèse ou d'habilitation à diriger des recherches, d'utiliser la forme "nous" pour identifier l'auteur. J'aimerais profiter de ces remerciements pour préciser que cette forme revêt ici pour moi un caractère particulier. J'ai en effet eu la chance de travailler avec un grand nombre d'étudiants venant de différents horizons et les travaux présentés ici leur doivent tant que je ne peux écrire "nous" sans penser à eux. Qu'ils soient tous ici particulièrement remerciés. Certains ont bien évidemment été des collaborateurs beaucoup plus proches que d'autres, que ce soit pour des raisons de temps (comment comparer un stage de deux mois et une thèse) d'implication personnelle ou d'intérêts scientifiques mais tous m'ont été précieux. Je les cite ici par ordre d'arrivée dans "l'équipe" (qui n'aurait pas d'existence sans eux); tous sauront reconnaître la part que ce mémoire leur doit.

Un grand Merci donc à : *Anders Jonson, Carine Bournez, Jean Laurens, Hédi Soula, Loïc Chrisostome, Frédéric Henry, Nancy Reymond, Sylvain Blachon, Sébastien Gripon, François Perrin, Carole Knibbe, Virginie Lefort, Kevin Frugier, Ludovic Potier, Thomas Tempé, Matthieu Bontemps, Aravind Alwan, Frédéric Mayot, Anis Tissaoui, Jaber Kouki, Antoine Coulon, Yolanda Sanchez-Dehesa Fernandez, Damien Peelman, Gaël Kaneko, David Parsons, Elie Rotenberg.*

Certains d'entre eux m'ont en outre fait l'amitié de relire tout ou partie de ce mémoire. Qu'ils apprennent ici que leurs remarques, critiques, conseils ou encouragements m'ont touché au plus haut point.

Un grand merci à mes rapporteurs, Gilles Bernot, Michel Morange et Michel Morvan, qui m'ont fait l'honneur d'évaluer ce mémoire. Merci aussi à Paul Bourguine, François Képès et Alain Mille qui ont accepté de participer à mon jury. Je sais combien leur temps à tous est compté et je leur en suis d'autant plus reconnaissant.

Ce travail n'aurait pas été possible sans un certain nombre de soutiens.

Joël Favrel m'a fait l'honneur de croire très tôt en cette démarche de modélisation. Son accueil et son ouverture d'esprit ont permis à ce travail de prendre son essor.

Jean-François Boulicaut a exercé une pression amicale mais ferme. Il m'a permis de ne pas me fourvoyer dans les circonvolutions d'un rapport qui sans lui aurait été un fleuve sans fond ni fin.

Ce que je dois à Jean-Michel Fayard dépasse de loin le cadre de ce mémoire. Il aura été un guide et un garde-fou; il est surtout devenu un ami cher.

Parmi tous les biologistes avec lesquels j'ai été amené à travailler (qu'ils en soient tous chaleureusement remerciés), deux têtes dépassent tout particulièrement. Chacun à leur

manière, Olivier Gandrillon et Hubert Charles ont eu, ont et auront encore une influence fondamentale sur mon parcours scientifique.

Des remerciements particulièrement appuyés s'envolent vers Nancy. Bertrand aura été un relecteur attentif de ce mémoire comme il l'avait déjà été pour mon mémoire de thèse. Il est je crois la seule personne à avoir consacré autant d'attention aux quelques 471 pages que représentent ces deux mémoires.

Il est 23h07. Je suis devant mon ordinateur. Je n'aurai jamais assez de mots pour remercier Laurence pour sa patience durant la rédaction de ce mémoire qui fut long avant d'être court ...

Table des matières

Introduction	11
I Modélisation individu-centrée de systèmes biologiques complexes	13
1 Introduction	13
2 Système biologique complexe	14
2.1 De la complexité	14
2.2 Système biologique complexe et biologie des systèmes	19
3 De la modélisation à la modélisation individu-centrée	22
3.1 Qu'est-ce qu'un modèle ?	22
3.2 Le cas de la simulation	25
3.3 La modélisation individu-centrée	25
3.4 Conclusion : modèles individu-centrés de systèmes biologiques complexes	35
4 Une méthodologie pragmatique pour la modélisation individu-centrée en biologie	37
4.1 Introduction	37
4.2 La modélisation individu-centrée, une modélisation expérimentale	38
4.3 Quelques règles pour la modélisation individu-centrée	40
4.4 Le cycle de vie d'un modèle individu-centré	47
5 Conclusion : une position paradoxale	51
II Modèles	53
1 Introduction	53
2 Modèles de l'énaction	55
2.1 Introduction	56
2.2 Problématique	56
2.3 Modèles et expériences	60
2.4 Résultats	63
2.5 Conclusion et perspectives	69
3 Modèles de l'évolution	71
3.1 Introduction	72
3.2 Problématique	72
3.3 Modèles et expériences	74
3.4 Résultats	78
3.5 De Aevol à RAevol : évolution des réseaux de régulation génétiques	82
3.6 Conclusion et perspectives	84

4	Vers des modèles cellulaires 4D	87
4.1	Introduction	88
4.2	Problématique	88
4.3	Modèles et expériences	94
4.4	Résultats	99
4.5	Conclusion et perspectives	102
Conclusion		103
Bibliographie		107
A Notice biographique		125
1	Informations Personnelles	125
2	Situation professionnelle	125
3	Cursus universitaire – Expérience professionnelle	125
4	Activités Scientifiques	126
5	Activités d’Encadrement Scientifique (depuis 1998)	127
5.1	Thèses	127
5.2	Stages de DEA/Master Recherche	128
6	Autres activités scientifiques	129
6.1	Organisation de conférences scientifiques	129
6.2	Conférences invitées	130
6.3	Évaluation scientifique - Jury de thèses	131
6.4	Vulgarisation scientifique	131
7	Activités d’Enseignement (pour l’année 2007/2008)	132
7.1	Au département Informatique de l’INSA de Lyon	132
7.2	Au département BioSciences de l’INSA de Lyon	132
7.3	Enseignement de troisième cycle	132
8	Activités Administratives, Responsabilités Collectives	132
B Bibliographie personnelle classée		133
1	Revue internationale	133
2	Revue nationale, chapitres d’ouvrages	133
3	Conférences internationales avec comité de lecture	134
4	Autres conférences, résumés, posters	137

Table des figures

I.1	Diagramme de bifurcations de l'équation logistique	18
I.2	Le modèle 4M de la biologie des systèmes	20
II.1	Expérience d'apprentissage avec le Khepera (robot réel)	64
II.2	Réponse du réseau à un stimulus périodique	65
II.3	Influence de la STDP sur la dynamique du réseau	66
II.4	Expérience d'apprentissage avec le Khepera (simulation)	67
II.5	Expérience d'apprentissage avec le Khepera (résultats)	68
II.6	Activité du réseau en fonction de la distance aux murs	68
II.7	Transcription/traduction dans aevol	77
II.8	Évolution des paramètres structuraux	79
II.9	Organismes obtenus sous deux taux de mutation différents	80
II.10	Nombre de gènes et de bases non-codantes en fonction du taux de mutation et du taux de sélection	81
II.11	Potentiel d'interaction moléculaire	98
II.12	Influence de l'interaction protéine-protéine sur la structure des agrégats multi-protéiques	101

Introduction

“Apprivoiser” c’est créer des liens. “Apprivoiser la vie” c’est donc créer des liens entre la vie et ... Compris dans une démarche scientifique, apprivoiser la vie c’est pour nous créer des liens entre disciplines, mais aussi des liens entre niveaux d’organisation au sein du vivant. Ces deux démarches se confondent d’ailleurs puisque la distinction opérée entre les disciplines scientifiques est souvent liée aux niveaux d’organisation auxquelles elles s’intéressent. Créer du lien c’est enfin chercher à expliquer la macrostructure à partir de la microstructure, la population à partir de l’individu, la pensée à partir du neurone, ou la cellule à partir de la molécule. C’est relier ensemble, ou re-lier ensemble, dans une acception historique, des éléments que les disciplines ont plus ou moins virtuellement isolés, découpés ou découplés.

Il est en effet clair que la biologie, qui a longtemps fait un usage quasi-exclusif de la méthode réductionniste avec les succès que l’on sait, arrive aujourd’hui à un tournant de son histoire. La masse de données accumulées sur les éléments des systèmes biologiques a indirectement ouvert une boîte de Pandore : il n’est plus possible de considérer que la compréhension de l’activité d’un système biologique proviendra de cette accumulation, pour la bonne et simple raison que *l’organisation* des constituants n’est pas contenue dans la description de ces mêmes constituants, aussi détaillée soit-elle. Or, les propriétés du système, du “tout”, dépendent au moins autant de cette organisation que des caractéristiques locales de ses constituants.

Il y aurait une prétention certaine à vouloir s’approprier – qui plus est en tant qu’informaticien – une démarche qui, depuis qu’elle est devenue “systémique” ou “intégrative”, traverse tous les courants de la biologie. Cependant, créer du lien c’est d’abord créer la *possibilité* du lien, c’est-à-dire proposer les concepts, les méthodes et les outils qui vont permettre à la biologie de raccrocher les morceaux. Il nous semble en effet que la biologie ne dispose pas, au sein de son propre corpus scientifique, des moyens de répondre pleinement à cette question et qu’elle doit pour cela s’approprier des outils développés dans d’autres champs, par d’autres disciplines. C’est en ce sens que doit être comprise la démarche décrite dans ce mémoire : en tant qu’informaticien, nous disposons d’un outil capable de représenter le monde par le calcul, de le simuler et de donner à voir le simulacre. Nous nous proposons de mettre cet outil au service de la compréhension des systèmes biologiques. En d’autres termes, nous nous proposons de développer des outils de modélisation informatique pour la biologie intégrative. Il serait erroné de comprendre notre démarche comme une volonté de lier informatique et biologie. Selon nous il s’agit plutôt de relier la biologie avec d’autres disciplines, par exemple la physique, ou de relier entre eux plusieurs domaines de la biologie (biologie moléculaire et biologie cellulaire) et cela à l’aide de l’outil informatique. L’informatique est donc ici un outil de médiation. Elle incarne elle-même le lien (nous reviendrons sur ce point) car elle permet d’intégrer dans un

même modèle les caractéristiques microscopiques et les caractéristiques macroscopiques d'un système et de montrer comment les secondes peuvent émerger des premières.

Ce mémoire présente les travaux que nous avons menés, depuis près de dix ans, dans ce domaine. Cependant, après dix ans consacrés à développer des modèles informatiques de systèmes biologiques, il nous semble nécessaire de prendre du recul et d'essayer d'exprimer en quoi, mais aussi à quel prix, un modèle informatique peut contribuer au développement des connaissances en biologie. C'est pourquoi nous avons souhaité organiser ce mémoire en deux parties. Dans un premier temps, nous essayerons de formaliser la démarche de modélisation que nous avons adoptée. Cette démarche est en grande partie issue de notre expérience de la modélisation de systèmes biologiques et, plus encore peut-être, de nos interactions avec "les" biologistes (plus peut-être qu'avec "la" biologie – qui n'existerait pas sans eux). Alors qu'il est habituel de parler du "développement d'un langage commun" dans toute démarche interdisciplinaire, il nous semble pour notre part que ce langage commun n'est que la partie immergée d'un iceberg auquel nombre d'informaticiens (nous y compris) se heurtent lorsqu'ils s'approchent de la biologie. En effet, la production, par l'informatique, de connaissances en biologie passe selon nous par une démarche qui réclame un investissement beaucoup plus important que l'acquisition d'un langage. Si le modèle n'est qu'un médiateur, un outil, il n'incarne pas la connaissance produite ; il peut juste la donner à voir, lorsqu'il est "réussi". Cependant, donner à voir n'est en rien suffisant. Il faut quelqu'un pour regarder, pour interpréter et pour, *in fine*, exprimer la connaissance. En d'autres termes, une fois le lien tissé, il faut le suivre et nous verrons que, selon nous, seul le modélisateur peut le suivre correctement.

La deuxième partie de ce mémoire sera plus classique. Nous y présenterons en trois sections les trois domaines scientifiques auxquels nous nous sommes intéressés. Nous avons pour cela adopté un style synthétique en limitant *a priori* l'espace consacré à chacune de ces sections. Cela nous a conduit à extraire les informations que nous jugeons les plus pertinentes dans chacune des thématiques, mais aussi, corollaire inévitable, à éluder de nombreux points ; nous renvoyons le lecteur vers les travaux publiés pour obtenir des informations plus complètes. Les trois points abordés dans cette deuxième partie seront donc la modélisation de l'énaction (c'est-à-dire de la relation animal-environnement et du substrat neuronal qui la sous-tend), la modélisation de l'évolution de la structure des génomes et la modélisation de la structuration spatiale du noyau cellulaire. Nous leur consacrons à chacun 16 pages du chapitre II.

Chapitre I

Modélisation individu-centrée de systèmes biologiques complexes

*La poésie ça sert surtout à compliquer les choses, non ?
Mais peut-être qu'en les compliquant on les comprend mieux.
Et en les comprenant on les simplifie,
au bout du compte.*

Fred Vargas “Dans les bois éternels”
(ed. Viviane Hamy, 2006)

1 Introduction

Nous avons, dans l'introduction, partiellement clarifié le titre de ce mémoire. Le sous-titre – Modélisation Individu-Centrée de Systèmes Biologiques Complexes – semble à première vue plus explicite. Il n'est pourtant pas inutile de se pencher sur sa signification. En effet, si la phrase, dans son ensemble, ne pose pas de problème au premier abord, une étude un tant soit peu attentive révèle vite ses limites : alors qu'elle est sensée préciser le champ scientifique auquel nous nous rattachons, elle utilise une terminologie dont la signification est tout sauf consensuelle. Qu'on pose, dans un milieu scientifique quelconque, la question de ce qu'est un modèle, une modélisation, et on sera surpris de l'absence de consensus autour d'un terme qui est parfois supposé recouvrir l'ensemble de la démarche scientifique. De même, si on s'interroge sur la définition d'un système biologique complexe, il est nécessaire de produire une définition de la complexité ce qui relève de la gageure. Encore supposerons-nous ici que la définition de la biologie est relativement consensuelle . . .

Dans ce premier chapitre, nous nous proposons de prendre appui sur notre sous-titre pour décrire l'organisation conceptuelle du champ scientifique auquel nous nous rattachons. Nous procéderons pour cela en deux temps : nous tenterons tout d'abord de définir la

complexité (ou, du moins, son acception dans le cadre de ce mémoire), ce qui nous permettra de produire une définition pour les “systèmes biologiques complexes” auxquels nous nous intéressons (section 2). Le deuxième temps nous verra définir la modélisation puis la modélisation individu-centrée (section 3). Nous nous attacherons en particulier à montrer l’intérêt de cette approche de la modélisation pour l’étude des systèmes biologiques complexes, ce qui nous permettra de mieux cerner notre domaine de recherche (section 3.4).

Une fois le domaine – la modélisation individu-centrée de systèmes biologiques complexes – défini, et avant de préciser nos contributions dans le chapitre suivant, nous tenterons d’esquisser une méthodologie pour cette approche de modélisation (section 4). Cette méthodologie est issue de notre pratique : après plus de dix années consacrées à la modélisation de systèmes biologiques, après de nombreuses erreurs et quelques réussites, nous tâcherons, dans cette dernière section, de faire le point sur les causes réelles ou supposées de ces échecs et de ces succès. Nous qualifierons cette méthodologie de “pragmatique” car elle est basée sur notre expérience mais aussi parce qu’elle ne formalise pas totalement une démarche de modélisation dont nous pensons qu’elle doit être avant tout personnelle dès lors que, et nous en ferons le constat en conclusion, l’acte de modéliser est au moins aussi important que le modèle lui-même.

2 Système biologique complexe

2.1 De la complexité

La complexité est aujourd’hui omniprésente. En témoignent l’éclosion récente d’une multitude d’instituts dédiés à l’étude des systèmes complexes (par exemple à Paris, Barcelone, Lyon . . .) ou le foisonnement d’ouvrages de vulgarisation sur ce domaine (Benkirane, 2002; Delahaye, 2006; Gribbin, 2006; Zwirn, 2006). Il y aurait beaucoup à dire sur cet engouement soudain, sur ses racines plus ou moins bien assumées, sur les différents courants se réclamant des “sciences de la complexité” tout en s’ignorant magistralement, ou sur les objectifs réels des différents acteurs de ces différents courants (la complexité a le vent en poupe dans les instances scientifiques ; comme toujours dans de telles circonstances, bon nombre d’équipes, de laboratoires, voire de filières universitaires changent . . . de nom ! Il est alors souvent difficile de distinguer l’opportunisme des réelles volontés scientifiques). Notre objectif n’est pas ici de retracer l’histoire plus ou moins récente du concept mais plutôt d’identifier ce qu’il recouvre, ce qui n’est peut-être pas beaucoup plus simple . . .

La “complexité” est partout donc, mais rares sont ceux qui se risquent à une définition formelle. De fait, la complexité reste un concept particulièrement flou, qu’on cerne généralement plus qu’on ne le définit. De plus, le concept étant manipulé par nombre de disciplines différentes, les définitions peuvent être plus ou moins formelles mais aussi – et c’est généralement inversement lié – plus ou moins universelles. Finalement, la complexité est souvent définie par des expressions toutes faites aussi vagues que consensuelles (c’est complexe quand “*le tout est plus que la somme des parties*”) ou par des exemples de systèmes complexes supposés représentatifs (un embouteillage, un vol d’étourneaux, un marché financier . . .). Pourtant, lorsque la définition est un tant soit peu discutée, on aboutit souvent au constat que “tout est complexe”, ce qui est difficilement conciliable

avec une définition à base d'exemples !

En 1999, Nigel Goldenfeld et Leo P. Kadanoff donnent, dans la revue *Science*, la définition suivante : “*Complexity means that we have structure with variations*”¹ (Goldenfeld et Kadanoff, 1999). Si elle a l'avantage de la concision, une telle définition est, selon nous, l'exemple même de la définition disciplinaire d'un concept qui devrait transcender les disciplines : structure et variations s'entendent ici au sens de la physique et une telle définition n'introduit aucune distinction claire entre la complexité de la vague et la complexité de l'écrevisse, par exemple, puisque toutes deux sont des “structures avec variations”. Dans le même numéro de *Science*, Georges M. Whitesides et Rustem F. Ismagilov, après avoir fait le constat de l'ambiguïté du terme et après avoir limité leur définition à un unique champ disciplinaire (la chimie), ne donnent pas moins de trois acceptations différentes pour la complexité :

- *A complex system is one whose evolution is very sensitive to initial conditions or to small perturbations*²,
- *A complex system is one in which the number of independent interacting components is large,*
- *A complex system is one in which there are multiple pathways by which the system can evolve.*

Ce qui ne les empêche pas de conclure qu'un système est complexe si “*it is complicated by some subjective judgment and is not amenable to exact definition, analytical or otherwise*” (Whitesides et Ismagilov, 1999), définition parfaitement orthogonale aux trois précédentes !

À l'opposée du champ disciplinaire de la physique, l'informatique théorique a elle aussi proposé des définitions plus ou moins fonctionnelles de la complexité. La plus classique d'entre elles est probablement la mesure de complexité de Chaitin-Kolmogorov qui définit la complexité d'un système comme la taille du plus petit programme susceptible d'engendrer cette structure. Cependant cette définition ne recouvre que très partiellement ce que le sens commun d'une part et les sciences expérimentales d'autre part entendent par “complexité”. En effet, la complexité de Chaitin-Kolmogorov est intimement liée à la notion de hasard, et plus précisément, à la notion de suite aléatoire : une suite totalement aléatoire est ainsi l'objet le plus complexe possible puisque le plus petit programme capable de la générer est elle-même ! Pour dépasser cette contradiction, plusieurs pistes ont été proposées, plus ou moins issues de la définition de Kolmogorov (par exemple la profondeur logique de Bennett), mais aucune aujourd'hui ne permet réellement de sortir la complexité du strict champ de la théorie du calcul³.

¹Dans l'ensemble du manuscrit, nous avons pris le parti de conserver les citations anglophones dans leur langue d'origine. Dans le cas où ces citations sont extraites d'ouvrages traduits en français nous utiliserons en revanche la version traduite.

²Inversement, les systèmes complexes sont souvent supposés doués de “résilience”, c'est-à-dire de la capacité à retrouver un comportement “normal” après avoir été perturbés, ce qui est difficilement conciliable avec cette première définition.

³Une fois de plus, il n'entre pas dans le cadre de ce mémoire d'étudier de façon plus approfondie les définitions “orientées calcul” de la complexité ni de comparer leurs mérites. Pour une introduction à cette thématique, on pourra se référer à Delahaye (1999). Il nous semble cependant que la plupart de ces définitions ignorent l'existence, dans un système complexe, de deux niveaux d'organisation (micro et

Pour éviter l'écueil disciplinaire, des définitions plus larges ont été produites qui visent à une certaine universalité vis à vis des disciplines. Ainsi, le New England Complex Systems Institute (NECSI) propose la définition suivante : “*Complex systems are systems with multiple interacting components whose behaviour cannot be simply inferred from the behaviour of the components*¹”. Cette définition est difficilement contestable, si ce n'est qu'elle repose sur une opposition entre le “complexe” (ici du système) et le “simple” (ici de l'inférence du comportement du système). Dispose-t-on pour autant d'une définition du simple ? La mécanique des fluides est-elle simple ? Une vague est-elle complexe ? ... et qu'en est-il de l'écrevisse ? En outre, est-il raisonnable, scientifiquement, de définir un objet non par ses propriétés intrinsèques mais par notre capacité à le comprendre ou à en inférer le comportement ? L'exhibition d'un modèle ou d'une théorie change-t-elle alors les propriétés de l'objet de la théorie ?

Dans la pratique, il nous semble nécessaire de renoncer à une définition universelle de la complexité et d'accepter la part d'incertitude qui entoure ce concept. Pour autant, est-il raisonnable de créer des instituts, de financer des projets, de piloter la recherche, sur la base d'un concept flou et mal défini ? Clairement oui. L'histoire des sciences regorge de concepts flous dont les retombées ont été considérables. Ainsi, le concept de gène date des débuts du vingtième siècle. Alors même qu'il n'était encore qu'une construction théorique mal définie, il a conduit à la création d'une science universellement reconnue : l'essor de la génétique date du milieu des années cinquante alors que la découverte du code génétique – et donc du gène comme structure – date des années soixante (Fox Keller, 2003). Même aujourd'hui, alors que cette science est devenue indiscutable du simple fait de ses résultats et de ses retombées économiques et médicales, le gène (comme concept ou comme structure) n'est toujours pas clairement défini : à l'occasion d'un dossier publié en 2001, le magazine “La recherche” a publié le résultat d'une enquête réalisée auprès de dix-huit généticiens. Tous ont été invités à donner leur définition du “gène”. L'enquête a montré l'absence totale de consensus sur cette question puisque les dix-huit réponses étaient toutes différentes².

Dans le cadre de ce mémoire, l'absence de consensus clair sur une définition précise de la complexité et le manque de clarté des définitions consensuelles nous impose de donner notre propre définition afin de caractériser nos systèmes d'intérêt³ (sans pour autant prétendre à l'universalité ou à l'originalité de cette définition). Nous considérerons donc ici que :

Un système est un système complexe dès lors qu'il est composé d'un grand nombre d'éléments en interaction et que la dynamique de ces interactions dirige le comportement du système en lui donnant une apparence d'unité aux yeux d'un observateur extérieur.

Bien évidemment, nous ne prétendons pas que cette définition soit meilleure que les nom-

macro), l'un émergeant de l'autre tout en lui étant incommensurable.

¹<http://www.necsi.org>

²Cité par (Stewart, 2004) ; sur le même sujet, voir aussi (Gouyon *et al.*, 1997), p. 193-197.

³Une solution alternative serait de proposer – ou d'adopter – une définition disciplinaire de la complexité qui soit propre à notre domaine d'intérêt : la biologie. Malheureusement, c'est probablement dans cette discipline qu'il est le plus difficile de produire une définition claire de la complexité (Adami, 2002).

breuses définitions qui l'ont précédée ou qui la suivront, ni même qu'elle soit fondamentalement différente d'une grande partie d'entre elles. Elle présente selon nous l'intérêt d'assumer le caractère subjectif de la complexité (seule l'*apparence* d'unité est nécessaire¹) tout en définissant avec une certaine genericité la *structure* des systèmes complexes. De plus, elle place la dynamique au cœur de la complexité : selon notre définition, un système n'est un système complexe qu'à partir du moment où c'est la dynamique des interactions entre les éléments du système qui crée cette "apparence d'unité" (on pourrait aussi parler de "cohérence"). Or, qui dit dynamique dit variations mais, dans le cas d'un système complexe, l'unité ne préexiste pas aux variations ; au contraire, ici ce sont les variations qui créent l'unité. On retrouve donc là, mais en l'élargissant, l'idée de "structure avec variations" de Goldenfeld et Kadanoff (1999). En outre, cette définition évite, selon-nous, l'écueil des définitions "émergentistes" (qui définiraient un système complexe comme étant un système présentant des caractères émergents). En effet, si la définition de la complexité est quelque peu problématique, celle de l'émergence l'est encore plus. Ainsi, Edgar Morin (1977), pour ne citer que lui, définit comme "émergentes" les caractéristiques d'un système "présentant un caractère de nouveauté par rapport aux composantes de ce système considérées isolément ou agencées différemment" (p. 106), pour se retrancher immédiatement derrière le fait que "tout état global présente des émergences" (p. 107) ... On en revient au problème des définitions englobantes qui considèrent que tout système (qui est nécessairement un "état global") est un système complexe ...

Selon notre définition, l'écrevisse est bien un système complexe, comme le sont l'écosphère, l'embouteillage ou le marché financier. Par ailleurs, toujours selon cette définition, ce sont des systèmes complexes indépendamment de toute discipline ou théorie scientifique². Une montre en revanche n'est en rien un système complexe car, bien que composée d'un grand nombre d'éléments, ce n'est pas la dynamique de leurs interactions qui dirige le comportement et fait l'unité de l'objet : la montre préexiste à ses propres variations³. De même, un cristal ne sera pas considéré comme un objet complexe car ce n'est pas son *comportement* qui lui confère son caractère d'unité (sauf à considérer que la solidité est un comportement⁴). Enfin, alors que les systèmes chaotiques – la fumée – sont généralement considérés comme des systèmes complexes (voire les définitions de Whitesides et Ismagilov (1999) ci-dessus), notre définition permet de distinguer les systèmes complexes (dont

¹Nous préférons la subjectivité de "l'apparence d'unité" à celle de la "simplicité de l'inférence" (dans le définition du NECSI) ou de la "possibilité d'une description analytique" (Whitesides et Ismagilov, 1999). Elle nous semble en effet être moins soumise au progrès scientifique : quelle que soit la science qui nous permet de le décrire, un objet complexe garde sa complexité. En outre, cette formulation permet aussi de poser la question de l'observateur et de ce qu'est, pour lui, une unité (et pourquoi). Il nous semble qu'en ce sens cette définition est susceptible de réunir plusieurs approches de la complexité.

²Ce qui ne donne cependant pas à cette définition un caractère universel : l'*apparence d'unité* peut être discutée, en particulier pour des raisons culturelles, historiques ou politiques. Ainsi, l'apparence d'unité de la classe sociale, évidente dans la théorie marxiste, sera-t-elle évidemment contestée dans une perspective capitaliste. Il en ira bien évidemment différemment de l'apparence d'unité du "marché". On m'accordera la liberté de croire qu'il s'agit plus ici de rejets idéologiques que d'entorses à notre définition : les classes sociales *et* les marchés financiers sont bel et bien des systèmes complexes.

³En outre, dans le cas de la montre, la dynamique et l'unité naissent d'une volonté extrinsèque qui se matérialise dans une partie de l'objet – ici la cage.

⁴Ce qui serait d'ailleurs tout à fait recevable mais nous ne prétendons pas proposer une définition parfaitement formelle. La définition proposée, avec ses limites, a selon nous l'avantage d'être exploitable pratiquement.

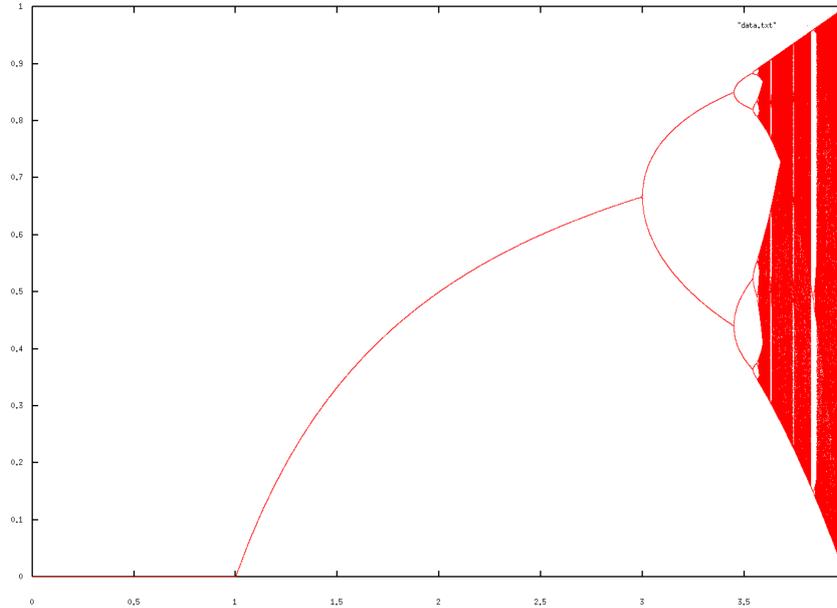


FIG. I.1 – Diagramme de bifurcations de l’équation logistique (equation I.1). Le diagramme représente les valeurs de X au cours du temps (après une phase de stabilisation) en fonction de la valeur de λ . Pour $0 < \lambda < 1$, la fonction se stabilise sur le point fixe 0; au-delà, tant que λ reste inférieur à 3, la valeur du point d’équilibre croit progressivement jusqu’à atteindre la valeur maximum de 0,66. L’augmentation de λ provoque alors un phénomène de doublement puis de quadruplement (à partir de $\lambda \simeq 3,45$) de période (la fonction logistique converge vers un cycle limite à deux puis quatre états). Enfin, pour $\lambda > 3,57$, le cycle limite est remplacé par un attracteur chaotique.

l’apparence d’unité peut être liée à des phénomènes chaotiques) des systèmes qui, pour être chaotiques, ne peuvent cependant pas être considérés comme complexes parce que ne répondant pas aux critères structuraux que nous avons énoncés. Cela n’interdit pas, bien sûr, l’utilisation de modèles non-complexes à fin de compréhension des systèmes complexes. Un exemple typique serait l’équation logistique (équation I.1).

$$\begin{cases} X(t+1) = \lambda X(t)(1 - X(t)) \\ X(0) \end{cases} \quad (\text{I.1})$$

Classiquement utilisée pour illustrer les comportements chaotiques et les bifurcations (figure I.1), elle n’est pas complexe au sens de notre définition mais elle permet de décrire le comportement du système complexe que forment un ensemble d’individus (dont le nombre au temps t est décrit par la variable $X(t)$) se reproduisant en présence de ressources limitées.

La distinction que nous venons d’établir entre une population et l’équation qui la décrit (l’une étant considérée comme complexe tandis que l’autre, qui ne l’est pas, en donne une description) nous conduit à définir, conjointement à la notion de complexité, ce que sont selon nous les “sciences de la complexité”. Une définition large en serait “sciences des systèmes complexes” mais elle se heurterait à la quasi-ubiquité des systèmes complexes qui

en ferait une “science du tout” absorbant entre autres, la biologie, les sciences humaines et une partie (au moins) de la physique ! Nous préférons donc ici une définition plus restrictive, faisant référence à notre définition d’un système complexe, en considérant que l’objectif des “sciences de la complexité” est d’expliquer comment l’*apparence de l’unité* émerge des interactions entre les entités composant un système complexe. Selon cette définition, il y a bien là un questionnement scientifique spécifique, certes lié à nombre d’autres (ainsi, dans l’exemple ci-dessus, à la théorie du chaos) mais doté d’un programme de recherche qui lui est propre.

2.2 Système biologique complexe et biologie des systèmes

Parmi l’immense variété des systèmes complexes, nous nous intéresserons ici à un ensemble de systèmes particuliers : les systèmes biologiques. Selon notre définition un système biologique complexe est donc un système vivant (ou une partie d’un système vivant), composé d’un grand nombre d’éléments en interaction, gouverné par la dynamique de ces éléments et dont le comportement donne l’*apparence* de l’unité.

Il est clair que les êtres vivants sont composés d’un grand nombre d’entités en interaction (on estime par exemple à 10^{14} le nombre de cellules composant un corps humain et à 10^{10} environ le nombre de molécules composant une cellule), qui plus est très hétérogène (ainsi, les 10^{10} fois 10^{14} molécules du corps humain appartiennent à plus de 10^6 familles différentes). Il est clair aussi que les systèmes biologiques ont l’*apparence* de l’unité. Ils peuvent donc être considérés comme autant de systèmes complexes. Il est cependant difficile d’aller au delà de telles généralités car, à part peut-être dans le cas très particulier des séquences moléculaires, on ne dispose pas d’outil permettant d’évaluer quantitativement la complexité d’un système biologique (Adami, 2002). Comment dès lors interpréter l’expression “système biologique complexe” que nous mentionnons en sous-titre de ce mémoire ? Nous avons évoqué dans la section précédente le programme de recherche des “sciences de la complexité” en le restreignant la recherche des racines de “l’apparence d’unité” des systèmes complexes. Nous opérons ici la même restriction : l’étude des systèmes biologiques complexes doit ici être comprise au sens d’étude de “ce qui fait” l’unité des systèmes biologiques ou, éventuellement, la défait, par exemple dans le cas du cancer. En d’autres termes, il s’agira de rechercher les principes structurant ces systèmes.

Historiquement, les grands courants de la science biologique ont longtemps ignoré ces questions. Jusqu’à la fin du vingtième siècle, les biologistes ont concentré leurs efforts sur la description des éléments composant les systèmes biologiques plus que sur l’explication du comportement global du système (son “apparence d’unité”) en tant que produit des interactions entre ces éléments. L’histoire de la biologie est cependant remplie de tentatives de remise en cause de cette vision réductionniste qui veut expliquer le comportement d’un système par la description de ces composantes. Au premier titre de ces tentatives, on trouvera bien entendu les approches émergentistes développées en Angleterre dans les années 1920 ou les approches cybernétiques du vivant développées après la seconde guerre mondiale (Anderl *et al.*, 2002). Aucune de ces approches n’a cependant su s’imposer.

La fin des années 90 a vu l’émergence d’une nouvelle approche de la biologie, souvent qualifiée de “biologie des systèmes” (Kitano, 2000, 2001, 2002) et, plus rarement, de “biologie

intégrative¹” (Liu, 2005). Cette approche vise à comprendre le fonctionnement des objets de la biologie (cellules, organes, organismes, populations), en tant que systèmes. Ainsi, Hiroaki Kitano, souvent considéré comme l’un des pères de la biologie des systèmes (bien que non biologiste lui-même) débute-t-il la préface de son ouvrage (Kitano, 2001) par la “définition” suivante : “*Systems biology aims at understanding biological systems at system level*”. Inversement, dans leur article séminal de 2001, Trey Ideker, Timothy Galistky et Leroy Hood définissent la biologie des systèmes dans une perspective méthodologique : elle y est présentée comme l’étude des systèmes biologiques par le biais de perturbations systématiques, d’acquisition de données à haut-débit (pour analyser les conséquences de ces perturbations) et, à partir de ces données, de l’inférence de la structure du système de régulation² (Ideker *et al.*, 2001). Cette définition “méthodologique” se retrouve partiellement dans les travaux de Palsson (2000) ou dans la définition “4M” (Measurement, Mining, Modeling and Manipulation) proposée par le groupe *Computational & Systems Biology*³ du MIT. Ces derniers ne vont pas, comme le faisait Ideker *et al.* (2001) jusqu’à restreindre la biologie des systèmes à l’étude des perturbations systématiques mais ils proposent un “cercle vertueux” similaire associant acquisition de données à haut-débit, analyse automatique et modélisation (figure I.2).

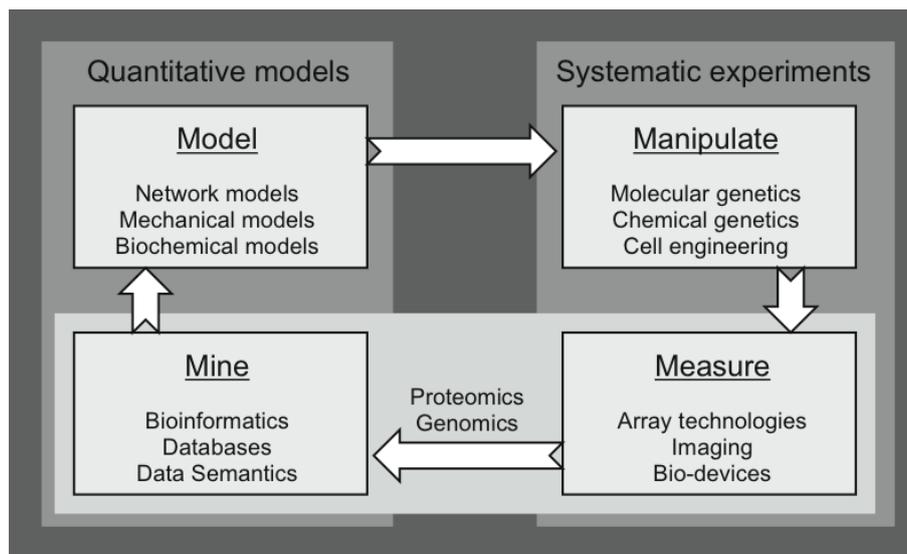


FIG. I.2 – Définition “4M” de la biologie des systèmes (d’après (Ideker *et al.*, 2006) et le *Computational & Systems Biology MIT Group*).

Les définitions méthodologiques ne définissent pas un objet d’étude tandis que les dé-

¹On notera que cette approche n’a de nouveau que son appellation et sa popularité dans la communauté des biologistes ; les fondements scientifiques eux-mêmes sont largement plus anciens et puisent leurs racines dans l’émergentisme, les sciences de la complexité, la “deuxième cybernétique” ou la biologie théorique (Bothwell, 2006). Il est possible que l’engouement actuel pour la biologie des systèmes soit plus dû au succès de l’activité de quelques lobbyistes qu’aux changements méthodologiques ou à la remise en cause soudaine du cadre réductionniste.

²On reconnaît derrière ce programme l’origine disciplinaire de la plupart des “inventeurs” de la biologie des systèmes : Trey Ideker, tout comme Hiroaki Kitano, est en effet issu d’une culture technologique. Les méthodes proposées portent la marque des techniques d’analyse issues de l’automatique.

³<http://csbi.mit.edu>

finitions de l'objet de la biologie des systèmes sont particulièrement redondantes et relèvent souvent plus de la formule incantatoire que de la définition formelle¹. En pratique, la biologie des systèmes est avant tout une recherche des principes organisationnels du vivant (Di Ventura *et al.*, 2006) ; il s'agit d'expliquer le comportement d'un système biologique complexe non plus en termes de constituants mais en termes d'agencement de constituants (Kirshner, 2005), d'où la popularité de la notion de réseaux en biologie des systèmes (Fox Keller, 2005). Or, qu'est-ce qu'un système biologique sinon un système ayant un "caractère d'unité" ? Une certaine boucle est ainsi bouclée : l'émergence simultanée d'un courant scientifique en "sciences de la complexité" et d'un autre en "biologie des systèmes" ne reflète selon nous qu'une distinction disciplinaire², les tenants des sciences de la complexité, souvent issus des disciplines "dures" de modélisation (mathématiques, physique, informatique) mettant en avant le caractère pluridisciplinaire de leur approche de la complexité tandis que les tenants de la biologie des systèmes³ cherchent à élargir le champ disciplinaire de la biologie en lui adjoignant de nouvelles méthodes, voire un nouveau "paradigme"⁴ (Aggarwal et Lee, 2003).

Le flou qui entoure la définition de la biologie des systèmes se retrouve logiquement dans les programmes méthodologiques des principaux acteurs de ce domaine. L'analyse des systèmes biologiques dans leur *globalité* peut en effet se comprendre de différentes façons. Ainsi, on distingue deux courants majeurs en biologie des systèmes : d'un côté une approche globalisante visant à intégrer l'ensemble des éléments *réels* du système tels qu'ils ont été observés, par exemple dans des analyses à haut débit ; c'est souvent à cette approche que se réfèrent les définitions méthodologiques telles que celle proposée par Ideker *et al.* (2001). De l'autre, une approche minimaliste dans laquelle on cherche, parmi les composants, le ou les éléments structuraux (idéalisés) qui conduisent à l'émergence des propriétés globales d'intérêt. Ainsi, si l'on considère le cas particulier des réseaux de régulation génétique (qui sont aujourd'hui au cœur de la biologie des systèmes), la première approche consisterait en une accumulation de données sur un organisme particulier⁵ (pour le plus grand nombre de gènes possibles et dans le plus grand nombre de situations possibles) en supposant que cette accumulation va permettre de déduire (par des approches nécessairement automatiques) les interactions réelles entre les éléments du

¹Aderem (2005) va jusqu'à conclure que la "*systems biology will fundamentally transform society*".

²À ceci près que les sciences de la complexité se sont évidemment pas limitées aux questions biologiques.

³Dont, nous l'avons vu, plusieurs sont issus de l'ingénierie, de l'informatique ou de la robotique.

⁴L'usage du terme n'est pas neutre puisqu'il invoque une "révolution scientifique" (Kuhn, 1983). Plusieurs auteurs considèrent cependant que cet usage est ici abusif (Bothwell, 2006) et que la biologie des systèmes ne constitue en rien un nouveau "paradigme", entre autres parce que les principes défendus par les tenants de la biologie des systèmes n'ont rien de "révolutionnaires" en biologie. Ainsi, Stengers et Bensaude-Vincent (2003) écrivent : "[le terme paradigme] fut très rapidement approprié par des disciplines émergentes qui ont cru voir dans le paradigme la clé du succès qu'elles ambitionnaient. Se donner un paradigme revenait alors à brandir l'étendard d'un point de vue global nouveau qui allait rassembler les chercheurs et les motiver. Manquait évidemment le trait essentiel du paradigme Kuhnien, sa capacité à engendrer des puzzles, c'est-à-dire des problèmes nouveaux, inconcevables indépendamment du paradigme, et susceptibles d'une solution."

⁵Ideker *et al.* (2001) soulignent ainsi l'intérêt des organismes "modèles" qu'ils qualifient de "Pierres de Rosette" de la biologie des systèmes. De fait, en réaction à la diversité des organismes vivants, la biologie s'est dotée de modèles "réels". Il s'agit, par exemple, de la bactérie "modèle" (*Escherichia coli*) mais on pourrait aussi bien évoquer la levure, le nématode, la drosophile, la souris ou le poulet.

réseau. La seconde approche, qui se pose plutôt en complément qu'en opposition avec la première, suppose que les réseaux de régulation présentent des régularités (structurelles, dynamiques, historiques ...) qui sont des propriétés émergentes (du fait du grand nombre d'interactions au sein du réseau) et que la compréhension de ces propriétés est indispensable à la compréhension du réseau d'une part mais aussi à son inférence à partir des données. Elle se propose donc de compléter les approches guidées par les données par des approches de modélisation *a priori* permettant de mettre en évidence ces régularités et, éventuellement, d'en proposer des explications. En ce sens, elle a des racines profondes dans les traditions philosophiques, que ce soit dans l'émergentisme anglais des années 20 ou dans le structuralisme français des années 50/60¹.

Que ce soit dans son acception "émergentiste", ou dans son acception "intégrative" (mais pas pour les mêmes raisons), la biologie des systèmes est productrice – et consommatrice – de modèles² (Palsson, 2000; Uhrmacher *et al.*, 2005). Nous proposons ici l'hypothèse selon laquelle la prise en compte des systèmes biologiques *dans leur complexité* demande que soient mis en œuvre des modèles à même d'appréhender cette complexité dans ses différentes dimensions. Sans anticiper sur la suite de ce mémoire, nous savons déjà que l'approche que nous préconiserons sera l'approche individu-centrée. Mais auparavant, effectuons un rapide retour sur ce que sont, pour le scientifique, un modèle et une modélisation.

3 De la modélisation à la modélisation individu-centrée

3.1 Qu'est-ce qu'un modèle ?

Il n'existe pas d'acception claire de la notion de modèle et une définition ne peut être donnée que dans le cadre d'un usage particulier et dans le contexte d'une discipline particulière. Ainsi, en science, la notion de modèle est fondamentalement différente dans les sciences théoriques et dans les sciences expérimentales. En effet, bien que la modélisation – en tant qu'activité de production de modèles – soit centrale dans la démarche scientifique (Nouvel, 2002), les objectifs poursuivis dépendent du rapport à l'entité modélisée : on ne modélisera évidemment pas de la même manière – ni pour les mêmes raisons – un objet accessible (directement manipulable expérimentalement), un objet inaccessible (au sens où son observation ne peut être qu'indirecte ; un atome par exemple), un concept ou un objet "à venir" (dans le cadre des sciences de la conception). Comment, dès lors, définir la notion de modèle et comment prétendre, comme le font par exemple souvent les informaticiens avec UML³, à un langage de modélisation universel !

¹“Or, par-delà les schémas d'association atomistes et celles des totalités émergentes, il existe une troisième position, qui est celle des structuralismes opératoires : c'est celle qui adopte dès le départ une attitude relationnelle, selon laquelle ce qui compte n'est ni l'élément ni un tout s'imposant comme tel sans qu'on puisse préciser comment, mais les relations entre les éléments, autrement dit les procédés ou processus de composition (selon qu'on parle d'opérations intentionnelles ou de réalités objectives), le tout n'étant que la résultante de ces relations ou compositions dont les lois sont celle du système” (Piaget, 1968).

²Pour un aperçu de la diversité des modèles développés dans le cadre de la biologie des systèmes, on se référera par exemple au cycle d'écoles thématiques organisées en France par l'atelier d'épigénomique d'Évry depuis 2002 (Amar *et al.*, 2002, 2003, 2004, 2005, 2006, 2007).

³Unified Modeling Language

Tous les auteurs et toutes les sources bibliographiques s'accordent pour attribuer à la notion de modèle une relative jeunesse (d'après Rey *et al.* (1983) le terme n'aurait pas plus de 600 ans) en même temps qu'une importante polysémie. Cependant, quelle que soit l'acception du terme¹, le modèle se trouve toujours au cœur d'un processus d'imitation : il est tantôt l'objet à imiter, tantôt l'imitation de l'objet voire – plus rarement – un intermédiaire entre l'objet et son "imitation" (ainsi, si le peintre contemporain a souvent abandonné toute volonté figurative, cela n'interdit pas l'usage du modèle comme *outil* permettant de faciliter le passage du concept à la forme²). Cette approche est d'ailleurs très proche de la vision moderne qui considère les modèles comme des médiateurs entre un objet et une théorie.

À l'instar de ce que nous avons déjà constaté pour la notion de complexité, la notion de modèle est donc, en science, très différente d'une discipline à l'autre. Une première dichotomie, majeure, s'opère entre les sciences dites de "conception" et les sciences de la nature. Pour les premières, et en particulier pour l'informatique, un modèle est avant tout *normatif* (Walliser, 1977; Simon, 2004) tandis que pour les secondes, le modèle sera *descriptif* et destiné à "expliquer" des phénomènes observés par ailleurs en en proposant une description la plus concise possible. En d'autres termes, le modèle de conception sera destiné à produire, tandis que le modèle d'observation sera destiné à *re-produire*. Au delà de cette première dichotomie, chaque discipline va proposer son acception propre de la notion de modèle, voire, pour certaines d'entre elles, poser des définitions spécifiques.

Pour l'informaticien, pour qui le modèle entre souvent dans le cadre d'un dialogue interdisciplinaire (Garbay, 2003), ces acceptions multiples sont difficilement acceptables. En effet, elles imposent de redéfinir la notion de modèle selon la discipline interlocutrice alors que l'informatique dispose de sa propre acception du terme, ce qui n'est pas sans poser quelques problèmes dans les interactions avec les tenants d'autres disciplines. Les exemples de telles "tensions" interdisciplinaires sont fréquents dans les discussions entre biologistes et informaticiens : pour les premiers, le modèle décrit – évidemment – un phénomène biologique, quelle que soit la démarche de conception qui a permis de l'obtenir (schématiquement, on pourrait dire que le modèle du biologiste correspond à un code machine réellement exécuté). Pour les seconds, le modèle a une vocation différente puisqu'il s'agit souvent du modèle sous-jacent au programme exécuté (on parle ainsi de modèle de données, de modèle de calcul, de modèle de connaissance, de modèle conceptuel, etc.). Or, le choix de ce modèle, qui est, pour les biologistes, un méta-modèle, pourra influencer les caractéristiques du programme (par exemple sa généralité, sa portabilité, son évoluti-

¹Le dictionnaire *Le Robert* donne sept acceptions différentes :

- Ce qui sert ou doit servir d'objet d'imitation pour faire ou reproduire quelque chose,
- Personne ou objet dont l'artiste reproduit l'image,
- Objet, fait, personne possédant au plus haut point certaines qualités et caractéristiques qui en font le représentant d'une catégorie,
- Catégorie, variété particulière définie par un ensemble de caractéristiques et à laquelle peuvent se rapporter des faits ou des objets réels,
- Objet, type déterminé selon lequel des objets semblables peuvent être reproduits en de multiples exemplaires,
- Objet de même forme qu'un autre objet, mais exécuté en réduction,
- Représentation simplifiée d'un processus, d'un système.

²Pablo Picasso considérait ainsi le modèle comme la réalisation d'un être imaginé au préalable.

tivité ...) mais, au moins en théorie, il ne modifiera pas le modèle biologique exécuté par ce programme.

Comme pour la complexité, nous choisirons de sortir de cette opposition entre disciplines en adoptant une définition suffisamment générale pour être englobante au vu des différentes disciplines qui nous intéressent (la conséquence étant qu'elle peut aussi bien être *rejetée* par les disciplines lorsque celles-ci adoptent un point de vue strict). Pour notre usage, nous adopterons la définition proposée par Marvin Minsky :

To an observer B, an object A is a model of an object A to the extent that B can use A* to answer questions that interest him about A (Minsky, 1995).*

Au premier abord, la définition de Marvin Minsky peut sembler si générale qu'elle en sera improductive. En pratique, elle limite cependant fortement l'étendue des modèles possibles en donnant une interprétation fonctionnelle de la notion de modèle (le modèle doit *servir* à "B" pour produire de la connaissance). Il place donc la notion de modèle au sein d'une démarche scientifique. En conséquence, la qualité d'un modèle dépendra de sa capacité à permettre la production d'une "bonne" connaissance, les questions sous-jacentes (qu'est-ce qu'une "bonne" connaissance ?) étant repoussées hors-champ, c'est-à-dire vers l'épistémologie. En d'autres termes, une telle définition fonctionnelle permet d'intégrer les différences entre les disciplines comme relevant de différences épistémologiques, toutes les disciplines n'ayant pas la même approche de la connaissance. Ici, B est supposé appartenir à une discipline particulière et c'est donc en regard de *cette* discipline que le modèle doit, en dernière instance, être jugé (et non au regard d'une discipline annexe qui aurait contribué à produire le modèle¹).

Repousser la question de la qualité du modèle dans le champ de l'épistémologie et, plus précisément, de l'épistémologie particulière d'une discipline (et donc de ses habitudes) ne suffit pas : il importe de préciser quelle est cette discipline et son épistémologie. Dans notre cas, il s'agira évidemment de la biologie et nous devrions donc nous poser la question du "bon usage" d'un modèle en biologie. Cette question est cependant beaucoup trop vaste pour être traitée dans ce mémoire. En effet, il existe un foisonnement de modèles et de méthodes de modélisation, chacune ayant sa propre histoire, ses propres méthodes, et chacune étant plus ou moins bien acceptée par la communauté des biologistes. Dans notre cas, nous limiterons notre discours à la modélisation *individu-centrée*. Cette approche de la modélisation sera présentée dans les sections suivantes. Ce n'est qu'ensuite que nous aborderons les questions de son épistémologie et que nous présenterons une méthodologie de modélisation pour cette approche particulière. Pour une présentation plus générale de la modélisation en biologie, de son historique ou de son épistémologie, nous renvoyons le lecteur à des ouvrages ou à des travaux de référence tels que (Pavé, 1994; Legay, 1997; Fox Keller, 2004; Varenne, 2004, 2007)

¹Il ne s'agit cependant pas de renier, ici, tout pouvoir aux disciplines modélisatrices : l'informatique ou les mathématiques, par exemple, ont la capacité de juger de la qualité d'un modèle (par exemple si ce modèle est mathématiquement faux ou s'il exécute un code biaisé). Ce n'est cependant pas à elles de juger de la pertinence du modèle au regard de l'objet d'étude.

3.2 Le cas de la simulation

Au sein de la ménagerie des modèles, la simulation dispose d'une place particulière, relativement ambiguë. On parle en effet de simulation dès lors que l'on "met le modèle en action", mais aussi dès lors que l'on "sauve les phénomènes¹", selon l'expression de Pierre Duhem (Duhem, 1908). Or, si la "mise en action d'un modèle" n'est que le passage de la description du modèle au calcul et à la visualisation des conséquences de cette description (on peut alors considérer, à l'instar de Parrochia (2000), que la simulation ne fait que prolonger la modélisation en permettant de réaliser des expériences sur le modèle), la "sauvegarde des phénomènes" se borne à une description phénoménologique indépendamment de toute explication scientifique (la simulation n'est alors qu'un simulacre). La science moderne a rejeté cette approche et, jusqu'à une époque récente, une "bonne" simulation ne pouvait être "que" le prolongement, par le calcul numérique, d'un modèle mathématique pré-établi. Le développement de l'informatique et, surtout, des outils graphiques a remis la question des rapports entre simulation et modélisation à l'honneur puisqu'il devient possible de "mettre le modèle en action" et de "sauver le phénomène" dans une même simulation. Le calcul numérique sur la base d'un modèle ne se contente plus de nous donner un résultat mathématique : il donne à voir une reproduction du phénomène modélisé. Dès lors, la simulation redevient ambiguë : sur quoi mettre l'accent ? sur le modèle initial ou sur l'image (ou l'animation) donnée à voir à l'issue du calcul ? Le développement de la réalité virtuelle accentue le malaise puisqu'il s'agit alors d'immerger les sens de l'observateur dans une *reproduction* de la réalité. On retrouve alors le principe d'une description phénoménologique du monde, puisque les sens peuvent être piégés indépendamment de la pertinence du modèle sous-jacent à la simulation. Le scientifique risque alors de devenir esclave de ses sens et d'accepter un modèle sur la seule foi de ce qu'il lui donne à voir.

Il serait paradoxal que l'amélioration des outils de visualisation se traduise par un affaiblissement du pouvoir de la simulation en science. En effet, s'il est certain qu'une simulation ne peut pas être, seule, un outil de démonstration scientifique, il est non moins certain que la plupart des disciplines scientifiques utilisent la simulation comme auxiliaire de la construction d'hypothèses, de la découverte, de la validation ou de la prévision. Pour éviter cet écueil, il nous semble nécessaire de conserver à la simulation graphique son statut d'*instrument* prolongeant un modèle et de veiller à évaluer avant tout le modèle (ce qui ne signifie pas qu'il ne faille pas évaluer *aussi* la simulation, comme on calibre et évalue un instrument avant de lui faire confiance²).

3.3 La modélisation individu-centrée

Parallèlement au développement de la simulation, l'informatique a ouvert la voie à une forme de modélisation nouvelle, elle-même intimement liée à la simulation. Il s'agit de la modélisation individu-centrée dans laquelle le modèle porte sur un niveau d'organisation particulier, les individus, mais où la simulation d'un ensemble de ces individus va conduire

¹Au sens où la simulation est une mémoire, une copie, du phénomène simulé.

²La validation ou, plus simplement, la comparaison, des outils de simulation constitue un champ de recherche à part entière en génie logiciel et en informatique scientifique. Nous n'aborderons pas ce thème plus longuement dans ce mémoire.

à l'émergence d'une structure globale. Parmi les différentes techniques de modélisation, la modélisation individu-centrée occupe une place à part. En effet, elle prétend proposer la conduite d'expériences à part entière et donc disposer d'un pouvoir explicatif *per se*. Dans cette section, nous présentons plus en détails ce que recouvre précisément la modélisation individu-centrée. Nous discuterons en particulier des liens qu'elle entretient avec l'Intelligence artificielle distribuée (IAD) et les Systèmes multi-agents (SMA). Nous concluons en abordant la question du statut épistémologique spécifique de la modélisation individu-centrée.

Définition

Contrairement aux définitions présentées précédemment, la définition de la modélisation "individu-centrée" semble assez consensuelle. Ce consensus fait d'ailleurs étonnamment écho à la diversité de la terminologie, les différentes disciplines utilisant ici des termes différents pour désigner une même approche de la modélisation. Ainsi, le terme "modélisation individu-centrée" est plutôt issu de la biologie ou, plus précisément, de l'écologie (Grimm et Railsback, 2005) tandis que les informaticiens, les sociologues ou les économistes préféreront souvent le terme de "modélisation agents" (Macal et North, 2005, 2006), sans d'ailleurs toujours se référer à la même définition de l'"agent". Les physiciens préféreront quant à eux le terme de "micro-simulation". Quoi qu'il en soit, tous ces termes recouvrent une approche identique de la modélisation dans laquelle la dynamique d'un système est calculée à partir de la modélisation et de la simulation des constituants de ce système. C'est pourquoi on parle souvent de modèles "bottom-up" au sens où le modèle est défini "par le bas" et que la simulation dudit modèle permet d'observer le niveau "haut". Plus formellement, nous nous référerons à la définition suivante, inspirée de celle proposée par Éric Bonabeau (2002) :

Dans la modélisation individu-centrée, un système est modélisé comme une collection d'entités autonomes, les "individus". Chaque individu dispose d'informations sur son environnement local (perceptions), ces informations sont traitées par un ensemble de règles de décision pour produire une action qui va modifier en retour l'environnement local (celui-ci pouvant être composé de ou contenir d'autres individus). Les "perceptions", "actions" et "règles de décision" peuvent être très diverses en fonction du système que les individus représentent. La répétition des interactions entre les individus est le cœur de la modélisation individu-centrée qui utilise des outils automatiques pour explorer les dynamiques de ces interactions, dynamiques hors de portée des outils mathématiques classiques¹.

Cette définition met l'accent sur plusieurs points caractéristiques de la modélisation individu-centrée. Premièrement, et c'est la caractéristique principale de cette approche, un

¹La définition originale proposée par Eric Bonabeau (2002) est "In agent-based modeling (ABM), a system is modeled as a collection of autonomous decision-making entities called agents. Each agent individually assesses its situation and makes decisions on the basis of a set of rules. Agents may execute various behaviors appropriate for the system they represent – for example, producing, consuming, or selling. Repetitive competitive interactions between agents are a feature of agent-based modeling, which relies on the power of computers to explore dynamics out of the reach of pure mathematical methods". Nous avons choisi de modifier légèrement cette définition pour lui donner un caractère plus générique quant aux systèmes modélisés.

modèle individu-centré comporte toujours au moins deux niveaux d'organisation : le local et le global. Ensuite, les caractéristiques spécifiques de ces deux niveaux sont définies : le niveau local est composé d'entités autonomes (les "individus¹") modélisées explicitement et dont les capacités de perception et d'action sont limitées par rapport au niveau global. Inversement, le niveau global n'est défini qu'implicitement puisqu'il est produit par la réunion d'un ensemble d'individus. Il est toujours possible d'énoncer des règles au niveau global (par exemple les caractéristiques de l'environnement pour la modélisation d'un écosystème). Ces règles globales ne devront cependant pas déterminer totalement le comportement des individus. En d'autres termes, le principe de la modélisation individu-centrée est de définir un système implicitement par la conjonction des interactions entre ses constituants.

La deuxième partie de la définition insiste sur le caractère dynamique des interactions locales et sur la production, par la répétition de ces interactions (dans le temps et dans l'espace), d'une dynamique globale caractéristique. Sans la nommer, c'est la notion d'émergence qui est ici évoquée : la modélisation individu-centrée suppose en effet que le collectif d'individus modélisés va faire preuve d'une dynamique spécifique, non modélisée explicitement (même si, à l'échelle individuelle, le comportement est totalement explicite). L'immersion de l'individu dans le collectif va, en effet, créer, via les interactions entre les actions des individus et leurs perceptions, une dynamique qui dépend de la dynamique individuelle mais qui la dépasse et crée la nouveauté). L'idée sous-jacente est ici la même que celle que nous avons évoquée section 2 en parlant d'"apparence d'unité" : on suppose que la dynamique du niveau global n'est pas explicable directement (à partir de la seule *description* des entités locales) mais qu'elle émerge des interactions entre ces entités.

On notera que la définition précédente définit un type de modèles mais aucunement une technique de modélisation particulière (la référence à l'ordinateur – plus exactement à des "outils automatiques" – montre la nécessité d'un moyen de calcul, elle ne nous renseigne pas sur la façon de modéliser les agents, par exemple). De fait, une grande variété de modèles différents peuvent se réclamer de la modélisation individu-centrée, d'où la variété des appellations et la difficulté de produire une définition claire de cette approche. Deffuant *et al.* (2003) cite les automates cellulaires comme "prototype du modèle individu-centré" ce qui est, selon nous, discutable². Nous proposons plutôt de considérer comme prototype les modèles de "grid-world" tels que ceux utilisés dans les simulations de DLA (Diffusion Limited Agregation, (Witten et Sander, 1983)) ou, pour des raisons au moins historiques, les modèles en essaim tels qu'ils ont été proposés par Craig Reynolds (1987). Ces modèles "prototypiques" ne doivent cependant pas masquer la très grande variété d'approches en modélisation individu-centrée (ainsi, nous proposerons dans le chapitre suivant de considérer les algorithmes évolutionnaires comme des modèles individu-centrés de l'évolution). En particulier, alors que de nombreux auteurs opposent les modèles individu-centrés aux modèles mathématiques (ou aux modèles formels), nous ne pensons pas que cette opposition soit réelle. Rien n'interdit en effet la mathématisation (ou la formalisation) d'un modèle individu-centré. En résumé, comme le précise Éric Bonabeau : "*Agent-Based Model*

¹"Individus" doit ici s'entendre au sens de "individualités" et non dans son sens anthropomorphique.

²La modélisation des entités n'est en effet pas équivalente : dans un modèle individu-centré, les entités sont modélisées explicitement et distribuées dans l'espace (c'est l'entité qui est "active") alors que dans un automate cellulaire, l'espace est actif et dense (chaque cellule est une entité). Voir à ce propos les pages de Craig Reynolds : <http://www.red3d.com/cwr/ibm.html>

is a mindset more than a technology” (Bonabeau, 2002).

Modélisation individu-centrée et systèmes multi-agents

La modélisation individu-centrée est souvent considérée comme appartenant à la “grande famille” de l’Intelligence artificielle distribuée (IAD) et des Systèmes multi-agents (SMA). Ainsi, dans l’introduction du premier chapitre du rapport de l’Observatoire Français des techniques avancées consacré aux systèmes multi-agents (OFTA, 2004), les auteurs citent la simulation de systèmes sociaux ou naturels parmi les “multiples domaines d’application des systèmes multi-agents¹”.

En pratique, même si des liens forts existent, il nous semble réducteur de considérer la modélisation individu-centrée comme une composante de l’IAD. En effet, l’objectif premier de l’Intelligence artificielle distribuée est de produire des dispositifs artéfactuels dotés de propriétés particulières (“intelligentes”). L’IAD a donc nécessairement une approche normative. Or, nous avons vu (section 3.1) qu’il était nécessaire de clairement différencier les modèles normatifs des modèles descriptifs et que la construction d’un modèle, en particulier informatique, pouvait faire appel à des *techniques* impliquant des méta-modèles spécifiques (modèles de données, modèles de calcul ...). Trop rapprocher la modélisation individu-centrée de l’IAD et des SMA risque d’entraîner des confusions entre les questions liées au modèle (biologique) et les questions liées au méta-modèle (informatique); ces confusions peuvent, par exemple, conduire à des définitions restrictives des agents/individus (Deffuant *et al.*, 2003). Or, dans une démarche de construction d’un savoir scientifique, le méta-modèle est ici un *outil*, qu’il faut connaître et dont il faut maîtriser les limites, mais dont l’impact sur le phénomène observé (donc ici sur le modèle) doit être minimisé (ou, pour le moins, connu). Dans ce contexte, un système multi-agents peut être considéré comme un outil d’instanciation d’un modèle individu-centré mais certainement pas comme le modèle individu-centré lui-même. Reste alors à déterminer si, comme le propose Alexis Drogoul, les systèmes multi-agents sont les plus adaptés à une telle instanciation (Drogoul *et al.*, 2002). En pratique, il nous semble, au moins pour l’instant, dangereux de faire reposer les modèles individu-centrés sur des implémentations “agents”. En effet, les questions clés en modélisation individu-centrées sont liées à la validation des résultats des modèles. Dès lors, l’utilisation de méthodologies ou de plate-formes “agents” pour implémenter les modèles individu-centrés, parce qu’elle a tendance à masquer les détails de l’implémentation voire certains aspects du modèle de l’individu, nous semble être une stratégie particulièrement risquée².

¹Un “agent” est ici considéré comme “une entité physique ou virtuelle (a.) qui est capable d’agir dans un environnement, (b.) qui peut communiquer directement avec d’autres agents, (c.) qui est mue par un ensemble de tendances (sous la forme d’objectifs individuels ou d’une fonction de satisfaction, voire de survie, qu’elle cherche à optimiser), (d.) qui possède des ressources propres, (e.) qui est capable de percevoir (mais de manière limitée) son environnement, (f.) qui ne dispose que d’une représentation partielle de cet environnement (et éventuellement aucune), (g.) qui possède des compétences et offre des services, (h.) qui peut éventuellement se reproduire et (i.) dont le comportement tend à satisfaire ses objectifs, en tenant compte des ressources et des compétences dont elle dispose, et en fonction de sa perception, des ses représentations et des communications qu’elle reçoit. (Ferber, 1995). Cette définition, très anthropomorphique est très limitante par rapport à la définition de l’individu comme “entité autonome modélisée explicitement”.

²En fait, il nous semble que les implémentations basées sur des plate-formes “agents” portent en elles-mêmes leurs propres contradictions. En effet, d’un côté elles sont basées sur des approches émergentes et

Motivations pour la modélisation individu-centrée

Selon (Grimm, 1999), deux motivations différentes peuvent conduire à choisir la modélisation individu-centrée : une motivation “pragmatique” et une motivation “paradigmatique”. La première motivation suppose qu’un modèle individu-centré va pouvoir rendre compte de propriétés inaccessibles aux modèles “classiques” : Bankes (2002) parle de “restrictions imposées par les modèles classiques”, restrictions auxquelles ne seraient pas soumis les modèles individu-centrés. Il serait donc, dans certains cas, simplement pragmatique d’utiliser un tel type de modèle – celui-ci n’étant alors vu que comme “un modèle de plus dans la boîte à outil du modélisateur” (Grimm, 1999). La motivation paradigmatique serait plus profonde : elle supposerait en effet que les modèles “classiques”, macroscopiques, décrivant d’emblée le niveau global de l’objet d’étude, nous donnent une description faussée de la réalité en mettant l’accent sur les caractéristiques moyennes d’un grand nombre d’individus. Ce serait donc la capacité à “capturer les phénomènes émergents” (Bonabeau, 2002) qui ferait l’intérêt de la modélisation individu-centrée.

La motivation “pragmatique” nous semble difficilement recevable puisque les modèles individus-centrés ne sont pas exclusifs des modèles “classiques” (sous-entendu, généralement, les modèles mathématiques basés sur les équations différentielles). Nous avons vu que la modélisation individu-centrée peut faire appel à des outils de modélisation différents. Rien n’interdit a priori d’utiliser une approche individu-centrée en modélisant les individus par des *outils* mathématiques puis d’intégrer (éventuellement numériquement) l’ensemble des équations pour obtenir le comportement global du système (Bonabeau, 2002).

La motivation “paradigmatique”, en revanche, semble plus claire : les modèles individu-centrés seraient à même de proposer une vision différente de la réalité, d’une part en intégrant les différences inter-individuelles (Ferber, 1995; Grimm, 1999) et, d’autre part, en proposant une approche ascendante (“bottom-up”) de la modélisation par opposition à l’approche descendante (“top-down”) qu’imposeraient les approches de modélisation macroscopiques. L’approche ascendante permettrait d’approcher les phénomènes émergents en les reproduisant là où une description macroscopique (initiant une approche descendante) serait muette car incapable de relier le comportement du tout – émergent – aux propriétés des parties.

Cependant, dès que l’on cherche à préciser ce qu’elle recouvre, la motivation “paradigmatique” se heurte à de lourdes difficultés. En effet, la capacité à prendre en compte les différences inter-individuelles ne semble pas être une motivation centrale pour bon nombre de modèles individu-centrés (nous avons cité le cas de la DLA ; les individus sont

sur l’argument selon lequel “de petites différences de comportement individuel peuvent avoir un impact dramatique sur le comportement collectif” (Bonabeau, 2002), de l’autre elles revendiquent la généralité des implémentations agents. Or, cette généralité va toujours être accompagnée d’un cadre et celui-ci sera d’autant plus dangereux qu’il sera intégré implicitement à une méthodologie ou à une plate-forme et qu’il échappera ainsi au concepteur du modèle. Ce cadre limite en effet les “différences de comportement individuel” accessibles au modèle mais, surtout, il peut donner aux individus modélisés des comportements inadaptés par rapport aux individus réels, sans que le modélisateur n’en ait jamais conscience puisque ces comportements deviennent alors implicites. Il est alors très difficile de déterminer si les comportements “émergents” observés au niveau global sont liés ou non à ces caractéristiques implicites. En résumé, l’utilisation de plate-formes génériques augmente le risque d’artefacts de simulation et diminue la capacité du modélisateur de distinguer ces artefacts des résultats du modèle.

ici tous identiques et c'est uniquement la possibilité de prendre en compte explicitement l'espace qui fait la particularité du modèle) et ne peut donc pas, selon nous, être considérée comme un avantage majeur (d'autant plus que plusieurs outils mathématiques intègrent – au moins partiellement – les différences inter-individuelles). La capacité à “capturer les phénomènes émergents”, pour reprendre l'expression d'Éric Bonabeau (2002), semble plus intéressante et fait directement écho aux sciences de la complexité. Mais l'“émergence” reste un concept très mal défini et la plupart des auteurs utilisant cet argument en faveur des modèles individu-centrés se contentent, pour étayer leur argumentation, de proposer des exemples de phénomènes émergents (on retrouve ici un des travers que nous avons déjà signalé pour la définition de la complexité). En d'autres termes, l'émergence est un concept invoqué pour argumenter en faveur des modèles individu-centrés sur la base d'exemples ad-hoc (c'est-à-dire sur la base d'exemples où ces modèles ont effectivement été des réussites), ce qui est évidemment contestable. En outre, l'émergence est, au moins dans le cas général, non-mesurable et le plus souvent les modèles ne sont validés que par des sorties graphiques analysées “à l'œil” (Bankes, 2002). Même si, indubitablement, les modèles individu-centrés ont “à voir” avec l'émergence, il reste difficile de se baser sur ce seul concept pour les défendre !

En fait, selon nous, l'intérêt des modèles individu-centrés est à rechercher ailleurs. Ces modèles ont en effet trois particularités que ne possèdent pas les modèles macroscopiques : ils sont relativement simples à manipuler, ils permettent de modéliser l'“inconnu” et ils utilisent directement l'ontologie du domaine :

Des modèles simples à manipuler. Une fois les principes généraux admis, la construction d'un modèle individu-centré est relativement aisée (au moins pour quiconque maîtrise suffisamment un langage de programmation). En effet, l'approche individu-centrée permet de réaliser des “modèles minimaux” très simples, mais présentant “déjà” des comportements remarquables et qui peuvent servir de point de départ à des programmes de recherche plus ambitieux. Les exemples de tels modèles minimaux sont extrêmement nombreux¹. Par rapport aux difficultés qu'il y a à maîtriser les modèles mathématiques, il peut donc sembler simple de s'improviser “modélisateur individu-centré”. Cette relative simplicité ne doit cependant pas masquer la difficulté de construire un discours scientifique recevable sur la base d'un modèle individu-centré (voir ci-dessous et section 3.3).

Des modèles permettant de modéliser l'“inconnu”. La modélisation d'un objet demande une connaissance explicite de cet objet (ainsi, en climatologie, la modélisation de l'atmosphère demande une connaissance des règles de déplacement des masses d'air à l'échelle continentale). Dans le cas de la modélisation individu-centrée, cette contrainte peut être partiellement relâchée puisque l'échelle de description du modèle est différente de son échelle d'observation. Il devient alors possible de modéliser des phénomènes “inconnus” (au sens d'inexpliqués) lorsque ceux-ci mettent en jeu des éléments connus, modélisables explicitement sous la forme d'individus. La modélisation change alors de statut puisque de technique de validation d'hypothèses elle devient force de proposition d'hypothèses.

¹On pourra se référer aux bibliothèques de “curricular models” du logiciel “NetLogo” (Resnick, 1995) pour avoir un aperçu de l'étendue des modèles simples accessibles via la modélisation individu-centrée (<http://ccl.northwestern.edu/netlogo/>).

Des modèles utilisant l'ontologie du domaine. Un des principaux avantages (bien qu'indirect) des modèles individu-centrés est qu'ils permettent une communication directe entre le modèle et le spécialiste du domaine (Garbay, 2003). En effet, contrairement aux modèles mathématiques qui vont généralement faire appel, dans leur expression comme dans leurs résultats, à une terminologie propre au modélisateur, les modèles individu-centrés utilisent directement une description comportementale des agents (donc exprimable directement dans l'ontologie du domaine) et leurs résultats sont directement "présentables" puisqu'il s'agit soit d'images (fixes ou animées) représentant le comportement de la population, soit d'indicateurs mesurés sur le système au cours de son évolution. En outre, ces indicateurs sont potentiellement identiques à ceux utilisés dans la discipline "cible". Parce que son ontologie est celle de l'objet modélisé et non celle de la technique de modélisation (ou de la technique d'analyse du modèle), la modélisation individu-centrée simplifie l'analyse du comportement du modèle.

Paradoxalement, ce sont les avantages des modèles individu-centrés qui constituent (au moins historiquement) leur principale faiblesse. En effet, leur simplicité apparente de conception, d'analyse et de manipulation, leur capacité à présenter des résultats graphiques directement comparables aux observations réelles, ont pu conduire à une certaine précipitation dans la construction de modèles individu-centrés. On a ainsi pu assister à un foisonnement de tentatives de modélisation sans que celles-ci ne soient intégrées à un réel questionnement scientifique. De même, la proximité ontologique entre le modèle et l'objet modélisé a pu être source de confusion : en effet, l'utilisation d'une ontologie commune (mais tacite) ne suffit pas puisqu'elle ne garantit en rien que les interprétations de cette ontologie seront réellement partagées. En cas de divergence, les risques d'interprétations abusives sont nombreux puisque le modèle et l'objet modélisé n'ont plus, alors, que des rapports de ressemblance, d'analogie ou de métaphore (Fox-Keller, 1999; Nouvel, 2002; Varenne, 2006). En fait, les problèmes méthodologiques et épistémologiques posés par la modélisation individu-centrée sont nombreux. Or, nous l'avons dit, un modèle n'a de sens que s'il permet de répondre à une question ; il ne suffit pas de produire une image "ressemblante", encore faut-il que cette image puisse être intégrée à un corpus scientifique, y faire sens et répondre à des questions d'intérêt *au sein de ce corpus*.

Plusieurs auteurs ont proposé de résoudre ces problèmes méthodologiques en formalisant la démarche de construction du modèle et/ou en développant des plate-formes de modélisation individu-centrées "génériques", celles-ci étant supposées permettre une amélioration de la qualité des modèles (Michel, 2004). Selon nous, ces démarches sont illusoire car elles oublient qu'un modèle est avant tout un modèle *de quelque-chose*. Ce n'est donc pas l'implémentation dudit modèle dont il faut améliorer la qualité (même si une telle amélioration ne peut pas nuire) mais sa relation à ce "quelque-chose". Or, cette relation passe par la médiation du scientifique car non seulement le modèle est un "modèle de quelque-chose" mais il est aussi modèle de quelque-chose *pour quelqu'un* (voir section 3.1). La formalisation de la démarche de modélisation, parce qu'elle prive selon nous le modélisateur d'un contact "charnel" avec son modèle, casse cette relation tripartite entre l'objet A , le modèle A^* et le modélisateur B . Or, selon nous, c'est uniquement dans le cadre de cette relation qu'une connaissance peut être produite.

Épistémologie de la modélisation individu-centrée

Ce qui est simple est toujours faux. Ce qui ne l'est pas est inutilisable.

Paul Valéry “mauvaises pensées et autres”,
(Œuvres, Tome II, Gallimard, Bibliothèque de la Pléiade, 1960, p. 864)

En première approximation, la question de l'épistémologie de la modélisation individu-centrée pourrait se résumer à la question suivante : “Est-ce qu'un modèle individu-centré peut constituer une preuve scientifique” ? Selon nous, il faut impérativement se garder de simplifier ainsi la question sous peine d'y apporter une réponse négative et, partant, de rejeter en bloc l'outil au nom de son incapacité à prouver. Il nous semble en effet évident – et à de multiples titres – qu'un modèle individu-centré ne constituera jamais une preuve scientifique. En premier lieu, bien sûr, il s'agit d'un modèle et à ce titre d'une description approximative, nécessairement fautive, de la réalité (de plus, il s'agit aussi d'un objet artificiel toujours susceptible d'être soumis au bon vouloir de son concepteur). Il ne peut donc pas constituer une preuve du phénomène modélisé (Gruene-Yanoff, 2006). Cependant, au delà des limites de l'outil “modèle”, c'est surtout le sort que l'épistémologie, de Popper à Feyerabend en passant par Kuhn ou Lakatos, a fait subir à la notion de preuve scientifique qui interdit de répondre par l'affirmative : un modèle individu-centré ne peut pas être une preuve scientifique tout simplement parce qu'il n'existe pas de preuve scientifique absolue¹. Alors que la première objection tendrait à rejeter l'usage de la modélisation individu-centrée au profit d'une pratique scientifique expérimentale directe², la deuxième nous conduit au contraire à considérer les deux pratiques d'une façon équilibrée à la manière du “tout est bon” de Paul Feyerabend (Feyerabend, 1975).

La question de l'épistémologie de la modélisation individu-centrée ne peut donc pas se résumer à son éventuel statut de preuve scientifique. Il est nécessaire d'élargir notre pro-

¹Nous n'entrerons évidemment pas ici dans une discussion approfondie sur ce point – ce n'est ni possible dans l'espace de ces lignes, ni de notre domaine de compétence. Nous renvoyons le lecteur, par exemple, à Chalmers (1998) pour une présentation des principaux auteurs cités ci-dessus. Pour une présentation de l'“explication scientifique” (et en particulier des différentes acceptions en fonction des disciplines), on se reportera à (Vergnioux, 2003). Sur la question de l'impossibilité de la preuve absolue, on citera Stengers et Bensaude-Vincent (2003) (p. 305-306) : “L'idée que la preuve est le régime “normal” en science est donc néfaste. D'une part, elle méconnaît la variété des pratiques d'argumentation en sciences, d'autre part, elle conduit à tenir pour preuve des arguments statistiques opaques, à justifier au nom de la preuve des montages pseudo-expérimentaux qui soumettent unilatéralement et sans risque leur objet d'étude”.

²Le débat sur la pertinence de la modélisation par rapport à l'expérimentation mérite cependant d'être nuancé. Certes, les modèles individu-centrés (et les modèles “tout court”) ne contribuent pas directement à la preuve scientifique, certes, ils mettent en scène des situations fictives, mais sont-ils en cela fondamentalement différents des expériences “de laboratoire” ? À côté des expériences nécessairement idéalisées dans lesquels un objet réel est arraché à son environnement pour être “soumis à la question” (Stengers et Bensaude-Vincent, 2003), la modélisation individu-centrée apparaît comme une expérience idéalisée sur le plan de l'objet (remplacé par son modèle : l'individu) mais dans laquelle le modèle de l'objet est placé au sein d'une collectivité “réelle”. Nous n'irons pas ici jusqu'à défendre, à l'instar de Feyerabend, une position d'équivalence entre les différentes pratiques (“tout est bon” ne signifie pas pour nous “tout se vaut”) et il nous semble évident qu'une expérience *réelle* dans un environnement idéalisé aura toujours plus de pouvoir explicatif (au moins aux yeux de la communauté scientifique) qu'une expérience de simulation, fut-elle individu-centrée. Néanmoins, cette position extrême présente l'intérêt de montrer le caractère subjectif de *toute* expérience et de relativiser le caractère fictif du modèle.

blématique pour aborder la question de l'intégration de la modélisation individu-centrée dans une *pratique* scientifique valide. En d'autres termes, si les modèles individu-centrés ne sont pas porteurs d'un pouvoir explicatif en eux-mêmes, peuvent-ils néanmoins être utilisés dans le cadre d'une procédure scientifique et contribuer à l'avancée de celle-ci ? À cette deuxième question, notre réponse est clairement affirmative. En effet, les modèles individu-centrés ont la particularité d'être articulés sur deux niveaux : un niveau descriptif (l'individu) et un niveau d'observation (la population). À ce titre, ils peuvent *surprendre* leur concepteur¹ et lui poser des questions . . . auxquelles celui-ci peut répondre scientifiquement. Cette particularité des modèles individu-centrés permet de réaliser, sur le modèle, des expériences concrètes, bien que virtuelles. Franck Varenne (2003) parle ainsi d'"expériences concrètes du second genre", par rapport aux expériences concrètes du premier genre qui se déroulent, elles, dans le monde réel. Ces expériences concrètes permettent donc de "faire de la science" avec les modèles individu-centrés, c'est-à-dire d'aller au delà du simple rôle heuristique (donner des idées) pour permettre de déployer une réelle méthodologie scientifique basée sur l'observation de la nouveauté, la production d'hypothèses et la validation de ces hypothèses par l'expérimentation (ici par l'expérimentation sur le modèle). Il reste cependant, un problème majeur : nous venons d'opérer un glissement sémantique dangereux similaire à celui dans lequel un grand nombre de modèles individu-centrés se laissent entraîner. En effet, à une question portant sur un objet réel *modélisé*, nous venons de substituer une question portant sur le modèle lui-même. Nous avons donc remplacé l'objet réel A par un autre A^* qui n'était, dans le questionnement initial, qu'une construction théorique (Varenne, 2001). Cela nous ramène à notre première objection : la réponse à la question posée au modèle n'implique (au sens logique) *jamais* la réponse à la question posée à l'objet (fut-elle la même question).

Pour transposer la réponse donnée par le modèle vers l'objet lui-même, il faut nécessairement passer par le filtre de l'analyse. C'est en effet l'analyse qui va permettre cette transposition et ce, même si (surtout si), comme nous l'avons précisé ci-dessus, la modélisation individu-centrée *semble* pouvoir se passer de cette analyse en vertu de la proximité ontologique qu'entretient (ou que semble entretenir) le modèle avec l'objet modélisé. En résumé, et en faisant référence à la définition que nous avons adoptée pour le "modèle", ce n'est en aucun cas à ce dernier de répondre à une question scientifique ; la question doit rester la propriété de l'observateur B et c'est à lui d'analyser le comportement du modèle A^* et d'en *déduire* (ou non) des conclusions sur l'objet A . Dès lors, la contribution du modèle à la progression d'un questionnement scientifique doit être mesurée à l'aune de la capacité du modélisateur à transposer ce qu'il a pu observer sur le modèle vers le domaine scientifique d'origine, c'est-à-dire celui de l'objet A . Cela ne signifie bien évidemment pas que tous les modèles se valent et que seul compte le modélisateur mais plutôt que la qualité d'un modèle ne peut être comprise que relativement à la capacité du modélisateur à exploiter ce modèle. La qualité d'un modèle n'est donc pas – ou pas exclusivement – intrinsèque au modèle, ni même au couple objet/modèle². Nous rejoignons

¹Cette capacité à surprendre a d'ailleurs été proposée par Ronald et Sipper (2001) comme critère caractéristique de l'émergence.

²Voir la position que nous avons déjà exprimée à propos de la simulation (section 3.2) : ce n'est pas la capacité à reproduire une dynamique particulière qui fait la qualité d'un modèle, même si elle ne nuit pas. On notera cependant que la validation d'un modèle ou d'une simulation doit être accordée avec l'usage qui en sera fait. La position que nous défendons ici est adaptée aux modèles scientifiques. Le cas des

en cela la position formulée par Grimm (1999) qui affirme que : “*The decisive thing with modelling is not the model per se, but what the model and working with the model **does to our mind***”¹. (c’est nous qui soulignons). Ce dernier va d’ailleurs plus loin puisqu’il affirme que : “*It could be argued that a criterion to determine good models is that they are no longer needed afterwards*”.

De la conclusion à laquelle nous venons d’aboutir dérivent des conséquences pratiques importantes. En effet, si le concept de preuve doit être considéré avec défiance, il n’en reste pas moins qu’une contribution scientifique, quelle qu’elle soit, est considérée comme “prouvée” lorsque se dégage un accord collectif de la communauté scientifique concernée par cette contribution (Stengers et Bensaude-Vincent, 2003). Dès lors, si la qualité d’un modèle A^* se mesure à la capacité du modélisateur B de transposer les résultats obtenus sur le modèle vers le domaine scientifique propre à l’objet A , cette transposition implique, de la part du modélisateur, qu’il s’adresse effectivement à cette communauté et qu’il en adopte les règles méthodologiques et épistémologiques (sous peine d’être rejeté par elle). Dans le cas qui nous intéresse, la biologie, cela signifie que les résultats du modèle – ou les conséquences que le modélisateur tire de ce modèle – doivent être communiqués aux biologistes. Cela impose en particulier que soient adoptées les règles de communication propres à ce domaine scientifique². En outre, ces résultats devront être confrontés au corpus de connaissance de ce domaine, ce qui implique une discussion – donc une bibliographie – thématiquement orientée vers le domaine d’intérêt et non vers une hypothétique qualité intrinsèque du modèle.

Quelle est, dans ce schéma, la place du modèle ? Selon nous, elle est – au moins – double : c’est le modèle qui a stimulé l’intuition du modélisateur, à travers les “expériences” qu’il a pu conduire dessus (même si c’est aussi l’intuition du modélisateur qui lui a dicté le modèle). Mais le modèle (ou un modèle dérivé) peut aussi être utilisé pour communiquer cette intuition à la communauté. Cette deuxième utilisation du modèle doit cependant être prudente : en présentant son modèle à la communauté, le modélisateur court le risque de se voir reprocher des faiblesses méthodologiques. Il doit en effet alors avouer la part intuitive de son travail, part qui, si elle fait partie de toute construction scientifique, est généralement gommée a posteriori pour ne conserver que ce qui est admis par la “méthode scientifique” (Stengers et Bensaude-Vincent, 2003). Selon nous, cette obligation à reconnaître une démarche scientifique sinueuse, brouillonne, voire opportuniste a au moins autant contribué aux difficultés de reconnaissance de la modélisation individu-centrée (ou, préalablement, de la vie artificielle) en biologie que les faiblesses méthodologiques évoquées ci-dessus³. À tel point que les seuls modèles reconnus par la communauté sont ceux aux-

modèles prédictifs destinés à l’aide à la décision serait bien évidemment différent.

¹Le point de vue de Grimm (1999) est ici très proche de ce qui avait été énoncé par Legay et Lebreton (1973) puis Le Moigne (1977), à savoir que le premier critère de qualité d’un modèle est son “utilité” (pour J.-M. Legay) ou sa “pertinence” (pour J.-L. Le Moigne).

²Il n’est pas “suffisant” de communiquer les résultats à une audience de modélisateurs (puisque ceux-ci ne sont pas à même de juger de la qualité du modèle). De fait, “il y a peu de controverses entre simulateurs car il y a peu de travail collectif. Les simulateurs sont rassemblés par l’équipement informatique qui leur est nécessaire, mais ils fonctionnent plutôt à la manière de petits artisans : chacun son problème, son modèle, son programme” (Stengers et Bensaude-Vincent, 2003).

³On notera que la modélisation mathématique est moins sujette aux reproches sur ce point, probablement parce qu’elle entre plus clairement dans le cadre méthodologique tacitement admis : le modèle

quels se sont associés des biologistes suffisamment renommés pour ne pas être suspectés de déroger à la sacro-sainte méthodologie expérimentale.

Nous n'avons fait qu'effleurer la question de l'épistémologie des modèles individu-centrés (comment aurions-nous pu faire plus?). Cependant, cette discussion rapide nous apporte déjà trois conséquences majeures : la première est que parler d'une épistémologie *des* modèles individu-centrés n'a probablement pas de sens : chaque champ d'application va en effet apporter avec lui sa propre épistémologie à laquelle le modèle sera prié de se plier (Varenne, 2001). La deuxième est la reconnaissance que le processus de modélisation ne se limite pas à la construction d'un modèle : il commence bien en amont par la spécification d'une question ; il prolonge la construction du modèle par une série d'expérimentations sur ce modèle (l'objectif de ces expérimentations n'étant pas de "valider" le modèle mais plutôt de stimuler l'intuition du modélisateur) et se termine bien en aval par la reconnaissance, par la communauté, de la réponse à cette question. Cette deuxième conséquence nous amènera à proposer (section 4) une méthodologie pragmatique pour la modélisation individu-centrée en biologie, méthodologie qui prendra la forme d'un cycle de vie tel que celui que nous venons d'ébaucher. La troisième conséquence est l'affirmation de la part fondamentale de la pluridisciplinarité dans le processus de modélisation. Outre que c'est à la communauté étudiant l'objet A qu'il faudra s'adresser pour valider les avancées autorisées par le modèle, nous avons vu que c'est à l'intuition du modélisateur que "parle" le modèle. Or, si la nécessité de restituer les résultats implique que le modélisateur se *fonde* dans une communauté (pour nous de biologistes), l'appel à l'intuition implique aussi qu'il se *fasse* biologiste¹ : pour qu'un modèle puisse le surprendre, le modélisateur doit en effet être en mesure d'être surpris puisque "*le hasard ne favorise qu'un esprit préparé*" (Louis Pasteur).

3.4 Conclusion : modèles individu-centrés de systèmes biologiques complexes

Nous l'avons vu dans la section 2, la biologie des systèmes est gourmande de modèles. Ceux-ci peuvent cependant être de natures très différentes, depuis les modèles mathématiques classiques utilisés, par exemple, pour modéliser les voies métaboliques dans la cellule (Loew et Schaff, 2001), jusqu'aux modèles formels inspirés de la théorie du calcul. Ainsi, le π -calcul a été utilisé pour modéliser les interactions concurrentes au cours de la transcription de l'ADN (Kuttler, 2007). Pour un inventaire des modèles utilisés en biologie des systèmes, on se référera par exemple à Uhrmacher *et al.* (2005). Dans cette "ménage-

mathématique est construit en fonction de l'hypothèse (il permet de la vérifier, de la valider ou de la préciser, mais pas de la dépasser) alors que le modèle individu-centré est susceptible de proposer de nouvelles hypothèses. Là où le modèle mathématique *confirme*, le modèle individu-centré *étonne*.

¹Il n'y a pas là contradiction avec la nécessité de travailler en étroite collaboration avec des biologistes. Une telle collaboration n'est cependant pas suffisante car la position du biologiste étant centrée sur A et non sur A^* , il ne sera pas surpris par le modèle : le biologiste *attend* (s'attend à) la ressemblance de A^* avec A mais il ne peut pas interpréter cette ressemblance puisqu'il n'a pas lui-même construit le modèle (le cas idéal d'une interaction parfaite entre un biologiste et un modélisateur, chacun restant strictement dans son rôle, est, d'après notre propre expérience, illusoire). Cela n'interdit en rien à un biologiste de suivre le chemin inverse de celui que nous énonçons – et donc se se faire modélisateur – mais le regard que nous portons ici est celui issu de notre propre parcours.

rie” de formalismes, les modèles individu-centrés sont, semble-t-il, légèrement en retrait. Ainsi, ils ne sont même pas mentionnés dans la liste établie par Ideker et Lauffenburger (2003). Pourtant, il nous semble qu’ils disposent d’atouts importants pour l’étude des systèmes biologiques complexes. En effet, la question centrale qui anime la communauté est la compréhension de ce qui fait d’un système biologique une totalité organisée (Di Ventura *et al.*, 2006). Or, il est probable que les principes organisationnels à l’œuvre dans les systèmes biologiques soient encore très mal compris, voire, pour certains d’entre eux, encore à découvrir¹. En d’autres termes, la biologie des systèmes complexes doit encore *explorer* pour dégager ou préciser ces principes organisationnels. Parce qu’ils intègrent les principes organisationnels *a priori*, les modèles dits “classiques” sont mal adaptés à une telle exploration. En revanche, parce qu’ils parlent directement à l’intuition du scientifique et parce que les principes organisationnels y sont observés et non prédéfinis, les modèles individu-centrés constituent une alternative prometteuse. De plus, ils permettent de dépasser les contraintes d’échelles : il est possible de faire communiquer au sein d’un même modèle des échelles temporelles ou spatiales différentes, voire des schèmes explicatifs différents. Or, nous avons vu que de telles interactions constituent une des caractéristiques importantes des systèmes biologiques complexes (en d’autres termes, dans de tels systèmes, les “micro-émergences” sont au moins aussi importantes que les “émergences²”). En outre, Morange (2005) montre la nécessité de concilier des schèmes explicatifs différents pour parvenir à dégager les principes organisationnels des systèmes vivants.

Parallèlement à l’étude des mécanismes structurant les systèmes biologiques, l’approche individu-centrée permet d’étudier, par la modélisation, des structures difficilement observables *in vivo*. En effet, même si la biologie dispose de moyens toujours plus puissants pour observer les systèmes qui l’intéressent, de nombreuses échelles d’observation lui sont encore inaccessibles. Il peut ainsi s’agir d’échelles spatiales (par exemple l’échelle méso-scopique pour la cellule) ou temporelles (par exemple l’échelle évolutive). Souvent, c’est l’observation *couplée* de différentes échelles qui est impossible (par exemple la dynamique du système). Dans de tels cas, les modèles individu-centrés permettent de réaliser des expériences, certes totalement artificielles, mais impossibles à réaliser en pratique sur les systèmes biologiques réels (Cohen, 2004). Même s’il ne s’agit pas là, selon nous, d’un atout majeur (car les progrès des techniques expérimentales en biologie sont si rapides que ces “tâches aveugles” ont de fortes chances de ne pas le rester très longtemps), les modèles

¹Pour ne citer que les thèmes que nous aborderons dans les trois prochains chapitres, l’équilibre robustesse-variation est un principe organisationnel encore très mal connu en biologie évolutive (Lenski *et al.*, 2006) ; le couplage entre l’organisation spatiale de la cellule et son métabolisme est une des clés de la compréhension de la cellule – mais très peu de modèles permettent d’aborder ces questions (Bork et Serrano, 2005) ; la circularité par laquelle le comportement d’un organisme détermine ses percepts, donc son apprentissage, donc son comportement reste encore largement méconnue (Di Paolo et Iizuka, 2008). Ces trois principes organisationnels, concernant trois systèmes biologiques différents, demandent, pour être compris, que soient prises en compte simultanément les structures, la dynamique et les fonctions des systèmes biologiques, ce que peu de formalismes de modélisation permettent.

²Si une “émergence” est définie par Morin (1977) (p. 106) comme une caractéristique d’un système “présentant un caractère de nouveauté par rapport aux composantes de ce système considérées isolément ou agencées différemment”, une “micro-émergence” est une “qualité inhérente aux parties d’un système donné [qui est] absente ou virtuelle quand cette partie est à l’état isolé ; elle ne peut être acquise et développée que par et dans le tout” (d’après Morin (1977), p. 108). D’autres auteurs ont proposé des terminologies différentes pour qualifier cette rétroaction du global vers le local (on trouvera ainsi souvent le terme “immergence” ou, plus occasionnellement “émergence locale”).

individu-centrés peuvent permettre de “préparer l’intuition” des biologistes en explorant *in virtuo* ces échelles inaccessibles. Nous avons évoqué dans la section précédente la nécessité, pour le modélisateur, d’être “préparé” à l’observation de son modèle ; il s’agit ici de renverser cette perspective en utilisant le modèle pour préparer l’expérimentateur aux “surprises” que lui procureront ses expériences, voire pour lui permettre de choisir les expériences à réaliser.

Si les modèles individu-centrés semblent à même d’apporter leur pierre à l’étude des systèmes biologiques complexes, cela ne doit pas nous faire oublier la conclusion à laquelle nous a amené la section précédente, à savoir que le modèle individu-centré ne répond jamais *seul*. Nous avons en particulier insisté sur la pluridisciplinarité que l’utilisation des modèles individu-centrés implique et sur la nécessité d’une méthodologie adaptée à leur utilisation. Dans la dernière partie de ce premier chapitre, nous tâcherons d’ébaucher une telle méthodologie en nous appuyant sur les intérêts et sur les limites de la modélisation individu-centrée, tels que nous les avons évoqués dans les sections précédentes.

4 Une méthodologie pragmatique pour la modélisation individu-centrée en biologie

Une autre erreur c’est la réduction péremptoire de la connaissance aux arts et méthodes, réduction qui améliore rarement les sciences ; car de même que les jeunes gens grandissent rarement lorsque leur corps et leurs membres sont pleinement formés, ainsi la connaissance, tant qu’elle constitue en aphorismes et en observations, reste en état de croissance ; mais une fois qu’elle est façonnée par des méthodes, bien qu’elle puisse encore être améliorée, illustrée et mise en état de servir, elle n’augmente plus en volume et en subsistance.

Francis Bacon¹.

4.1 Introduction

Nous avons fortement insisté sur la nécessité d’une méthodologie pour la modélisation individu-centrée. Il pourrait nous être objecté que de telles méthodologies existent déjà dans une discipline “sœur”, à savoir les systèmes multi-agents (par exemple, les approches “Voyelles” (Demazeau, 1995), “Belief-Desire-Intention” (Muller *et al.*, 1997), “Cassiopée” (Collinot et Drogoul, 1998) ou “Agents-Groupes-Rôles” (Ferber et Gutknecht, 1998)). Selon nous, ces méthodes de conception ne prennent pas suffisamment en compte l’interaction entre le modélisateur et le modèle (qui plus est, elles proposent souvent une vision très anthropomorphique des “agents”). Elles sont en effet adaptées à la conception de modèles normatifs (voir section 3.1) quand nous cherchons une méthodologie adaptée à la conception de modèles descriptifs. En d’autres termes, il s’agit de proposer une méthodologie permettant la production de connaissances, par le modélisateur, sur un objet

¹cité dans (Feyerabend, 1975), p. 171.

biologique réel et non la production d'un artefact, quand bien même celui-ci présenterait un comportement similaire à un objet biologique. Ainsi, aucune des méthodologies de conception "agents" n'évoque les liens que le modèle – et le modélisateur – doivent tisser avec le domaine d'origine de l'objet modélisé alors que ces liens sont au cœur de la démarche de production de connaissances¹.

Dès lors qu'on souhaite proposer une méthodologie de modélisation, plusieurs approches sont possibles : on peut proposer des règles générales auxquelles le modélisateur devra se conformer ou, au contraire, décrire l'ensemble des étapes qu'il devra strictement suivre pour parvenir à ses fins (la méthodologie étant alors souvent décrite sous la forme d'un diagramme d'états). Dans la perspective qui nous intéresse ici, nous devons garder à l'esprit que la méthodologie proposée doit permettre la co-construction du modèle et de la connaissance. Cela signifie que la méthodologie proposée ne pourra pas être trop stricte : chaque modèle – mais aussi chaque modélisateur – va devoir développer ses propres spécificités en vue de cette co-construction. En d'autres termes, la méthodologie proposée devra fixer un cadre et permettre, à l'intérieur de ce cadre, une liberté d'action en fonction de l'intuition du modélisateur (et donc de l'impact que le modèle a déjà pu avoir sur cette intuition). Nous avons déjà évoqué le "tout est bon" qui résume l'opposition de Feyerabend (1975) à une quelconque méthode scientifique. L'idée est ici, à nouveau, que si "tout est bon", "tout ne se vaut pas" dans le cadre d'une pratique scientifique ici, maintenant, dans *cette* communauté.

Nous avons choisi de présenter ce cadre méthodologique en trois points : dans un premier temps, nous reviendrons sur le caractère *experimental* de la modélisation individu-centrée. Cela nous permettra, dans un deuxième temps, de proposer des règles générales auxquelles le modélisateur devra se conformer afin de produire une connaissance *admissible* par une communauté scientifique. Ces règles seront ensuite déclinées en un "cycle de vie d'un modèle individu-centré" qui constituera la forme la plus contraignante de la méthodologie proposée.

4.2 La modélisation individu-centrée, une modélisation expérimentale

Nous avons évoqué, au cours des sections précédentes, les particularités de la modélisation individu-centrée et sa capacité à *surprendre* le modélisateur. Nous avons adopté une position proche de celle énoncée par Varenne (2003), à savoir que la modélisation individu-centrée, de par sa capacité à réaliser des expériences concrètes "du second genre", dépasse le simple statut de représentation de la réalité pour acquérir celui d'objet d'étude à part entière.

C'est cette caractéristique spécifique de la modélisation individu-centrée qui constituera le socle de notre méthodologie. En effet, si on peut (i) être surpris par le modèle qu'on a soi-même construit et (ii) conduire des expériences sur ce même modèle, alors il devient

¹En fait, une classification des méthodologies en fonction de leur "granularité" pourrait probablement être proposée : les méthodologies agents seraient adaptées à la production du modèle informatique et constitueraient donc une des étapes de la méthodologie de modélisation que nous proposons. Cette idée mérite d'être discutée mais il n'est, selon nous, pas clair qu'une approche normative puisse s'insérer ainsi dans une démarche descriptive.

possible de reproduire *in silico* (Zemirline *et al.* (2002) parlent de *in virtuo* ou de *in machina*) la méthodologie expérimentale qui constitue, depuis Claude Bernard, le socle de la méthode scientifique. L'acquisition de connaissances *sur* le modèle puis *par* le modèle passe donc par la conduite d'expériences¹ systématiques au cours desquelles les idées sont méthodiquement soumises à "l'expérience des faits" (Vergnioux, 2003) (rappelons que nous avons souligné la nécessité de l'analyse pour, ensuite, transposer les résultats obtenus sur le modèle vers l'objet modélisé).

Plusieurs auteurs ont récemment plaidé pour une approche expérimentale de la modélisation individu-centrée (Deffuant *et al.*, 2003; Amblard et Phan, 2006). Cependant, dans ces travaux, l'expérimentation consiste essentiellement en une exploration paramétrique systématique, par exemple en adaptant la méthode des plans d'expérience à l'étude du modèle (Amblard et Phan, 2006). L'idée est alors de *valider* le modèle en vérifiant la répétabilité et la robustesse des résultats qu'il fournit. Il nous semble que cette approche ne va pas assez loin : elle oublie selon nous qu'un modèle ne peut pas être validé en lui-même, ni même en vertu de sa capacité à reproduire le comportement de l'objet modélisé. La validation doit être tripartite : le *modèle* est validé si il permet au *modélisateur* de répondre à une question d'intérêt sur l'*objet modélisé* (et non sur le modèle). Les expériences menées peuvent pour cela être très diverses : de même que la conclusion d'une expérience sur un objet peut conduire à proposer de nouvelles expériences sur ce même objet *ou sur un autre objet*, les expériences conduites sur un modèle peuvent conduire à proposer de nouvelles expériences sur ce modèle ou à la création d'un nouveau modèle : "des expériences virtuelles peuvent être menées qui n'interdisent nullement que des formalismes moins polymorphes leur soient progressivement substitués" (Varenne, 2003). L'approche expérimentale que nous préconisons ici couvre l'ensemble de ce processus et non la seule validation numérique d'un modèle.

L'approche expérimentale implique donc une vision "historique" des modèles, avec, pour point de départ et pour point d'arrivée, l'objet *A* (tel que défini par Marvin Minsky) et pour acteur le modélisateur *B*. Le modèle (en tant que *représentation*) n'est alors pas une construction unique. Au contraire, l'objectif poursuivi – répondre à une question posée par *B* sur *A* – demanderait que soit progressivement transformée la connaissance (sous la forme de modèles successifs) de façon à relier, *in fine*, *A* avec *A*. Dans ce cadre, la construction d'un modèle de *A* relèverait d'un processus d'apprentissage (pour le modélisateur) correspondant à des itérations successives de complexification et de simplification. À chacune des phases de cet apprentissage, le modélisateur (*B*) va successivement propo-

¹On pourrait contester que la manipulation d'un modèle puisse constituer une "expérience" : Legay (1997) définit en effet une expérience comme "toute procédure organisée d'acquisition d'information qui comporte, dans la perspective d'un objectif exprimé, une confrontation *avec la réalité*" (c'est nous qui soulignons). La question est alors de savoir ce que recouvre la notion de réalité (en d'autres termes, qu'est-ce qui est "concept" et qu'est-ce qui est "réel"). Il nous semble évident, d'une part, que la "réalité" n'implique pas la matérialité (même si des substrats matériels peuvent être présents – par exemple dans le cas de la conscience) mais aussi que la déduction logique ou discursive à partir d'un ensemble de concepts ne produit que de nouveaux concepts. Nous adoptons à nouveau la position de Varenne (2003), à savoir que, dans le cas de la modélisation individu-centrée, l'utilisation de moyens de calculs pour manipuler les concepts que constituent les modèles d'individus, leur donner une évolution temporelle et les faire interagir, permet de dépasser le simple cadre logico-discursif et confère donc au résultat de la simulation un caractère *réel*, bien que non matériel.

ser un modèle et en analyser les propriétés pour construire, au tour suivant, un modèle plus proche de son objet, ou plutôt plus clair dans son esprit, jusqu'à ce que le modèle lui-même soit indissociable de la connaissance qu'il a permis d'acquérir (on retrouve l'idée de Grimm (1999) selon laquelle le modèle devient inutile lorsqu'il a été "utilisé" : il se confond alors avec la connaissance qu'il a conduit à produire). Cette approche historique de la modélisation rappelle la notion de "délégation" telle qu'elle a été proposée par Isabelle Stengers : *La délégation réussie est le résultat de ce qu'Andrew Pickering compare non sans raison à une "danse" où se succèdent, de la part du scientifique, des moments "actifs", où il construit, met en scène, agence, et des moments "passifs" où ce qu'il a tenté de mettre en scène se manifeste activement, satisfait parfois, et déçoit souvent, le "désir" immanent à l'agencement, celui d'une "capture" de ce qui opère (capture of agency)* (Stengers, 1997).

Nous avons déjà souligné le risque que comporte un usage trop "intuitif" des modèles individu-centrés (Deffuant *et al.* (2003), citant (Grimm, 1999), parle de "manque de temps et d'efforts passés à tester le modèle après son élaboration"). Ce risque correspond en fait à l'absence de méthode quant à l'analyse des modèles. L'adoption d'une méthode expérimentale constitue ici une réponse. Il faut cependant veiller à ne pas tomber de Charibde en Scylla. En effet, adopter une méthode expérimentale sur un modèle individu-centré risque de masquer la réalité première de celui-ci, à savoir d'être avant tout un modèle. On serait alors tenté d'extrapoler trop rapidement les résultats (expérimentaux) obtenus sur le modèle vers la réalité de l'objet modélisé. L'objectif de la méthodologie que nous proposons est de conduire le modélisateur à garder un juste milieu : ne pas oublier sa pratique scientifique lorsqu'il s'adresse au modèle, mais ne pas non plus être tenté de ne plus s'adresser qu'à lui.

4.3 Quelques règles pour la modélisation individu-centrée

Afin de proposer un cadre méthodologique pour la modélisation individu-centrée tout en laissant au modélisateur toute la souplesse nécessaire à la pratique expérimentale telle que nous venons de l'exposer, nous avons choisi de structurer notre méthodologie suivant deux axes orthogonaux. Le premier de ces axes est constitué d'un ensemble de règles générales qui devront être respectées durant l'ensemble de la démarche. Le deuxième axe sera un axe temporel : le modélisateur doit en effet aborder le problème en suivant un cheminement conforme à la méthode scientifique (sans cependant trop idéaliser celle-ci). Une vision simpliste, inspirée par la notion de "juste milieu" que nous avons évoquée en conclusion de la section précédente, serait que ce cheminement est là pour garantir le respect de la méthode expérimentale tandis que les règles sont destinées à maintenir le modélisateur au contact de son objet d'étude. Bien que partiellement justifiée (voir en particulier la règle No. 0 ci-dessous), cette séparation entre les deux axes n'est qu'apparente et, aussi bien dans les règles que dans le cycle de vie du modèle, on trouvera des éléments propres à ces deux aspects.

Les sections suivantes présentent les cinq règles générales. Celles-ci sont partiellement inspirées de Miller (1995) et de Grimm (1999), mais elles s'en éloignent aussi en ce qu'elles tentent de tenir compte des spécificités de l'approche expérimentale telle que nous venons de la définir. En outre, bien que nous n'ayons que peu évoqué ce point jusqu'ici, ces

règles tiennent compte des caractéristiques spécifiques des modèles individu-centrés, en particulier du fait qu'ils sont basés sur une implémentation informatique.

Règle No. 0 : un modèle individu-centré est avant tout un modèle

C'est la règle fondamentale : un modèle individu-centré est avant tout un modèle, donc un modèle de quelque-chose (A) pour quelqu'un (B) et, plus précisément, pour que B réponde à une question d'intérêt sur A . Il s'agit donc d'un modèle et non d'un programme ou d'une simulation (ou, du moins, c'est un modèle *avant* d'être un programme ou une simulation). Cela signifie entre autres que les contraintes liées à la programmation du modèle devront toujours être considérées comme secondaires devant les contraintes liées à l'expressivité du modèle (ce qui ne signifie pas qu'elles devront être ignorées – les contraintes de temps de calcul ou d'encombrement mémoire seront souvent cruciales – mais leur prise en compte devra toujours se faire relativement au modèle).

Rappeler qu'un modèle individu-centré est avant tout un modèle, c'est aussi conserver à l'esprit que l'objectif devra toujours être la production de connaissance ; donc que le lien avec l'objet modélisé devra être maintenu à toutes les étapes de la démarche de modélisation, d'exploitation du modèle et de production des résultats. Bien qu'elle puisse paraître triviale, cette règle est rarement respectée, que ce soit dans les faits (les modèles individu-centrés sont trop souvent construits indépendamment de toute question explicite sur un quelconque objet) ou dans les textes. De nombreux auteurs (principalement issus de l'informatique) proposant des méthodologies ou des tutoriaux pour la modélisation "agents" centrent en effet leur approche sur la seule production du code du simulateur. Or, si le modèle individu-centré est *avant tout* un modèle, cela implique que la conception du code n'est qu'une phase particulière du cycle de vie du modèle (Miller, 1995).

Règle No. 1 : un modèle individu-centré doit être le théâtre d'expériences scientifiques

L'objectif d'un modèle individu-centré est la mise en évidence de mécanismes structurants issus des interactions entre les individus (nous pourrions parler de mécanismes structurants *émergents*) puis l'analyse de ces mécanismes afin de répondre à la seule "vraie" question : "ce mécanisme est-il susceptible d'être responsable des structures observées sur l'objet A ?". Il est pour cela nécessaire de dépasser le simple stade de la ressemblance avec l'objet modélisé, ressemblance qui est, certes, un indice positif mais qui ne permet en aucun cas d'assurer la validité du modèle (quelle que soit "la précision des décimales"). Il s'agit donc avant tout de comprendre ces mécanismes, de déterminer les conditions dans lesquelles ils sont actifs (ou pas), d'estimer leur pouvoir structurant "réel", leur robustesse face à d'autres éléments structurants ou, au contraire, face aux perturbations, etc.

La compréhension de ces mécanismes passe par la manipulation du modèle. Manipulation qui peut concerner les différents paramètres mais aussi les éléments structuraux eux-mêmes. En outre, la manipulation du modèle doit satisfaire aux critères de la méthode scientifique, d'autant plus que les modèles individu-centrés, parce qu'ils incluent souvent des éléments stochastiques (dans les modèles d'agents, dans les conditions initiales, dans l'environnement ...) doivent généralement faire l'objet d'études statistiques avant toute conclusion, y compris avant toute conclusion intermédiaire. Un des risques méthodologiques majeurs, avec les modèles individu-centrés, est en effet l'utilisation des résultats

intermédiaires (par exemple une première campagne de simulation alors que plusieurs campagnes sont prévues) comme autant de conclusions utilisées pour planifier les expériences futures¹. Il est impératif d'adopter une attitude extrêmement stricte à cet égard sous peine d'accumuler des résultats inexploitable et, finalement, d'explorer le modèle "au petit bonheur la chance". Il est à ce propos intéressant de constater qu'une des problématiques les plus en vogue, chez les informaticiens, est la parallélisation des modèles "multi-agents" alors qu'en pratique, il existe une parallélisation naturelle des expériences, qui plus est particulièrement efficace. La parallélisation du modèle permet d'accélérer la conduite d'une simulation (donc de conduire les simulations "en temps réel"), ce qui encourage les pratiques empiriques préjudiciables. Inversement, la parallélisation des expériences permet d'obtenir simultanément tous les points d'une campagne de test et favorise donc une exploration raisonnée du modèle².

La nécessité de conduire une étude expérimentale du modèle va avoir, sur le développement du modèle informatique, des conséquences importantes. En effet, elle impose que l'exploration paramétrique (et, éventuellement, structurelle) du modèle puisse être automatisée. L'étude expérimentale rend en effet illusoire le suivi manuel des "expériences". Le modèle va donc devoir se plier à la production aveugle d'indicateurs numériques (alors que, souvent, les modèles individu-centrés sont simplement utilisés pour produire des "images"). Les interfaces type "GUI"³, si elles sont nécessaires (voir ci-dessous), ne doivent donc pas être le seul moyen d'accéder au modèle.

Règle No. 2 : stimuler l'intuition et la dépasser

Nous venons d'argumenter en faveur de la mise en œuvre d'une démarche expérimentale stricte. Cette volonté, si elle doit être respectée au maximum est cependant un vœu pieux. En effet, l'exploration du modèle ne peut pas être simplement basée sur des indicateurs numériques. Parce que le modèle peut surprendre, parce que l'objectif est d'identifier des mécanismes structurants, l'exploration doit comporter une phase intuitive au cours de laquelle le modélisateur doit acquérir une connaissance empirique de son modèle. C'est cette connaissance qui lui permettra d'identifier les mécanismes structurants, de définir les expériences à réaliser⁴ ou les indicateurs les plus pertinents. On notera que cette phase intuitive, même si elle a souvent été décriée, est nécessaire à la pratique scientifique qui doit nécessairement développer une "vue de l'esprit" (une hypothèse) avant d'en vérifier

¹Miller (1995) évoque "*the temptation to run tomorrow's computer simulation before yesterday's has been fully understood.*"

²Bien entendu, cet argument n'est en rien un réquisitoire contre la parallélisation, celle-ci pouvant se révéler indispensable pour l'exécution de très gros modèles. C'est plus contre l'intérêt des simulations "en temps réel" que nous nous inscrivons (du moins dans une pratique scientifique; la réponse serait différente, par exemple dans une pratique pédagogique).

³Graphical User Interface

⁴Il a été proposé d'utiliser les plans d'expérience pour définir les expériences à réaliser sur le modèle. Cette proposition se heurte cependant au très grand nombre de paramètres présents dans les modèles individu-centrés (voir (Amblard et Phan, 2006) pour une discussion sur ce point) et, surtout, à la nécessité, dans une simulation informatique, de cerner les valeurs des paramètres dans un espace beaucoup plus grand que dans le cas des expériences "réelles". La connaissance empirique du fonctionnement du modèle permet d'identifier les zones d'intérêt dans l'espace paramétrique, même si l'analyse va ensuite devoir confirmer (ou infirmer) cette première estimation.

l'exactitude par l'expérience (Vergnioux, 2003). Rappelons, une fois de plus, que, si il est nécessaire, ce recours à l'intuition ne saurait constituer *seul* une pratique scientifique¹. En pratique, l'intuition du fonctionnement du modèle (l'hypothèse donc) est souvent présente préalablement à la construction du modèle² (ne serait-ce que parce que le modélisateur *suppose* certains mécanismes structurants d'être présents dans son objet d'étude). Il convient cependant de conserver à l'esprit que les modèles individu-centrés ont la capacité de surprendre, y compris leur concepteur. En d'autres termes, ils sont des outils de production d'hypothèses plus que des outils de validation d'hypothèses. Le modélisateur doit donc toujours veiller à "garder les yeux ouverts" sur le fonctionnement de son modèle, c'est-à-dire qu'il doit rester dans une position suffisamment curieuse pour pouvoir être surpris par lui. Notre expérience nous montre que, si les modèles individu-centrés sont "toujours" le siège de mécanismes émergents, il est en effet rare que ceux-ci confirment totalement les hypothèses initiales³.

Nous avons utilisé l'expression "garder les yeux ouverts" dans son sens métaphorique. Elle s'applique cependant aussi dans son sens premier car l'observation visuelle des résultats du modèle est souvent la première étape de l'acquisition de connaissance. Alors que la règle précédente nous conduisait à rejeter les interfaces visuelles, nous faisons donc ici le constat inverse : elles sont nécessaires au développement de l'intuition (Grimm, 2002; Payet *et al.*, 2005). À nouveau, nous aboutissons ici à la recherche d'un "juste milieu" entre l'utilité de l'impression visuelle et la nécessité de la dépasser : comme dans toute science expérimentale, il ne suffit pas de *voir* le phénomène, il faut (i) le voir plusieurs fois⁴, (ii) être capable de le caractériser et (iii) déterminer ses conditions d'apparition.

Règle No. 3 : rechercher la complémentarité avec d'autres modèles

Les modèles individu-centrés sont souvent présentés en opposition aux modèles "classiques" (c'est-à-dire, généralement, aux modèles mathématiques). En pratique, nous avons vu que cette opposition n'a pas lieu d'être puisqu'un modèle individu-centré peut être basé sur un formalisme mathématique et, surtout, que, de par son caractère expérimental, la modélisation individu-centrée va jouer, dans une pratique scientifique, un rôle différent. Nous voudrions ici défendre l'idée que la production de connaissance par des modèles individu-centrés (nous devrions d'ailleurs plutôt écrire *impliquant* des modèles individu-centrés) passe par l'utilisation conjointe de modèles de natures différentes et, en particulier, de modèles mathématiques (quel que soit le formalisme mathématique utilisé). Nous avons en effet insisté sur la nécessité d'une analyse pour transférer la connaissance acquise

¹"L'attitude religieuse n'est pas la seule caractéristique de l'approche non scientifique de la réalité. Une autre caractéristique est la confiance *excessive* que l'on peut accorder à son intuition ou à l'introspection [...]" (Bricmont et Franck, 2007) (c'est nous qui soulignons).

²"Les déterminations numériques de la science repassent sur le pointillé d'une constitution du monde déjà faites avant elles." (Merleau-Ponty, 1945).

³Trop souvent les modèles individu-centrés sont utilisés pour reproduire des résultats par ailleurs parfaitement connus, donc dans lesquels les principes structurants sont déjà identifiés (ainsi, un exemple classiquement utilisé est le fourrage chez les fourmis ou une de ses variantes). Ils nous semble que la reproduction de mécanismes connus ne doit en aucun cas être utilisée pour développer une démarche méthodologique. En effet, elle gomme les phases d'exploration puisque le modèle est alors construit pour *illustrer* et non pour comprendre.

⁴Sous-entendu, dans des conditions différentes.

sur le modèle vers l'objet d'étude initial. Or, cette analyse peut demander que des modèles soient utilisés comme médiateurs, jouant ici le même rôle que celui qu'ils jouent dans les sciences expérimentales, à savoir la médiation du passage de l'observation à la théorie (Varenne (2003) écrit : "on ne doit pas renoncer à condenser le réel *ou les simulations* pour comprendre"). Les simulations individu-centrées ne sont pas, nous l'avons vu, adaptées à ce rôle car elles ne font "que" raconter des histoires individuelles soumises à la contingence. Les mathématiques, en revanche, offrent un formalisme plus strict et plus "englobant". En d'autres termes, alors qu'une simulation individu-centrée ne présentera jamais qu'un ensemble de trajectoires individuelles (chaque simulation correspond à une trajectoire, avec sa part de contingence et de nécessité¹), le modèle mathématique va permettre d'étudier les propriétés générales du système, c'est-à-dire l'ensemble des trajectoires possibles. Il permet donc d'une part de passer de la simulation à la théorie, mais aussi de passer plus facilement de la théorie *sur l'objet A^** à la théorie *sur l'objet A* . Schweitzer (2003) parle à ce propos de "complémentarité expérimentations-simulation-théorie"².

Il est malheureusement très rare qu'un modèle individu-centré se prête à la formalisation, ce qui rend difficile l'analyse du modèle. La première réaction est alors souvent de simplifier le modèle de façon à permettre sa formalisation (approche "KISS" : Keep It Simple and Stupid). Cette approche a cependant ses limites puisqu'elle conduit à éliminer ce qui fait la spécificité du modèle individu-centré et limite son champ d'étude à l'intersection avec celui des mathématiques. À l'opposé, une approche descriptive a pu être proposée ("KIDS" : Keep It Descriptive and Stupid.) qui tend au contraire à se rapprocher d'un modèle descriptif de la réalité (la formalisation étant alors abandonnée au profit de la comparaison directe des résultats du modèle avec la réalité – nous avons déjà évoqué les limites d'une telle approche).

Il nous semble que, s'il est indispensable de chercher à analyser mathématiquement le modèle individu-centré, cette analyse ne doit pas oublier le caractère nécessairement partiel et idéalisé d'un modèle. En d'autres termes, le modèle mathématique est un objet $(A^*)^*$ qui doit lui-même être une idéalisation de la "réalité" que représente le modèle individu-centré A^* . Dès lors, il participe à son analyse mais ne le remplace pas ; il complète et précise les réponses apportées mais il ne répond pas lui-même . . . Cela n'interdit bien évidemment pas de modifier le modèle individu-centré pour aider à sa compréhension, par exemple en modifiant le niveau d'agrégation (Grimm, 1999), mais cette modification n'a plus pour objectif de rendre possible la mathématisation. L'intérêt du modèle mathématique est alors son caractère englobant : $(A^*)^*$, construit pour être un modèle de A^* peut en effet simultanément être un modèle de A au sens où il peut répondre (permettre de répondre) à des questions d'intérêt que B se pose à la fois sur A^* et sur A (on notera que cette proposition est réversible, ce qui autorise aussi d'utiliser, pour analyser le modèle individu-centré, des outils et modèles développés pour analyser A directement³). Il joue alors pleinement son rôle de médiateur en permettant de construire une théorie générale

¹"La simulation est un renoncement, au sens où l'évolution simulée est aveugle à son environnement, alors que la solution exacte d'équations mathématiques définit le paysage de toutes les solutions possibles, selon la valeur des paramètres et la condition initiale" (Stengers et Bensaude-Vincent, 2003).

²L'ensemble des travaux présentés dans (Schweitzer, 2003) constitue d'ailleurs un exemple de bonne complémentarité entre différentes approches de modélisation.

³Miller (1995) encourage d'ailleurs cette approche. Même si nous partageons cette position, il est important de rappeler que le modèle individu-centré nous donne un accès direct à des informations dont

à partir d'expérimentations réalisées sur le modèle individu-centré.

Règle No. 4 : insérer le modèle dans une pratique scientifique

La modélisation individu-centrée est une pratique nécessairement pluridisciplinaire impliquant au moins (dans notre cas) des biologistes, des informaticiens et des mathématiciens (voir règle No. 3) ; elle est même souvent trans-disciplinaire car la discipline du niveau “micro” n’est pas nécessairement la même que celle du niveau “macro” (ainsi, dans la section II.4, le modèle des individus sera en grande partie issu de la physique tandis que le modèle global relèvera de la biologie cellulaire). Il importe cependant de toujours conserver à l’esprit que l’objectif est toujours, *in fine*, d’accroître la connaissance sur un objet particulier et que cet objet relève généralement d’une discipline particulière.

Or, cette discipline dispose déjà d’un corpus de connaissances organisé sur cet objet, les questions d’intérêt sont généralement déjà connues, de même que les réponses *acceptables*¹. Pour intégrer ce corpus, une connaissance (quelle qu’elle soit) devra se plier à ses lois ; elle devra éventuellement *se rendre acceptable* (ce sera souvent le cas de la modélisation individu-centrée qui, parce qu’elle semble n’être d’une reproduction de la réalité, est toujours susceptible de n’être qu’un artefact et, à ce titre, risque d’être rejetée a priori). Quelles que soient les formules incantatoires, il est en particulier évident qu’un résultat issu “du dehors” n’a absolument aucune chance de parvenir seul à remettre en cause le paradigme en vigueur. Le changement de paradigme est en effet lié à l’accumulation de faits contradictoires. Or un modèle ne constituera un fait – sous-entendu “acceptable” – que dans la mesure où il *n’est pas* contradictoire.

Face à cette nécessaire trans-disciplinarité, il est possible d’énumérer plusieurs points auxquels le modélisateur devra être particulièrement attentif afin de garantir que sa démarche produira des résultats tangibles :

Règle 4.1 : identifier une question d’intérêt. La plupart des auteurs s’accordent sur ce point. Nous voudrions cependant insister sur la difficulté qu’il y a à identifier une question d’intérêt. En effet, une telle question est rarement isolée. Elle a des implications, qu’il faut comprendre ; elle est déjà traitée par une communauté, qu’il faut connaître et elle n’est pas forcément susceptible de faire l’objet d’une modélisation (ou, au contraire, elle l’a déjà été). Une question n’est donc jamais indépendante de son contexte. En particulier, elle est souvent l’objet d’hypothèses par rapport auxquelles le modèle devra se situer. En conséquence, l’identification de la question passe par un travail bibliographique complet, qui dépasse la question elle-même. Celui-ci ne doit en particulier pas porter uniquement sur les approches de modélisation mais couvrir l’ensemble du “puzzle” lié à la question posée.

Règle 4.2 : travailler avec de “vrais” biologistes. À nouveau, ce point est relativement consensuel mais il cache plusieurs difficultés. On entend souvent des appels à

le biologiste ne dispose généralement pas (par exemple, en évolution artificielle, la liste de l’ensemble des mutations subies par tous les ancêtres d’un individu). Il importe donc aussi de développer des outils nouveaux pour analyser ces informations.

¹Une réponse acceptable est une réponse qui s’intègre correctement dans le corpus de connaissances pré-établi mais aussi une réponse qui a été obtenue par des moyens acceptables aux yeux d’une communauté scientifique particulière (Stengers et Bensaude-Vincent, 2003).

la définition d'un "langage commun" entre modélisateurs et biologistes. Plus qu'un langage, c'est une épistémologie commune qu'il est nécessaire de développer : les critères qui font la qualité du modèle, les critères de "preuve", jusqu'aux pratiques bibliographiques, sont en effet très variables d'une discipline à l'autre. En outre, le développement d'un langage commun (nécessairement minimaliste) n'est en rien suffisant : parce qu'il faudra communiquer avec une communauté scientifique, il est nécessaire d'acquérir le langage de cette communauté¹. En effet, travailler avec un biologiste n'est pas suffisant : c'est une communauté qu'il va falloir convaincre avec des moyens souvent bien pauvres en regard de la richesse des échanges que permet la coopération directe.

Règle 4.3 : construire un modèle pertinent. Dégager ("faire-émerger" au sens de la phénoménologie) les éléments pertinents à inclure dans le modèle est difficile. En effet, le modèle devra pouvoir faire l'objet d'expériences qui seront ensuite relatées et qui, après analyse, devront être suffisamment convaincantes. Or, pour être convaincantes, ces expériences devront répondre aux différentes objections que la biologie (au sens de la communauté des biologistes) pourra leur opposer. Cela signifie que les éléments pertinents ne sont pas seulement les éléments actifs dans la dynamique du modèle (ce que préconiserait une approche minimaliste). Ils devront permettre de relier le modèle au corpus de connaissances actuel des biologistes (et non à un corpus de connaissances supposé appartenir à un "nouveau paradigme"). Ils seront donc ceux qui vont permettre de réaliser les "bonnes" expériences (y compris des expériences "négatives"), c'est-à-dire les expériences exigées par la communauté pour accepter le modèle en son sein. Celui-ci doit en quelque sorte être coopté pour passer du statut de modèle informatique à celui de modèle biologique, seul susceptible de produire de la connaissance.

Règle 4.4 : publier ! La publication fait partie du travail scientifique. Mais dans un contexte de modélisation, les cibles de publication sont multiples. Si l'objectif est de répondre à une question sur l'objet A , la publication des résultats doit se faire dans une revue spécialisée sur cet objet (et non sur un certain type de modèle) et se soumettre à un processus de review disciplinaire². Cela n'interdit évidemment pas de communiquer sur le modèle ou la démarche de la modélisation mais de telles communications ne peuvent en aucun cas être considérées comme suffisantes³.

Pour résumer, la production de connaissance en biologie implique d'adopter les pratiques de la biologie, l'épistémologie de la biologie, voire l'éthique de la biologie afin que les connaissances développées lors de la modélisation soient acceptables par la biologie. Le modèle doit donc réellement *s'insérer* dans la pratique biologique. Adopter une position

¹On pourrait objecter ici que, puisque nous travaillons avec des biologistes, ce sont eux qui vont devoir communiquer avec leur communauté. Cette objection est partiellement valide mais elle oublie la nécessaire interprétation du modèle. C'est le modèle qui, via le modélisateur, doit donner au biologiste les éléments pertinents pour convaincre sa communauté. Ce point doit donc être pris en compte lors de la modélisation.

²"*The burden of proof is on us to explain our results to biologists in their own language and in their own journals*" (Miller, 1995).

³Trop souvent, on peut être tenté de publier dans des domaines techniques et de laisser les biologistes importer les résultats dans leur domaine propre. Cette pratique est illusoire car, dans ce cas, les résultats publiés ne seront tout simplement pas lus (du moins pas par les bonnes personnes). On pourrait comparer cette démarche à une gesticulation muette dans le dos des biologistes.

extérieure n'interdirait certes pas de répondre à la question posée mais reviendrait à garder la réponse pour soi, ce qui, dans une démarche scientifique, est équivalent à ne pas répondre.

4.4 Le cycle de vie d'un modèle individu-centré

Les cinq règles que nous venons d'énoncer ne constituent pas une méthodologie. Elles ne fixent qu'un cadre très global à la démarche du modélisateur. Nous allons donc préciser ce cadre en proposant un cycle de vie, comportant cinq étapes, pour la réalisation d'un modèle individu-centré.

Premier stade : le temps d'avant

Il est clair depuis Bachelard que toute démarche scientifique doit commencer par l'énoncé d'un problème et "quoi qu'on en dise, dans la vie scientifique, les problèmes ne se posent pas d'eux-mêmes. C'est précisément ce *sens du problème* qui donne la marque du véritable esprit scientifique. Pour un esprit scientifique, toute connaissance est une réponse à une question. S'il n'y a pas eu de question, il ne peut y avoir de connaissance scientifique. Rien ne va de soi. Rien n'est donné. Tout est construit" (Bachelard, 1947). Dans le cadre d'une démarche de modélisation, l'énoncé du problème est particulièrement important puisque, en l'absence de problème, le modèle perd sa fonction scientifique pour ne garder qu'une fonction de représentation.

Préalablement à toute démarche de modélisation, il est donc nécessaire d'identifier clairement les objets, les liens entre ces objets et les questions sur ces objets. Conformément à ce que nous venons d'énoncer (règle No. 4.1), cette première phase doit être suffisamment complète pour permettre d'identifier le contexte dans lequel ces objets et ces questions se situent. Le contexte sera en effet indispensable lors du choix des éléments du modèle (règle No. 4.3) mais aussi lors de la formulation des résultats (règle No. 4.4).

Il serait illusoire de prévoir une transition stricte entre ce premier stade et le suivant. D'une part parce que le contexte est susceptible d'évoluer à tout moment, d'autre part parce que les premières phases de construction du modèle (voire d'expérimentation avec le modèle) apporteront leur lot de nouvelles questions. Comme pour toute pratique scientifique, la bibliographie est donc ici un processus continu, ce qui concerne aussi la bibliographie du domaine d'intérêt (c'est-à-dire le domaine de l'objet *A*).

Deuxième stade : le temps de l'exploration

Selon notre règle No. 2, le modèle doit en premier lieu stimuler l'intuition du modélisateur. Nous pensons qu'il faut accepter, dans la démarche de modélisation individu-centrée, une part d'expérimentation anarchique (pour reprendre les termes de Paul Feyerabend¹), durant laquelle le modélisateur va jouer (au sens de manipuler librement, instinctivement) avec le modèle de façon à affiner ses questions, à émettre ses premières hypothèses, voire à cerner les conditions de fonctionnement de son modèle.

¹"Même une science fondée sur la loi et l'ordre ne réussira que si des mouvements anarchiques ont occasionnellement le droit de se manifester" (Feyerabend, 1975).

Cette phase est nécessaire à toute démarche scientifique mais elle revêt, selon nous, un caractère particulier dans le cas de la modélisation individu-centrée. En effet, elle y est plus longue car elle demande que soient construits et manipulés des modèles successifs (l'objectif étant entre autres de déterminer, par la simulation de modèles successifs, la structure du modèle final (Ferber, 1995)). Elle ne peut donc pas être réalisée "en temps masqué" comme elle peut l'être dans d'autres approches de modélisation. En outre, dans le cas de la modélisation individu-centrée, c'est au cours de cette phase exploratoire que va se produire l'effet de "surprise" propre à ce type de modèle. Ceci implique que, même si, comme nous venons de le dire, la frontière entre la première et la deuxième phase est floue, l'ordre de ces deux phases doit impérativement être respecté sous peine d'être surpris par ce qui n'est en rien surprenant, ou l'inverse . . .

La phase exploratoire est nécessaire mais elle n'a de sens qu'à condition d'être ensuite dépassée. Autant la transition entre les phases 1 et 2 pouvait être floue, autant la transition vers la phase 3 doit être nette : une fois l'exploration suffisamment avancée, une fois les grands choix réalisés (choix des questions, choix des modèles, etc.), le modélisateur va construire le modèle sur lequel il va s'appuyer pour réaliser ses expériences.

Troisième stade : le temps du modèle

Le "temps du modèle" c'est le temps de la construction du modèle et de l'outil logiciel qui sera utilisé pour conduire les expérimentations. Il s'agit de déterminer la structure des individus (des "agents"), de leurs interactions, leurs comportements, les spécificités de l'environnement, mais aussi les méthodes de visualisation, les indicateurs numériques, etc. C'est dans cette troisième phase que peuvent éventuellement être utilisées des méthodologies de conception agent. Dans la plupart des cas cependant, la conception du modèle réutilisera de larges pans issus de la phase d'exploration, ce qui rend illusoire la mise en œuvre d'une méthodologie agent stricte (le modèle est en quelque sorte déjà conçu implicitement et il ne reste "qu'à" l'explicitier). Il est souhaitable, en revanche, d'adopter ici des méthodes issues du génie logiciel (méthodes de conception, méthodes de test, démarche qualité . . .). En effet, le caractère imprévisible des modèles individu-centrés rend très difficile la validation globale du logiciel (sauf erreur grave, il donnera toujours un résultat). Il est donc particulièrement important de valider individuellement tous les composants logiciel du modèle, mais aussi de prévoir des cas de tests dans lesquels le comportement du modèle est connu a priori (on peut, par exemple, remplacer certains comportements par des marches aléatoires, ce qui permet de vérifier que les individus se dispersent correctement dans l'espace).

Nous ne reviendrons pas ici sur le choix des éléments pertinents. Les critères de choix sont en effet trop dépendants du problème posé. Nous rappellerons simplement que les éléments pertinents doivent être déterminés en fonction de l'objectif (règle No. 0) mais aussi en fonction des critères propres à la communauté scientifique (règle No. 4.3). Le principe d'une simulation minimale, s'il s'applique toujours en modélisation pour permettre le passage du modèle à la théorie (Schweitzer, 2003), doit donc être tempéré. La simulation minimale ne contient en effet que les éléments *nécessaires* mais cette nécessité peut avoir des origines multifactorielles. Le choix des éléments du modèle demande en particulier des compétences pluridisciplinaires : une expertise du domaine, bien sûr,

mais aussi une expertise en modélisation de façon à discerner les éléments d'intérêt. Il est notamment important d'éviter au maximum d'introduire, au cours de cette démarche, des éléments implicites, c'est-à-dire des choix de modélisation qui ne soient pas explicites parce que considérés comme "allant de soi" par le modélisateur ou par l'expert du domaine. Il convient d'être extrêmement vigilant lors de la réutilisation d'un modèle ou lors de l'utilisation d'une plate-forme générique. Dans les deux cas, le risque est grand d'extrapoler des choix de modélisation sans que ni le modélisateur, ni l'expert du domaine, n'aient une complète perception des choix implicites qu'ils sont en train de faire.

Il nous semble important de préciser ici que ce troisième temps correspond à la définition du *modèle* et que celui-ci ne se résume pas au code de la simulation individu-centrée. Nous avons déjà évoqué la nécessité de manipuler le modèle afin d'analyser les phénomènes observés. Cette manipulation peut rendre nécessaire le développement d'outils complémentaires, certains d'entre eux pouvant être basés sur des formalismes différents (tout en représentant une implémentation du même modèle). Il est même possible de prévoir des implémentations simplifiées (correspondant à des cas particuliers du modèle) et de les utiliser pour valider les résultats de la simulation individu-centrée. Suivant les simplifications envisagées, les implémentations simplifiées pourront, par exemple, prendre la forme d'un modèle mathématique.

Le "temps du modèle" se termine lorsque l'ensemble des outils logiciels est stable (nous n'osons pas écrire ici "terminé") et qu'il permet une expérimentation systématique. Cela ne signifie pas que le code du logiciel ne va plus être modifié, puisque certaines expériences vont nécessiter de telles modifications. Mais ces modifications correspondront à des choix expérimentaux et ne devront pas remettre en cause les expériences antérieures. En d'autres termes, la transition vers la phase d'exploitation peut avoir lieu lorsque le modélisateur peut conduire des expériences avec une confiance *suffisante* dans les outils qu'il utilise. Vu les délais imposés par le calcul (il est généralement indispensable de conduire plusieurs répétitions et plusieurs reproductions des expériences), il est en effet impératif que le code soit stabilisé avant de se lancer dans des expérimentations "à haut débit". Par ailleurs, il est aussi particulièrement important que toutes les expériences soient réalisées sur le même objet (ou, plus exactement, sur des objets dont les différences sont choisies, connues et maîtrisées). Toute modification du code non dictée par les choix de l'expérimentateur (donc, en particulier, les corrections d'erreurs) risquent d'invalider l'ensemble des expériences réalisées auparavant et d'entraîner des pertes de temps considérables.

Quatrième stade : le temps de l'exploitation

À ce stade, le modélisateur dispose d'un modèle et d'une implémentation stabilisés. En outre, il a déjà développé une bonne intuition du fonctionnement de ce modèle et des liens qu'il entretient avec l'objet réel. Dans de très nombreux cas, en modélisation individu-centrée, les "expérimentations" s'arrêtent malheureusement à ce stade. Plusieurs auteurs ont ainsi déploré le manque d'efforts consacrés à l'analyse des modèles (Grimm, 1999; Deffuant *et al.*, 2003). Selon nous, cet "abandon prématuré" est très souvent dû à une faiblesse lors de la première phase de construction du modèle (le "temps d'avant"). Le problème étant mal posé, voire mal choisi (trop souvent les modèles individu-centrés sont utilisés pour reproduire des résultats déjà connus, voire déjà modélisés par d'autres

approches), il n’y a pas réellement de questionnement scientifique, donc pas d’expérience possible! Pour exploiter le modèle il ne suffit plus d’être modélisateur. Il faut aussi être expert du domaine¹.

Alors qu’elle est souvent négligée, cette phase est probablement la plus importante. C’est ici que le modélisateur va réellement expliquer les phénomènes dont il a déjà l’intuition. En outre, il s’agit d’une phase complexe au cours de laquelle des expériences vont être simultanément réalisées, interprétées, décidées, voire modélisées² (règle No. 3). Toute la panoplie des outils de l’expérimentation scientifique va pouvoir être utilisée de façon à interpréter les résultats³ et à *comprendre* le fonctionnement du modèle et les liens qu’il entretient avec l’objet modélisé.

Au cours de la phase d’exploitation, il est possible que le modélisateur soit surpris par son modèle et que celui-ci lui pose, en quelque sorte, de nouvelles questions. En soi cette éventualité n’a rien de négatif, nous pensons au contraire que le modélisateur doit rester en permanence dans une position permettant cette surprise : son objectif doit en effet être de “faire entrer l’inconnu dans le connu et en même temps rester ouvert à la nouveauté⁴”. Elle pose cependant des problèmes méthodologiques : l’émergence de nouveaux questionnements doit en effet être considérée avec prudence car, d’une part, elle ne doit pas perturber la démarche expérimentale en cours, d’autre part elle doit conduire à une reprise complète du cycle de modélisation (à cette nouvelle question correspondra une nouvelle étude, un nouveau modèle et de nouvelles expériences). C’est donc un nouveau cycle qui débute alors, indépendamment du processus en cours qui, lui, doit être mené à terme.

Cinquième stade : le temps d’après, production de la connaissance du domaine

Le cycle de vie du modèle ne s’arrête pas à la compréhension du phénomène. Il importe ensuite de relier cette compréhension à un corpus scientifique, ce qui, en pratique, revient à faire accepter le résultat par la communauté. Nous avons déjà insisté sur la nécessité de publications ciblées (règle No. 4.4) ; nous voudrions ajouter ici que cette étape peut demander à *juste titre* la conduite d’expériences complémentaires. En effet, les expériences qui ont permis au modélisateur de comprendre le phénomène ne se révéleront pas toujours suffisantes pour convaincre la communauté. Or, en dernier ressort, c’est bien convaincre qui est nécessaire. Le modélisateur doit alors être conscient qu’il devra convaincre deux fois : sur son modèle d’abord et sur son résultat ensuite car “[La notion de preuve] traduit la nécessité d’un accord méthodologique sur ce qui sera tenu pour acceptable.” (Stengers et Bensaude-Vincent, 2003). Or, il importe de garder à l’esprit que la démarche de modélisation individu-centrée n’est pas acceptable en elle-même. En ce sens elle ne constitue

¹Il est d’ailleurs symptomatique que la plupart des auteurs soulignant ce point soient précisément ceux qui conjuguent expertise en modélisation individu-centrée *et* expertise “du” (de leur) domaine.

²“L’expérience surgit en même temps que les suppositions théoriques et non avant elles” (Feyerabend, 1975).

³Ainsi, il pourra s’avérer utile d’importer en modélisation individu-centrée les outils de validation statistique des résultats d’expériences afin de vérifier leur reproductibilité. Il faut cependant rester conscient que ces outils permettent de valider la description du comportement du modèle mais n’apportent rien à sa validité en tant que modèle.

⁴Attribué à Emile Meyerson.

pas un *instrument* scientifique autour duquel un consensus peut être établi une fois pour toute. Chaque modèle individu-centré doit donc faire la preuve de sa pertinence et le succès antérieur d'autres modèles individu-centrés ne lui est pour cela d'aucun secours.

Le temps itéré : le modèle n'est jamais clos

Le cycle de vie que nous venons d'énoncer comporte cinq phases. Il est important de conserver à l'esprit que certaines transitions sont relativement floues alors que d'autres (en particulier en ce qui concerne la phase de construction du modèle) doivent être clairement identifiées. Nous avons aussi émis l'hypothèse que le modèle surprenne à n'importe quelle phase de son utilisation (il est "utilisé" lors des phases 2, 4 et 5) et qu'un nouveau cycle de modélisation soit alors initié. Il est aussi évident que le modèle sera soumis à la connaissance du domaine et que toute progression de cette connaissance peut conduire à une nouvelle itération du processus de modélisation (Palsson, 2000). En outre, le modèle peut lui-même être utilisé pour guider l'expérimentation "humide" (Bernot *et al.*, 2004; Batt *et al.*, 2005) et être ensuite amené à intégrer les résultats de ces expériences.

En ce sens, le modèle n'est jamais clos : il est toujours susceptible d'être réutilisé pour répondre à de nouvelles questions qu'il aura éventuellement lui-même contribué à poser ou pour conduire de nouvelles expériences qu'il aura éventuellement lui-même suggérées. Il est important cependant de rappeler ici que ces itérations doivent être réalisées avec précaution et qu'en particulier elles doivent systématiquement conduire à une reprise complète du cycle de modélisation. Ferber (1995) parle de "simulation multi-agents [qui] ne se réduit pas à l'implémentation d'un modèle puis à l'analyse de la réponse de ce modèle en fonction des paramètres d'entrée, mais participe au processus de recherche de ce modèle". Selon nous, il ne s'agit pas là d'une itération dans le processus de modélisation mais uniquement de la phase d'exploration du modèle. Nous préférons ici adopter la position de Le Moigne (1977) pour qui la modification de l'intention du modélisateur doit systématiquement conduire à une remise en cause de la définition du modèle (même si cette remise en cause peut conduire à adopter, par la suite, le même modèle). L'itération est donc toujours possible, à condition qu'elle ne néglige aucune des phases du cycle de vie, que ce soit pour le modèle initial (la surprise ne doit pas remettre en cause l'expérience en cours, même si elle peut modifier les expériences futures) ou pour le modèle itéré (puisque'il n'est pas concerné par l'expérience en cours).

5 Conclusion : la position paradoxale du modélisateur-expérimentateur

Ce premier chapitre a été pour nous l'occasion de faire le point sur l'environnement scientifique dans lequel nous plaçons nos travaux. Il nous a en particulier conduit à réfléchir aux "règles du jeu" de la modélisation individu-centrée, en particulier lorsqu'elle est appliquée aux systèmes biologiques complexes. Avant de passer à la suite de ce mémoire, c'est-à-dire aux modèles eux-mêmes, aux questions que nous leurs avons posées, aux réponses qu'ils nous ont données – ou pas – nous voudrions revenir sur deux points clés de ce premier chapitre.

En premier lieu, nous voudrions assumer pleinement les faiblesses de ce premier chapitre : les épistémologies croisées de la complexité et de la modélisation ne sont pas d'un abord facile et nous ne prétendons avoir apporté ici ni des idées originales, ni même des idées justes mais juste des idées. Il nous semble cependant que tout modélisateur se doit de réfléchir sur sa propre pratique, probablement comme tout scientifique, mais avec une difficulté supplémentaire liée au caractère à la fois réel et artéfactuel des modèles. Ce que nous avons exposé dans ce chapitre n'est que l'état de nos réflexions ici et maintenant. Nous espérons bien sûr que ces réflexions aient un minimum de sens mais, en tout état de cause, elles sont toujours "en devenir" puisqu'influencées à la fois par une pratique personnelle de la modélisation, par les questions que cette pratique pose, et par la recherche de réponses à ces questions. Elles sont donc ouvertes et soumises au débat.

Le deuxième point est probablement plus fondamental. Nous avons insisté à plusieurs reprises sur le caractère nécessairement pluridisciplinaire de la modélisation. Il est habituel, dans la plupart des communautés scientifiques, de revendiquer, dans un travail pluridisciplinaire, le rattachement disciplinaire des différents acteurs. Dans le cadre de la modélisation individu-centré de systèmes biologiques, et en particulier dans le cadre de la méthodologie pragmatique que nous avons exposée, les disciplines s'entrelacent tant qu'il devient selon nous difficile de les distinguer. À tout moment le modélisateur doit être ouvert aux termes des différentes disciplines : son objet n'est pas le modèle ; son objet est le modèle *et* l'objet modélisé, leur relation, leur "danse" et leur entrelacement. C'est à cet objet complexe ("lié avec") qu'il s'adresse et c'est ce que cet objet complexe "does to our mind" (Grimm, 1999) qui importe. Le modélisateur-expérimentateur doit pour cela manipuler son modèle, il doit s'intéresser à lui au delà de son seul aspect technologique ; il doit "jouer" avec lui. "L'activité ludique initiale est une condition essentielle de l'acte final de compréhension" dit Feyerabend (1975). Or, l'activité ludique, contrairement à la connaissance scientifique, ne se transmet pas. Celui qui comprend est donc celui qui a joué. Il n'est alors plus possible de se contenter d'une approche pluridisciplinaire. La démarche de modélisation individu-centrée implique de passer de la pluridisciplinarité à la transdisciplinarité¹.

Nous avons débuté ce mémoire par la question du lien. "Apprivoiser la vie", c'est créer du lien, mais lequel ? Nous pouvons déjà répondre partiellement à cette question. Lorsqu'il crée son modèle et l'apprivoise, lorsqu'il joue avec lui, lorsqu'il se laisse surprendre par lui puis lorsqu'il expérimente avec lui, le modélisateur ne crée pas le lien ; il est le lien.

*Caminante no hay camino, se hace camino al andar,
Golpe a golpe, verso a verso.*

Antonio Machado, "Cantares"

¹Nicolescu (1996) définit la *pluridisciplinarité* comme "l'étude d'un objet d'une seule et même discipline par plusieurs disciplines à la fois", ce qu'il convient de distinguer de l'*interdisciplinarité* ("transfert des méthodes d'une discipline à l'autre... (mais dont) la finalité reste aussi inscrite dans la recherche disciplinaire") et de la *transdisciplinarité* qui "concerne, comme le préfixe *trans* l'indique, ce qui est à la fois entre les disciplines, à travers les différentes disciplines et au-delà de toute discipline."

Chapitre II

Modèles

*L'essentiel, dans l'ivresse, c'est le sentiment de la force accrue et de la plénitude. Sous l'empire de ce sentiment on donne aux choses, on les force à prendre de nous, on les violente, – on appelle ce processus : **idéaler**. Débarrassons-nous ici d'un préjugé : idéaliser ne consiste pas, comme on le croit généralement, en une déduction, et une soustraction de ce qui est petit et accessoire. Ce qu'il y a de décisif c'est, au contraire, une formidable **mise en relief** des traits principaux, en sorte que les autres traits disparaissent.*

Friedrich Wilhelm Nietzsche, “Flâneries d’un inactuel”
(Robert Laffont, coll. Bouquins, 1993 – p. 995)

1 Introduction

Le premier chapitre nous a permis de prendre du recul sur notre pratique de la modélisation individu-centrée. En ce sens, il constitue, selon nous, la principale contribution de ce mémoire. Dans ce deuxième chapitre nous allons jeter un regard plus factuel sur les modèles effectivement produits, en présenter les objectifs, les principes et les résultats. Au cours de nos travaux, nous avons abordé trois domaines clairement différents, à savoir l'apprentissage de comportements, l'évolution des génomes bactériens et la modélisation cellulaire. La structure de ce chapitre est calquée sur ce parcours, avec trois sous-parties présentées dans un ordre chronologique.

Sans chercher à réécrire l'histoire – il est évident que la plupart des travaux présentés ici ont été réalisés avant la formalisation de la méthodologie développée au chapitre précédent¹ – nous adopterons pour les présenter une structure respectant le cycle de vie présenté au chapitre I. Pour chacun des trois domaines, nous détaillerons donc successivement la problématique, les modèles et les résultats en précisant l'état d'avancement dans

¹Celle-ci étant même en grande partie issue d'eux.

lequel se trouvent les modèles (certains d'entre eux étant "aboutis" tandis que d'autres sont encore en phase exploratoire). Il ne s'agit cependant pas ici de retracer dans les détails l'ensemble des travaux effectués. Aussi, pour chacune des parties, nous nous sommes imposé une contrainte de volume en nous limitant *a priori* à seize pages pour présenter l'ensemble du cycle de vie. Il est évident que cette contrainte nous a conduit à faire des choix qui se sont conclus par de nombreuses impasses. Nous espérons avoir conservé le plus important. Nous avons en particulier choisi de consacrer une large part de ces seize pages à la problématique. Il nous semble en effet que, pour un mémoire d'habilitation à diriger des recherches, celle-ci est au moins aussi importante que les modèles, voire que les résultats. D'une part les problématiques sont rarement exposées globalement (tandis que les modèles et les résultats sont présentés dans les publications et les communications), d'autre part elles sont, selon nous, plus pérennes : les modèles passent et les résultats se dépassent mais les problématiques restent. Il nous semble donc que c'est l'aptitude à bien cerner les problématiques qui constitue la première mission du directeur de recherche (ce qui ne signifie d'ailleurs pas que nous y soyons totalement parvenu). Sa deuxième mission est d'être capable de dégager des ouvertures à partir de ces problématiques. C'est pourquoi nous concluons chacune des sections thématiques par un exposé des perspectives de recherche. Celles-ci constituent autant de sujet de thèses ou de masters que nous pourrions proposer à la communauté.

2 Modèles de l'énaction

Une métaphore plus fluide s'insinuait en lui, produit de ses anciennes traversées marines. Le philosophe qui tentait de considérer dans son ensemble l'entendement humain voyait sous lui une masse soumise à des courbes calculables, striée de courants dont on eût pu dresser la carte, creusée de plis profonds par les poussées de l'air et la pesante inertie des eaux. Il en allait des figures assumées par l'esprit comme de ces formes nées de l'eau indifférenciée qui s'assaillent ou se relaient à la surface du gouffre ; chaque concept s'affaissait finalement dans son propre contraire, comme deux houles qui se heurtent s'annihilent en une seule et même écume blanche. Zénon regardait fuir ce flot désordonné, emportant comme des épaves le peu de vérités sensibles dont nous nous croyons sûrs. Parfois, il lui semblait entrevoir sous le flux une substance immobile, qui serait aux idées ce que les idées sont aux mots. Mais rien ne prouvait que ce substratum fût la dernière couche, ni que cette fixité ne cachât point un mouvement trop rapide pour l'intellect humain.

Marguerite Yourcenar, "L'Œuvre au noir"
(ed. Gallimard, 1968)

Résumé

Depuis le milieu des années quatre-vingt, une approche émergentiste (souvent qualifiée de "dynamique") du système nerveux s'est progressivement imposée. Selon cette approche, la compréhension du comportement d'un système vivant – et donc la compréhension de son système nerveux comme substrat de ce comportement – passe par la compréhension de la dynamique des interactions que ce comportement implique et dans lesquelles il est impliqué. Ce bouclage du comportement et de la dynamique caractérise l'approche dite "énactives" du comportement.

Dans ce contexte, nous nous intéressons ici à la dynamique interne d'un réseau de neurones formels, au couplage de cette dynamique avec une dynamique environnementale et à la modification de ce couplage par un algorithme d'apprentissage. Pour cela, nous définissons un modèle de réseau de neurones formels et un algorithme de modification des poids synaptiques, puis nous "embarquons" ce modèle dans un robot afin d'observer les dynamiques émergentes. Parallèlement, le modèle peut faire l'objet d'expériences complémentaires, par exemple en le plaçant dans des conditions dynamiques contraintes, de façon à identifier les effets de l'algorithme d'apprentissage sur la dynamique interne du réseau. En utilisant ce modèle, nous montrons qu'une loi d'apprentissage type "Spike-Timing Dependent Plasticity", appliquée sur des neurones à impulsion, provoque des transitions de phase dans la dynamique du réseau. En outre, lorsqu'elles se produisent dans le cadre d'une interaction bouclée avec un environnement stimulant, ces transitions de phases peuvent être comprises comme l'élaboration d'une structure comportementale.

2.1 Introduction

Comment une structure comportementale et perceptive stable émerge-t-elle de l'ensemble des interactions entre un système neuronal et son environnement ? Même si nous ne l'aurions probablement pas exprimée ainsi à l'époque, cette question est historiquement la première que nous ayons abordée au cours de nos travaux. Cette première approche avait cependant un objectif applicatif : il s'agissait de développer des architectures de contrôle permettant à un robot d'acquérir un comportement structuré dans un environnement partiellement inconnu.

Cette question résume à elle seule une grande partie du programme de recherche des sciences cognitives et soumettre la construction d'un robot (comportant nécessairement une obligation de résultat dans la réalisation d'une tâche) à la réponse à une question aussi fondamentale serait illusoire. Nous avons donc dû faire un choix entre le questionnement scientifique et les résultats applicatifs. Nous avons opté pour le questionnement, c'est-à-dire que nous avons travaillé sur les principes de l'architecture neuronale au prix des performances du robot. C'est ce choix qui nous permet de présenter ici ces travaux comme relevant d'une démarche de modélisation¹.

2.2 Problématique

Qu'on parle d'un robot, d'un animal ou d'un "animat" – un animal-robot² (Watts, 1998), la problématique à laquelle il doit faire face est similaire : il doit produire une séquence d'actions motrices, coordonnée par ses perceptions, en vue de parvenir à une satisfaction (qu'on appellerait, pour un robot, la réalisation d'une tâche). En d'autres termes, il doit produire un *comportement*. Si on se réfère à la définition posée au chapitre I, nous sommes bien en présence d'un système complexe. Le comportement est en effet issu de l'interaction entre un grand nombre d'entités ; on peut citer, bien évidemment, l'animat et les "objets du monde" mais aussi les "neurones" de cet animat, ou toute autre structure décisionnelle, voire l'ensemble de ses attributs corporels puisque le corps détermine en grande partie le comportement (Pfeifer et Bongard, 2006). En outre, ces entités sont bien en interaction dynamique et de cette dynamique va naître une structure stable – le comportement – dont l'apparence d'unité permet à l'observateur de lui attribuer un but.

Notre problématique, conforme au programme que nous avons défini pour les sciences de la complexité, est alors la suivante : il s'agit d'isoler ce qui, à l'échelle neuronale, permet l'émergence d'une structure comportementale stable à partir d'un ensemble de neurones en interaction entre eux et avec un environnement. Il s'agira donc de relier, par la modélisation, différents niveaux d'organisation en partant du substrat neuronal pour remonter à la dynamique des réseaux de neurones puis à la dynamique de l'interaction sensori-motrice (cette dernière étant issue du couplage de la dynamique neuronale et de la dynamique propre de l'environnement).

¹On peut y voir une entorse grave à la méthodologie que nous avons exposée au chapitre I. De fait c'en est une. Mais cette méthodologie est, entre autres, le fruit de mon propre parcours et les travaux évoqués ici ont été initiés bien avant la formalisation de la démarche.

²Dans la suite de ce chapitre, on utilisera systématiquement le terme d'animat puisqu'il s'agira de développer des robots modèles d'animaux. Nous maintiendrons ainsi volontairement l'ambiguïté propre à ce type de modèles.

Le dilemme structure/apprentissage

Si l'on excepte les modèles cognitifs développés dans les années soixante et soixante-dix et qui ne permettent pas d'aborder la problématique présentée ci-dessus (puisque le but – donc le comportement – y est défini *a priori* quand nous cherchons à le voir émerger), deux types de modèles abordent la question de la structure des comportements : l'approche connexionniste (Pomerleau, 1993; Webb et Consi, 2001) et l'approche comportementale ou "Behavior-Based Control" (Brooks, 1986, 1991). La première tente de simuler le substrat neuronal du comportement afin d'en comprendre (et/ou utiliser) les propriétés d'auto-organisation. La seconde considère le comportement comme issu d'un ensemble de règles sensori-motrices interagissant entre elles et avec un environnement stimulant (Murphy, 2000).

Il est "vital", pour un animat, d'être capable de structurer ses comportements. C'est en effet ce qui lui permet d'atteindre ses buts dans un environnement qui possède sa dynamique propre ou dont la dynamique va être perturbée (y compris par le comportement de l'animat lui-même). La modification de l'environnement perçu ne doit en effet pas, ou pas toujours, entraîner la remise en cause du comportement en cours (qu'on songe, par exemple, à l'âne de Buridan). Les comportements devront donc pouvoir (i) sélectionner les informations pertinentes pour eux (une information pertinente pour un comportement ne l'étant pas nécessairement pour un autre), (ii) être modifiés individuellement et (iii) s'exclure mutuellement. Or, l'approche connexionniste permet la mise en œuvre de mécanismes d'apprentissage mais, parce qu'elle est basée sur un parallélisme massif, elle ne permet généralement pas de faire émerger des structures intermédiaires modifiables individuellement. À l'inverse, le Behavior-Based Control, va, par construction, permettre de structurer les comportements, mais il ne permet pas (ou mal) l'acquisition de ces comportements. Cela nous a conduit à évoquer un "dilemme structure-apprentissage" (Beslon, 1995) et cela nous permet de préciser notre problématique : il s'agira de rechercher un modèle neuronal qui permette l'acquisition de comportements différents, stables et indépendants. En d'autres termes, la structure du comportement devra être modulaire.

Énaction et codétermination du comportement

Le modèle de comportement que nous proposons ici se place clairement dans une logique énactive (Varela *et al.*, 1992; Barbaras, 2002), c'est-à-dire qu'on considère qu'un comportement est une structuration de l'action qui (i) se met en place au cours de l'interaction entre l'animat et son environnement, (ii) modifie cette interaction et (iii) est modifiée par elle en retour. La question centrale de l'énaction est, en effet, celle de la co-construction, par un organisme et son environnement, d'un monde signifiant, c'est-à-dire d'un monde organisé *pour* et *par* l'organisme, d'un monde qui fait sens pour lui. Dans cette logique, la structuration du comportement n'est pas déterminée par la structure du système de contrôle mais, au contraire, codéterminée (Varela *et al.*, 1992) par les interactions entre un animat¹ (structuré), un contrôleur (structuré) et un environnement (structuré).

S'il est évident qu'un comportement s'inscrit dans la temporalité (c'est une *séquence* d'actions organisée), la codétermination implique que l'ensemble des acteurs de la boucle comportementale soient inscrits dans cette même temporalité. En effet, il n'est plus pos-

¹au sens d'entité dotée d'un corps, de moyens de perception et de moyens d'action.

sible d'attribuer à un élément particulier (animat, contrôleur, environnement) la structure du comportement puisque c'est par leurs interactions qu'elle se développe et se maintient. Cependant, l'environnement, s'il participe à la structure du comportement, va aussi être la source de nombreuses perturbations dont certaines vont être causées par le comportement lui-même¹. Le contrôleur devra donc être à même d'entretenir la régularité comportementale au sein d'une circularité qui, elle, tend à la détruire. En outre, c'est la régularité comportementale émergente qui permet l'accès à des relations de plus en plus complexes (par exemple en passant de la stabilité comportementale à la stabilité perceptive puis à la stabilité des représentations²). Du point de vue du contrôleur, la structure est donc une micro-émergence (voir chapitre I, section 3.4) qui implique la structure du comportement et est impliquée par elle. Pour qu'un couplage puisse s'établir, le contrôleur de l'animat doit donc être capable de se structurer dynamiquement à travers ses relations avec l'environnement. Mais il doit aussi tenir compte, pour l'apprentissage d'un comportement donné, des interactions avec les structures précédemment acquises puisque les régularités perçues ne doivent pas être comprises en tant que régularités environnementales mais en tant que régularités comportementales antérieures dans cet environnement³.

La modularité ascendante comme réponse à la codétermination

La codétermination du comportement nous invite à une approche alternative de la modularité. Dans les modèles neuronaux modulaires "classiques" (Ronco et Gawthrop, 1995; Sharkey, 1996), la modularité est généralement définie a priori (elle précède l'apprentissage). Nous proposons, en alternative à cette approche qualifiée de "descendante", une modularité "ascendante"⁴ dans laquelle la macrostructure émerge des interactions animat/contrôleur/environnement au cours de l'apprentissage.

Alors que la modularité descendante est basée sur l'intégration de modules locaux échangeant des informations via des "connexions" spécifiques (prédéfinies), l'approche ascendante repose sur un mécanisme de spécialisation au sein d'un réseau global. La structure émergente regroupe alors des modules aux contours flous sans qu'il soit possible de différencier qualitativement les connexions inter-modules des connexions intra-modules. Schématiquement, la différence entre les deux approches repose sur l'ordre d'apparition des niveaux hiérarchiques. En modularité descendante, les modules préexistent et sont agrégés pour construire un contrôleur intégré. Inversement, en modularité ascendante, le réseau global préexiste et se scinde en modules spécialisés au cours de l'apprentissage.

¹Ainsi, l'action de tourner la tête pour fixer un objet d'intérêt provoque un déplacement de cet objet dans le champ visuel donc la "perte" de cet objet.

²Dans le cas de l'exemple précédent, la stabilité de l'objet malgré le déplacement de la tête contribue à constituer une stabilité de la représentation de l'objet (Merleau-Ponty, 1942).

³En ce sens, cette approche se place dans une filiation typiquement "piagétienne". Dans son article "Jean Piaget" du dictionnaire des philosophes, Pierre Gréco écrit : "Les structures ultérieures seront engendrées par le fonctionnement lui-même, puisque ce fonctionnement transforme le réel. Dans le modèle proposé en 1957 (*Logique et équilibre*), l'idée centrale est que toute structure opératoire nouvelle est rendue possible (et mieux : probable) en fonction des résultats de la précédente. L'apparente finalité n'est que l'aspect externe d'une causalité séquentielle interne. (Piaget a perfectionné ce modèle en y marquant davantage les interactions entre le sujet et l'objet, chaque action du sujet faisant apparaître de nouveaux "observables", tant sur l'objet lui-même que sur l'action qu'on lui a fait subir.) (Gréco, 2001), p. 1215.

⁴Nolfi (1997) parle de modularité "proximale" par opposition à une modularité "distale".

L'approche ascendante de la modularité nous a conduit à adopter le principe d'une modularité "faible" c'est-à-dire d'une modularité sans module explicite. Dans cette approche, la modularité est portée par des groupes de neurones structurellement identiques mais associés temporairement à des fonctions distinctes (Tononi et Edelman, 1998; Varela *et al.*, 2001). De tels groupes fonctionnels – ou "assemblées" – de neurones sont donc des regroupements temporaires de neurones participant à un comportement donné. Ils permettent d'introduire un niveau de codage intermédiaire entre le substrat neuronal du comportement et le contrôleur global. De plus, ils permettent d'exprimer le comportement en termes de relations intra- et inter-modules tout en conservant une structure globale uniforme (les connexions étant toutes structurellement équivalentes mais fonctionnellement différenciables).

Le code neural, les assemblées de neurones, le temps

Souvent qualifiée d'"approche dynamique" (Guillot et Daucé, 2002), cette vision temporelle de la modularité du comportement est en grande partie issue de réflexions menées, dans les années soixante-dix et quatre-vingts, par différents auteurs¹ (Nagel, 1981; Braintenberg, 1986; Varela *et al.*, 1992). Au cours des années quatre-vingt dix, ces réflexions ont reçu une audience de plus en plus forte au fur et à mesure que les bases neurales du comportement commençaient à être élucidées. En effet, alors que le paradigme admis jusqu'alors était que l'information circulant sur une fibre nerveuse est codée par la fréquence de l'activité neuronale, il est apparu que le code neural, bien plus complexe, intègre les caractéristiques temporelles des signaux et transmet ainsi une information plus riche. Ainsi, des synchronisations temporelles entre neurones éloignés ont été détectées dans la plupart des aires cérébrales (Gray *et al.*, 1989; Singer, 1993; Varela *et al.*, 2001). Qui plus est, ces synchronisations semblent liées au caractère unifié des stimuli affectant les neurones incriminés (la synchronisation des décharges caractériserait donc l'appartenance à un même objet). Parallèlement, le concept d'assemblée de neurones, proposé par Hebb (1949) a commencé à recouvrir une certaine réalité, qui plus est une réalité dynamique. En effet, plusieurs auteurs ont proposé que ces assemblées de neurones correspondent à la participation transitoire à une dynamique particulière (Skarda et Freeman, 1987; Freeman, 1991; Meyrand *et al.*, 1991; Tononi et Edelman, 1998; Freeman, 2000; Huerta *et al.*, 2001; Varela *et al.*, 2001).

Toutes ces observations tendent à accréditer la thèse selon laquelle l'information est codée dans le cerveau sous une forme spatio-temporelle. Restait cependant à identifier un mécanisme neuronal permettant l'apprentissage d'une telle modalité. Un tel mécanisme a été mis en évidence à la fin des années quatre-vingt dix (Bi et Poo, 1998, 1999). Il a en effet été montré que certaines connexions synaptiques sont sensibles à l'écart temporel séparant les potentiels d'action pré- et post-synaptiques. Cette loi, la STDP (Spike-Timing Dependent Plasticity), permet donc de modifier la dynamique d'un réseau de neurones (en modifiant les couplages entre les neurones) en fonction de cette dynamique elle-même. Or, comme, au niveau supérieur, cette dynamique est elle-même couplée à la dynamique sensori-motrice, il devient envisageable d'identifier un mécanisme permettant, à travers

¹On pourrait lui trouver des origines plus anciennes, en sciences – émergentismes, structuralisme (Piaget, 1968; Andler *et al.*, 2002) – ou en philosophie, par exemple avec la phénoménologie (Merleau-Ponty, 1942, 1945)

tous ces couplages, de réguler la dynamique comportementale en agissant localement sur les connexions synaptiques par une loi de type STDP.

Conclusion : vers un modèle de couplage dynamique

Si les différents éléments présentés ci-dessus peuvent sembler disparates, ils convergent tous vers un principe unique : exprimer le comportement sous la forme d'un couplage entre la dynamique d'un réseau de neurones et la dynamique de l'interaction animat-environnement, ce couplage émergeant au cours de l'apprentissage du fait de la modification de la dynamique du réseau. Cela nous permet de proposer une série de questionnements suffisamment clairs pour développer des modèles individus-centrés. Ces questionnements sont les suivants :

Dynamique spontanée Quelle est la dynamique propre d'un réseau de neurones récurrent en fonction de paramètres tels que le type de neurones, la taille du réseau, la distribution des connexions ?

Dynamique et apprentissage Comment une loi d'apprentissage (par exemple la STDP) va-t-elle modifier les paramètres du réseau et donc sa dynamique spontanée ?

Dynamique couplée Suivant le régime dynamique propre du réseau, comment celui-ci va-t-il réagir aux influences externes ?

Modification du couplage En cas de couplage avec une dynamique externe, quels vont être les effets de la STDP ? Va-t-elle renforcer le couplage ou le contrarier ?

Du couplage au comportement La modification du couplage par la STDP est-elle suffisante pour permettre à un robot, piloté par un réseau de neurones, d'acquiescer un comportement structuré ?

2.3 Modèles et expériences

Nous avons travaillé sur cette thématique pendant plus de dix ans en particulier dans le cadre des travaux réalisés avec Anders Jonson, Jean Laurens, Hédi Soula¹, Thomas Tempé, Aravind Alwan, Frédéric Henri et Antoine Coulon. Cela a donc été l'occasion de développer différents modèles, dont la majeure partie ne seront qu'évoqués ici.

Exploration

Dès les premiers temps, nous avons cherché à établir un couplage dynamique entre un réseau de neurones récurrent et l'environnement de ce réseau. Nous avons en particulier cherché à établir ce couplage en permettant une propagation asynchrone de l'information dans un réseau de neurones, rejoignant en cela les travaux de Simon Thorpe sur la vitesse de propagation de l'influx nerveux (Thorpe *et al.*, 1996; VanRullen et Thorpe, 2001). Il manquait cependant à nos travaux plusieurs composantes fondamentales : le modèle neuronal était en effet trop spécifique et les algorithmes d'apprentissage n'étaient pas suffisamment dépendants du temps. Si ce modèle nous a permis d'obtenir quelques réussites sur l'apprentissage de comportements (Beslon *et al.*, 2000; Soula *et al.*, 2001; Beslon et

¹La collaboration avec Hédi Soula s'est étalée sur plus de cinq ans (dont trois ans de thèse). Une grande partie des travaux présentés ici ont été réalisés dans le cadre de cette collaboration.

Daucé, 2002; Beslon et Soula, 2003), nous l'avons progressivement abandonné au profit d'un modèle plus structuré et plus proche de l'état de l'art¹. On notera qu'une partie des travaux réalisés dans cette période ont utilisé les algorithmes génétiques pour optimiser les connexions neuronales (Soula *et al.*, 2003). Cela nous a conduit à travailler sur des populations de robots en interaction (Beslon *et al.*, 1998) et sur les relations de coopération/compétition avec ou sans apprentissage (Bournez, 2001; Bournez *et al.*, 2001; Beslon, 2002). Ces travaux nous ont été d'une grande utilité pour approcher les questions évolutives (voir section II.3). Dans les sections suivantes, nous présentons le modèle neuronal stabilisé tel qu'il a été utilisé dans nos travaux depuis plusieurs années.

Le modèle de neurone : LI&F Neuron

Le choix d'un modèle de neurone est difficile. Outre le nécessaire compromis entre précision et calculabilité, il est nécessaire d'intégrer dans le modèle toutes les composantes nécessaires à l'expression d'une dynamique suffisante. Historiquement, la plupart des modèles exploitent des neurones "fréquentiels" (le neurone est gouverné par la fréquence des potentiels d'action reçus). Cependant, ces modèles ne permettent pas d'exprimer toute la dynamique des neurones biologiques, en particulier en termes de synchronisation de phase. Or, de nombreux mécanismes cognitifs semblent exploiter cette synchronisation, soit pour le codage de l'information (Singer, 1993), soit pour l'apprentissage (Bi et Poo, 1998, 1999). C'est pourquoi nous nous sommes tournés vers des modèles plus fins, dans lesquels les potentiels d'action sont modélisés explicitement.

Nous avons utilisé le modèle "Leaky Integrate & Fire" (LI&F) qui a l'avantage de conserver la plus grande part de la dynamique de la cellule neuronale tout en étant calculable en un temps raisonnable (ce qui permet de simuler des réseaux de plusieurs milliers de neurones). Il modélise l'évolution du potentiel de membrane V_i (pour le neurone i , $i \in [1, N]$), celui-ci variant en fonction des influx reçus par le neurone et d'un courant de fuite (*leak*). L'équation II.1 décrit le comportement temporel du potentiel de membrane en fonction des différents paramètres du neurone (Maas et Bishop, 2001).

$$\begin{cases} \tau_m \frac{dV_i}{dt} = V_{rest} - V_i(t) + I(t) + \sigma X_i(t) + \sum_{j,k} w_{ji} \delta(t - T_j^k) \\ \langle X_i(t + \tau) X_j(t) \rangle = \delta_{ij} \delta(\tau) \end{cases} \quad (\text{II.1})$$

où $I(t)$ est un stimulus extérieur, éventuellement nul, V_{rest} le potentiel de la membrane au repos, τ_m la constante de temps membranaire, σ le bruit du potentiel ($X_i(t)$ étant un bruit blanc gaussien centré réduit), w_{ji} le poids synaptique du neurone j vers le neurone i , T_j^k l'instant de la k -ième décharge du neurone j , et δ un Dirac.

Lorsque le potentiel de membrane du neurone i dépasse un seuil θ_i , le neurone émet un potentiel d'action (une impulsion ou "spike") et le potentiel est réinitialisé à la valeur V_{reset} (équation II.2) :

¹Les premiers modèles possédaient déjà la plupart des qualités dynamiques nécessaires à l'apprentissage de comportements structurés. Cependant ils étaient non-recevables par la communauté des neurosciences computationnelles car ils ne se reliaient à aucun formalisme connu. Une partie des règles énoncées au chapitre I sont inspirées de ce constat.

$$\text{si } V_i(t) \geq \theta_i(t) \text{ alors } V_i(t) \rightarrow V_{reset} \quad (\text{II.2})$$

Les équations II.1 et II.2 modélisent la dynamique rapide du neurone, c'est-à-dire la dynamique d'émission des trains de spikes en fonction de l'évolution du potentiel de membrane. L'équation II.1 en particulier exprime, pour sa partie gauche, la variation spontanée de ce potentiel : lorsque le neurone ne reçoit aucune impulsion, le potentiel de membrane va décroître exponentiellement vers sa valeur au repos (V_{reset}), sauf si le neurone reçoit une entrée directe $I(t)$. En outre, le neurone est bruité (bruit diffusif). La partie droite de l'équation II.1 exprime quant à elle les interactions entre les neurones d'un réseau : lorsqu'un neurone émet un potentiel d'action (équation II.2), celui-ci est transmis aux autres neurones par l'intermédiaire de synapses qui vont pondérer le signal (qui, lui, est stéréotypé) ; w_{ji} représente la pondération de la synapse reliant le neurone i au neurone j .

Outre leur dynamique rapide, les neurones peuvent présenter une dynamique lente. Celle-ci va modifier la valeur des poids synaptiques (fonction d'apprentissage) et, dans certains modèles, le seuil de décharge θ_i . On parle dans ce dernier cas de mécanisme d'adaptation. Ce mécanisme correspond à un phénomène d'habituation, les neurones fortement stimulés étant de moins en moins sensibles, tandis que les neurones peu stimulés sont susceptibles de décharger très facilement.

Le modèle de réseau

Dans notre modèle, les neurones sont les individus et le "modèle de réseau" va nous permettre de déterminer l'influence que chacun d'entre eux va avoir sur les autres (à travers la matrice de connexions W qui contient les N^2 poids synaptiques w_{ji}). Toutes les expériences évoquées ici ont été réalisées avec un réseau de neurones totalement connecté (W est une matrice carrée non-symétrique). Lorsque le réseau est connecté à des entrées ou à des sorties, deux solutions peuvent être utilisées : soit des neurones d'entrée/sortie sont ajoutés au réseau, soit les entrées sont directement envoyées à tous les neurones du réseau via les valeurs $I(t)$ (équation II.1).

Le modèle d'apprentissage : Spike-Timing Dependent Plasticity (STDP)

La plupart des lois utilisées en modélisation sont inspirées de la "loi de Hebb" qui énonce que la synapse va être modifiée en fonction des corrélations d'activité entre les neurones pré- et post-synaptiques (Hebb, 1949). La notion de corrélation doit cependant être considérée différemment suivant le modèle de neurone (Gerstner et Kistler, 2002). Dans le cas des neurones fréquentiels, la corrélation est liée à l'activité simultanée des neurones (on a alors généralement $\Delta w_{ji} = \alpha x_i x_j$ si x_i et x_j sont les fréquences de décharge des neurones pré- et post-synaptiques). Dans le cas des neurones à impulsion, la notion d'activité *simultanée* n'a plus de sens puisque les impulsions sont de durée infinitésimale (dans le modèle). La "Spike-Timing Dependent Plasticity" (STDP) permet alors d'exprimer une fonction de corrélation $F(\cdot)$ en fonction du décalage temporel entre les spikes pré- et post-synaptiques (Bi et Poo, 1998, 1999).

Il existe plusieurs variantes de la STDP suivant le calcul de corrélation et la variation du poids synaptique induite. Il existe ainsi des STDP additives ou multiplicatives, symétriques

ou antisymétriques, positives ou négatives (Abbott et Nelson, 2000). La forme la plus classique est celle initialement proposée par Bi et Poo (1998). Elle correspond aux équations II.3 pour le calcul de la fonction de corrélation et II.4 pour le calcul de la variation des poids :

$$F(\Delta t) = \begin{cases} A_+ e^{\Delta t/\tau_+} & \text{si } \Delta t < 0 \\ -A_- e^{-\Delta t/\tau_-} & \text{si } \Delta t \geq 0 \end{cases} \quad (\text{II.3})$$

où Δt est le décalage temporel entre le spike pré- et le spike post-synaptique. A_+ , A_- , τ_+ et τ_- sont les paramètres de la STDP.

$$\Delta w_{ji} = F(\Delta t) w_{ji} \left(1 - \frac{w_{ji}}{w_{max}}\right) \quad (\text{II.4})$$

où w_{max} est la valeur maximale des poids synaptiques.

Dans les travaux présentés ci-dessous, nous avons parfois utilisé d'autres formes de STDP, soit pour des raisons de rapidité de calcul, soit pour induire des dynamiques spécifiques. Nous avons en particulier utilisé une forme inversée dénommée "ANTI-STDP" (Soula, 2005) et une forme symétrique gaussienne dénommée "STDP-GABA" (Woodin *et al.*, 2003).

Dispositifs expérimentaux

Nous appelons ici "dispositifs expérimentaux" les situations dans lesquelles nous avons placé nos réseaux pour estimer leur dynamique. Ils vont de la situation la plus simple (un réseau non bruité, totalement isolé, auquel on applique une stimulation initiale : $I(0) \neq 0$ et $I(t) = 0, \forall t > 0$) à la situation la plus complète où le réseau est embarqué dans un robot (réel ou simulé), connecté aux capteurs et aux moteurs, et soumis à une loi d'apprentissage (voir figure II.1). Plusieurs situations intermédiaires ont été testées, en particulier des stimulations par un signal périodique, afin d'explorer les propriétés de couplage du réseau avec ou sans apprentissage.

2.4 Résultats

À tous les stades de l'expérimentation sur le modèle de réseau de neurones, celui-ci nous a permis d'obtenir des résultats intéressants. Nous les présentons ici dans les grandes lignes. Pour plus de précisions, on pourra se référer à (Beslon et Soula, 2003; Soula, 2005; Soula *et al.*, 2005a, 2006; Soula et Beslon, 2006).

Dynamique spontanée et réponse aux stimuli

Nous avons pu confirmer l'existence de trois régimes de fonctionnement pour les réseaux en mode spontané. Le régime atteint dépend en particulier de la connectivité¹ du réseau : dans les cas extrêmes, l'activité s'éteint rapidement (connectivité faible) ou, au contraire, entre en régime synchrone (connectivité forte ; 50% des neurones déchargent à leur fréquence maximale, les autres étant muets). Entre ces deux extrêmes, le réseau est dans un état

¹La notion de connectivité recouvre des acceptions différentes suivant les réseaux considérés. Ici, la connectivité correspond à la distribution des poids synaptiques dans un réseau totalement connecté. Elle diffère donc de la notion de connectivité utilisée, par exemple, en théorie des graphes.



FIG. II.1 – Expérience d’évitement d’obstacles avec le robot Khepera. Ce dernier doit apprendre à utiliser le signal fourni par une caméra linéaire pour détecter les obstacles et les éviter (principe du flot optique). Les murs sont recouverts de bandes blanches et noires de largeur aléatoire. Cette expérience est inspirée de celle réalisée par Floreano et Mattiussi (2001).

complexe, asynchrone, proche du chaos. Lorsque le taille du réseau est suffisante (les expériences ont été conduites avec 100 neurones), on peut en particulier montrer une dépendance du régime de fonctionnement à la variance de la distribution des connexions : plus la variance est grande, plus le réseau va avoir tendance à entrer en mode synchrone (Soula *et al.*, 2006).

À ces trois modes vont correspondre différentes sensibilités aux entrées du réseau. Ainsi, si on soumet au réseau une entrée sinusoïdale (figure II.2) on constate que la réponse du réseau est fortement dépendante de son régime de fonctionnement. En régime faible (figure II.2, en haut), le réseau est totalement soumis à ses entrées. En régime fort (figure II.2, en bas), il ignore ses entrées et conserve son régime propre. Dans ces deux cas, le réseau ne peut pas être utilisé comme support d’un comportement organisé puisqu’il n’est pas soumis à la dynamique externe (dans le cas d’un réseau fort) ou pas suffisamment puissant pour conserver une dynamique interne (cas d’un réseau faible). Nous pouvons donc proposer une plage de fonctionnement correcte pour la mise en œuvre de comportements sensori-moteurs (Soula et Beslon, 2006). De fait, seul le régime intermédiaire présente une dynamique suffisamment riche pour répondre aux stimulations de l’environnement. Qui dit “couplage” dit en effet “influence mutuelle”, ce que ne permettent aucun des deux régimes extrêmes.

Influence de la STDP sur la dynamique spontanée

L’influence de la STDP sur un réseau de neurones LI&F est qualitativement prévisible. Elle agit en effet sur les connexions reliant des neurones dont l’activité est corrélée et, si $A_+ > 0$ et $A_- < 0$ (équation II.3) elle provoque l’augmentation, en valeur absolue, des poids des neurones corrélés (qui vont donc converger vers $\pm w_{max}$) tandis que les poids symétriques (donc entre neurones décorrélés) vont décroître et converger vers 0. Globalement, on peut donc prédire que la variance de la distribution des poids va augmenter. La STDP va donc provoquer une augmentation graduelle de l’activité du réseau, jusqu’à une transition de

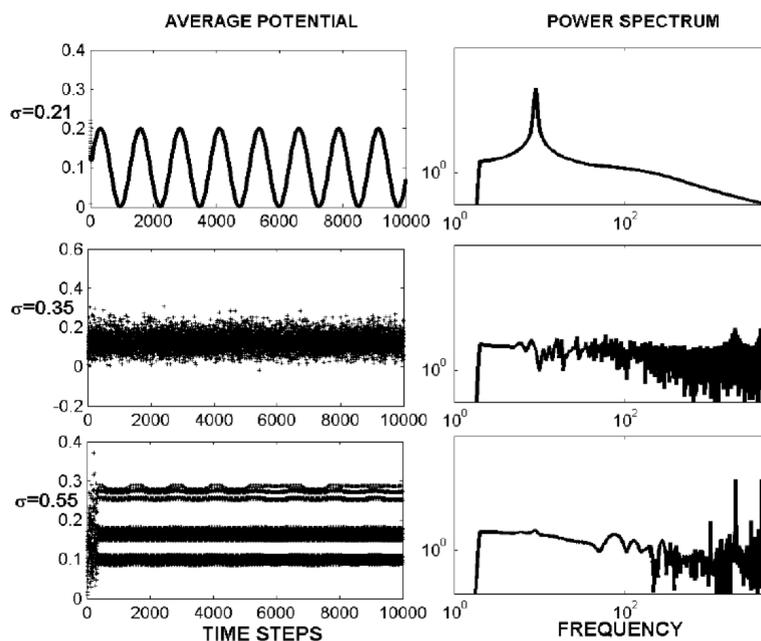


FIG. II.2 – À gauche : réponse moyenne d'un réseau de 100 neurones à un stimulus périodique. À droite : spectre de puissance de la réponse. En fonction de la variance de la connectivité, le réseau peut transmettre totalement le signal ($\sigma = 0,21$, en haut) ou ignorer le signal ($\sigma = 0,55$, en bas). Entre ces deux extrêmes, la dynamique du réseau est *influencée* par le signal ($\sigma = 0,35$, au milieu) (Soula et Beslon, 2006).

phase menant à la synchronie.

Nous avons vérifié en simulation cette propriété de la STDP qui, pour notre problématique, peut être assimilée à un apprentissage (figure II.3 en haut à gauche) puisque le réseau renforce progressivement sa dynamique spontanée jusqu'à la rendre totalement autonome (Soula, 2005; Soula et Beslon, 2006). Cet apprentissage est cependant inexploitable : nous avons vu (paragraphe précédent) qu'une fois que le réseau est passé en mode synchrone, il ne peut plus être le support d'un comportement organisé et réactif.

Nous avons testé d'autres formes de STDP, en particulier la loi ANTI-STDP (qui tend à décorrélérer les neurones corrélés). Contre toute attente, cette loi provoque qualitativement les mêmes effets que la STDP¹. Cependant, si ces deux lois agissent sur les mêmes connexions, elles provoquent des variations synaptiques opposées. Il est donc envisageable de les coupler, c'est-à-dire de les appliquer alternativement (avec des périodes longues – figure II.3, en bas à gauche – ou courtes – figure II.3, en bas à droite). Nous avons ainsi pu montrer que l'alternance des deux lois permet de provoquer des transitions réversibles entre les états synchrones et asynchrones (Soula *et al.*, 2005a).

¹A posteriori ce comportement est parfaitement compréhensible : du fait de la symétrie des variations synaptiques, STDP et ANTI-STDP provoquent toutes deux une augmentation de la variance de la distribution des poids.

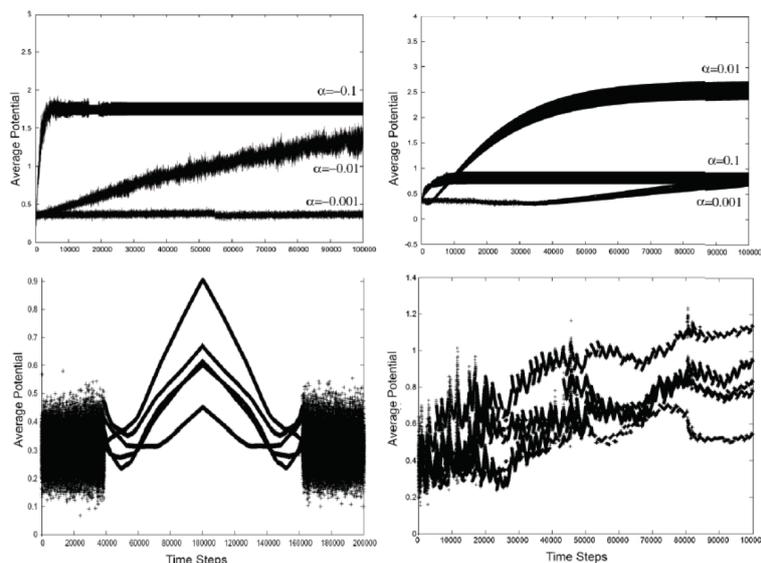


FIG. II.3 – Influence de la STDP sur la dynamique du réseau (représentée ici par le voltage moyen). **En haut à gauche** : augmentation de l’activité moyenne du réseau suite à l’application de la STDP. **En haut à droite** : augmentation moyenne de l’activité du réseau pour une ANTI-STDP. **En bas** : transition réversible par l’application alternée de STDP et ANTI-STDP (**à gauche** : transition à $t = 100\,000$; **à droite** : ou transition périodique de période 1000). D’après (Soula et Beslon, 2006).

Apprentissage de comportements

L’étude de la dynamique du réseau et de l’influence de la STDP nous permet de proposer une architecture cohérente avec les principes de l’énaction et de la co-détermination du comportement. Cette architecture repose sur l’existence, dans le réseau de neurones, d’une dynamique susceptible d’être influencée par les perceptions (le couplage ne doit donc pas être trop fort) et susceptible, en retour, d’influencer les perceptions (via la dynamique motrice qui va modifier la position du robot donc l’état des capteurs). Pour qu’un couplage puisse se mettre en place, il faut cependant que la dynamique interne réalise un filtre complexe sur la dynamique des perceptions (sans quoi le seul “comportement” possible est l’arc réflexe sans mémoire). C’est pourquoi il est nécessaire de maintenir le réseau dans un état intermédiaire, ce qui va lui permettre d’influencer la dynamique externe tout en étant influencé par elle (Soula *et al.*, 2005a).

L’apprentissage suppose que la dynamique interne du réseau puisse être modifiée. Or, nous avons vu que la STDP provoque systématiquement une transition vers un état synchrone, inadapté à notre objectif. En revanche, l’alternance STDP/ANTI-STDP permet de modifier la dynamique interne du réseau tout en maintenant une connectivité moyenne. Nous avons donc testé un modèle d’apprentissage basé sur cette alternance : dans un robot en interaction dynamique avec son environnement et gouverné par un réseau de 100 neurones LI&F, les deux lois sont appliquées suivant une alternance dépendant de l’interaction entre le robot et l’environnement¹.

¹Certains événements déclenchent une STDP, d’autres une ANTI-STDP. Malgré leurs dénominations, il est fondamental de ne pas voir dans ces deux règles une loi de “récompense” et une loi de “punition” :

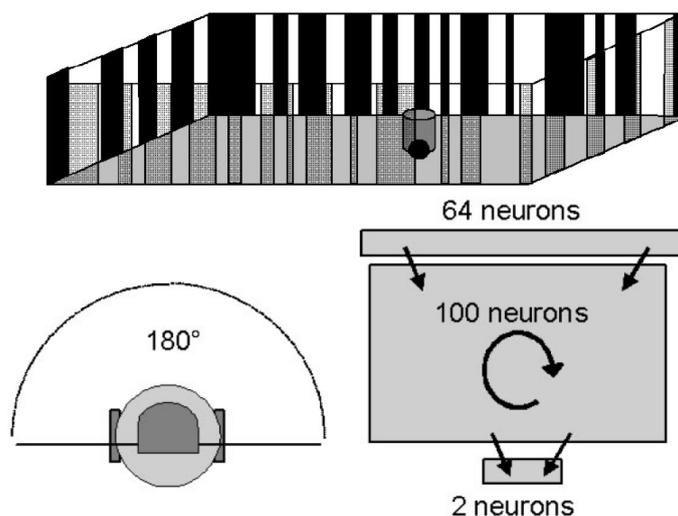


FIG. II.4 – Expérience d'évitement d'obstacles avec le Khepera. **En haut** : le robot (simulé) est placé dans une arène vide comportant des bandes blanches et noires sur les murs. **En bas à gauche** : le robot est équipé d'une caméra linéaire de 64 pixels et de deux roues motrices. Il est donc similaire au Khepera (voir figure II.1). **En bas à droite** : l'architecture neuronale comporte une couche de 64 neurones d'entrée (un par pixel), une couche cachée de 100 neurones LI&F et une couche de sortie de deux neurones reliés aux roues. D'après (Soula et Beslon, 2006).

Le protocole d'apprentissage est inspiré de Floreano et Mattiussi (2001). Les expériences ont été conduites sur robot réel (figure II.1) et en simulation (figure II.4). Durant l'expérience, le robot est placé dans une arène et contrôlé par le réseau de neurones pendant 100 000 pas de temps. La STDP et l'ANTI-STDP sont déclenchées sur des événements supposés signifiants. Dans notre cas il s'agit de la collision avec un mur ou d'un déplacement en ligne droite. Appliquées alternativement, ces deux règles vont modifier la dynamique du réseau jusqu'à trouver un point fixe – ou plus probablement un cycle limite – de la dynamique d'apprentissage. Considéré ainsi, l'apprentissage est donc la “recherche” d'une relation stable entre un animat et son environnement.

La figure II.5 présente le résultat du processus d'apprentissage au cours du temps. Le robot apprend bien à éviter les murs tout en se déplaçant. Il convient cependant de rester prudent. En effet, des simulations complémentaires nous ont montré l'extrême sensibilité du résultat au choix des événements signifiants déclenchant la STDP et l'ANTI-STDP¹.

Au delà de l'apprentissage de comportements, l'approche énaactive suppose l'émergence de symboles liés aux interactions dynamiques entre l'animat et son environnement : les catégories perceptives sont supposées liées aux invariants perceptuels, compte-tenu de la dynamique sensori-motrice. Dans le cas des expériences d'apprentissage menées avec le Khepera, nous avons observé l'activité moyenne du réseau de neurones en fonction de

toutes deux sont indissociables et leur couplage compose une même et unique loi d'apprentissage.

¹En termes de modèles, il n'y a pas, ici, contradiction : les événements signifiants peuvent être sélectionnés sur une base évolutive (Varela *et al.*, 1992; Di Paolo, 2003). En termes d'apprentissage automatique, cela rend difficile l'usage d'un tel mécanisme en robotique réelle, du moins en l'état, puisque nous n'avons pas d'autre moyen que l'expérience pour sélectionner les événements signifiants.

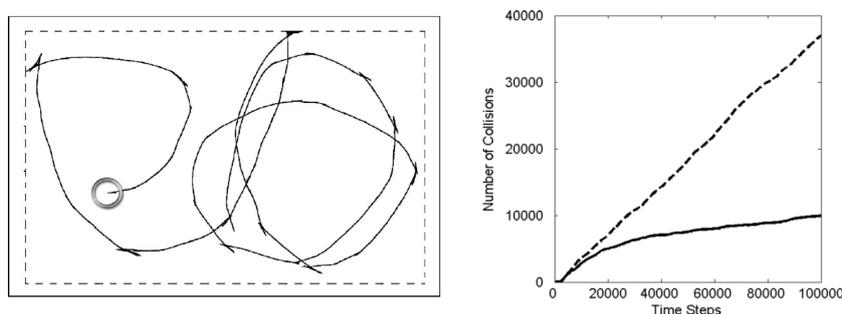


FIG. II.5 – Expérience d'évitement d'obstacles avec le Khepera. **À gauche** : trajectoire suivie par le robot après apprentissage. La trajectoire matérialise le déplacement du centre du robot. Le cadre pointillé correspond aux positions “interdites” (collision). **À droite** : cumul du nombre de collisions au cours de l'apprentissage avec STDP/ANTI-STDP (courbe pleine) et sans modification de la dynamique (en pointillés). L'apprentissage permet clairement au robot d'éviter les murs. Les courbes correspondent à des moyennes pour dix robots simulés. D'après (Soula et Beslon, 2006).

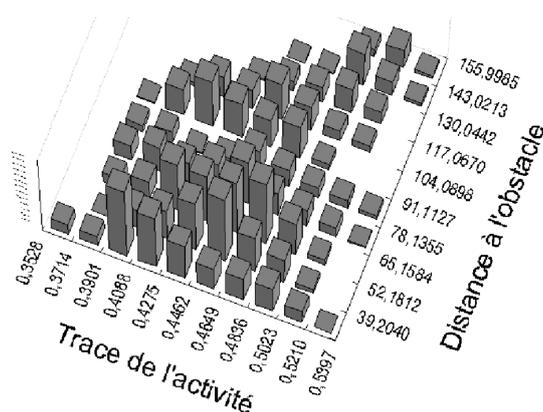


FIG. II.6 – Histogramme de l'activité du réseau en fonction de la distance. Plus la distance à l'obstacle est faible, moins le réseau est actif. Le réseau a donc “fait-émerger” une catégorie pertinente pour lui, alors que cette catégorie n'est jamais présente directement dans ses perceptions. D'après (Soula, 2005).

la distance au mur (figure II.6). Cette activité est clairement corrélée à la distance, ce qui montre que le réseau de neurones est parvenu à catégoriser son environnement. Il est cependant difficile d'aller au delà de ce premier constat : l'activité du réseau est en effet très complexe (d'autant plus que nous la maintenons volontairement “au bord du chaos”) et il est a priori impossible de déterminer de l'extérieur les catégories perceptives significantes pour ce robot dans cet environnement.

Influence de la STDP sur la réponse aux stimuli

Les résultats obtenus sur la dynamique des réseaux récurrents et sur leur propriété de filtrage (figures II.2 et II.3) nous ont conduits à étudier l'influence de la STDP sur le comportement d'un réseau de neurones en tant que filtre appliqué à un signal d'entrée.

Pour cela, le réseau est soumis à une entrée sinusoïdale d'amplitude α et de fréquence f variables ; soit $I(t) = \alpha \sin(2\pi ft)$. On observe alors la réponse du réseau à ce signal.

Ce travail a permis de confirmer des résultats connus sur les réseaux de neurones à impulsion (Mar *et al.*, 1999) et de montrer l'influence du codage en population sur la transmission du signal. L'analyse de la sensibilité du réseau au signal d'entrée a par ailleurs montré que le réseau se comporte comme un filtre passe-bas dont les caractéristiques dépendent de la connectivité (pour des connexions inhibitrices). Or, l'application de la STDP modifie la connectivité en fonction des caractéristiques du signal $I(t)$. La conséquence en est que la STDP va modifier les propriétés du filtre en fonction du signal filtré. Nous avons en particulier pu montrer que, dans le cas d'une STDP-GABA (Woodin *et al.*, 2003; Fino *et al.*, 2008), le filtre s'ajuste automatiquement au signal d'entrée de façon à le transmettre efficacement, et ce pour une large plage de signaux. Ce résultat pourrait être interprété en termes d'adaptation sensorielle¹.

2.5 Conclusion et perspectives

La rédaction de cette synthèse a été l'occasion de faire le point sur une thématique poursuivie depuis plusieurs années. La faible place consacrée ici aux travaux réalisés durant la première période souligne leur caractère "tâtonnant". Plusieurs années ont en effet été nécessaires pour que ce travail prenne réellement son essor. Pour autant, cette première période a-t-elle été inutile ? Nous sommes persuadés que tel n'est pas le cas. Le domaine des réseaux de neurones à impulsion appliqués à l'apprentissage de comportements est en effet particulièrement complexe et cette première phase était probablement indispensable pour parvenir à "naviguer" efficacement dans l'ensemble des concepts abordés.

En termes méthodologiques, nous n'avons pas abordé ici la complémentarité entre les simulations développées et les analyses mathématiques. Pour les travaux récents, nous nous attachons en effet, lorsque cela est possible, à valider les résultats obtenus par des modèles mathématiques. Nous sommes particulièrement redevables à plusieurs personnes (Olivier Mazet, Hédi Soula, Manuel Samuelidès et Antoine Coulon) d'avoir permis ces développements qui n'auraient jamais eu lieu sans eux. Par ailleurs, il est clair que les travaux exposés ici n'ont pas suivi la méthodologie proposée au chapitre I, sauf, peut-être pour les travaux les plus récents. Ainsi, nos liens avec le domaine d'expertise (les neurosciences) sont relativement lâches. Ce constat doit cependant être tempéré par les caractéristiques spécifiques du domaine de recherche. En effet, les "neurosciences computationnelles" se sont constituées en un champ de recherche autonome, relativement découplé de la neurobiologie. La raison en est probablement la difficulté qu'il y a à tisser des liens entre les modèles computationnels – tels que ceux que nous avons présentés ci-dessus – et les dynamiques neuronales "réelles".

En termes de résultats, les modèles développés permettent d'argumenter dans deux directions différentes. Sur la question de la dynamique neuronale, nous avons pu mettre en évidence l'action de la STDP sur les phénomènes de synchronisation et de filtrage. Nous pensons que ces travaux représentent un réel apport aux neurosciences (sans pouvoir cependant l'affirmer : conformément aux principes énoncés au chapitre précédent, ils

¹Ce travail a été réalisé en collaboration avec Antoine Coulon (alors en stage de master) et Hédi Soula. Il est en cours de publication.

devront être soumis à une revue d’audience “neurobiologique”). Sur la question de l’apprentissage de comportements et de l’énaction, les résultats obtenus sur le robot sont encourageants. Ils permettent de mieux comprendre les tenants et les aboutissants de la “co-détermination du comportement” en nous conduisant en particulier à nous détacher des modèles classiques de l’apprentissage de comportements. Ces modèles reposent en effet sur une classification *a priori* des “bonnes” ou “mauvaises” situations (Sutton et Barto, 1998). Dans les expériences que nous avons présentées ci-dessus, cette classification n’existe pas. Elle est remplacée par la notion d’éléments signifiants – nous devrions dire d’éléments signifiants primaires – sur lesquels l’apprentissage va se baser pour réguler la dynamique des interactions entre le robot et l’environnement (via la dynamique neuronale) et, à travers cette régulation, faire émerger des éléments signifiants de plus en plus complexes.

En termes de perspectives, nos travaux ouvrent trois perspectives majeures. La première est le prolongement des travaux sur la dynamique induite par la STDP dans un réseau de neurones. Il s’agit en particulier de caractériser la réponse d’un réseau à un signal suivant les différents types de STDP. La deuxième, qui bénéficierait directement de la première, est la poursuite des expériences d’apprentissage de comportements. En effet, nous avons vu que l’apprentissage est très dépendant du choix des éléments signifiants. Il est probable que cette sensibilité soit en grande partie liée au type de STDP utilisé et il serait particulièrement intéressant d’étudier des règles plus plausibles ou plus générales. Enfin, si l’approximation de connectivité totale est acceptable pour de petits réseaux, elle ne l’est plus dans les grands ensembles de neurones, d’autant plus que les réseaux de neurones réels présentent des schémas de connectivité spécifiques (Sporns *et al.*, 2005, 2007). Il nous semble donc particulièrement intéressant d’étudier les relations entre ces schémas, la dynamique des réseaux et la STDP. Pragmatiquement, il est malheureusement peu probable que nous poursuivions sur cette thématique. Il y a à cela plusieurs raisons. La principale est évidemment le manque de temps et la dispersion¹ mais des questions méthodologiques sont aussi en cause : nous sommes aujourd’hui plus enclins à concentrer nos efforts sur des questions pour lesquelles les liens noués avec les biologistes sont les plus forts.

Enfin, le peu de références faites ici à la robotique “applicative” peut surprendre. Après-tout, doter des robots de comportements organisés pourrait probablement conduire à des applications. Il faut cependant garder à l’esprit qu’un comportement est une relation non stéréotypée et que l’acquisition d’un comportement passe par un rapport à l’environnement plus que par à un rapport à la “consigne” ou au “programme”. Si l’acquisition de comportements est un jour utilisée dans des robots, alors il faudra envisager une approche de la programmation différente, peut-être plus proche du dressage que de l’informatique telle que nous la connaissons. Dans une telle perspective, il nous semble évident que l’application ne peut pas être l’objectif premier. Pour le dire avec les mots de Francisco Varela : “Il est clair que la stratégie de l’énaction n’est viable pour l’IA qu’à la condition d’oublier les contraintes opérationnelles de la résolution de problèmes spécifiques prédéfinis par l’ingénieur” (Varela, 1996).

¹Nous avons mentionné au premier chapitre la nécessité de suivre la “bibliographie du domaine”. Il ne semble pas raisonnable de suivre la bibliographie de trop de domaines différents.

3 Modèles de l'évolution

Si nous voulons que tout continue, il faut d'abord que tout change.

Giuseppe Tomasi di Lampedusa, "Le guépard"
(ed. Le Seuil, 1963)

Résumé

Afin d'explorer les liens entre évolution, robustesse et "évolvabilité", nous avons développé un modèle d'évolution expérimentale in-silico : aevol. Ce modèle est issu d'un double constat : d'une part, la robustesse et/ou l'évolvabilité d'un organisme sont liées – en partie au moins – à la structure de son génome (par structure, nous entendons les caractéristiques macroscopiques de la séquence génétique : taille, nombre de gènes, ordre des gènes, distances inter-géniques ...). D'autre part, il n'existait pas antérieurement de modèle d'évolution permettant d'étudier ces liens car les modèles d'évolution expérimentale in-silico ne présentaient pas une variabilité structurelle suffisante au niveau génétique. En outre, l'étude directe de ces questions, par exemple par des approches phylogénétiques, est difficile car il est impossible en pratique de différencier les pressions indirectes que sont la sélection pour la robustesse ou l'évolvabilité des pressions directes (par exemple une pression à la vitesse de répllication de l'ADN) ou de simples biais (par exemple un biais à la délétion).

La conception du modèle aevol respecte à la fois les principes généraux de la structure des génomes bactériens et de leurs principales sources de variation, le tout dans un contexte de sélection darwinienne. Dans aevol, les organismes sont composés de trois niveaux d'organisation : le génome, qui est une séquence circulaire double-brin, le protéome, qui est une collection d'éléments fonctionnels traduits des séquences géniques, et le phénotype, correspondant aux capacités fonctionnelles de l'organisme. C'est en particulier l'introduction du niveau "protéine" qui différencie clairement aevol des modèles d'évolution in-silico antérieurs. Ici, c'est la notion de protéine qui apporte conjointement la notion de gène et, in fine, la notion de structure du génome.

L'utilisation du modèle aevol dans des campagnes d'évolution expérimentale nous a permis de montrer que, en présence de réarrangements chromosomiques, la sélection indirecte de la robustesse conduit à une régulation de la taille des génomes suivant un principe proche de celui de l'"error threshold", mais s'appliquant aussi bien sur les séquences codantes que sur les séquences inter-géniques alors que l'"error threshold" classique ne s'applique que sur les séquences codantes. Ce premier résultat ouvre un grand nombre de perspectives et, si les plus immédiates ont déjà été exploitées, la plupart d'entre elles sont en cours d'exploration.

3.1 Introduction

La biologie est une science expérimentale. Mais elle est aussi, nécessairement, une science de l'évolution. Or, l'évolution échappe largement à l'expérimentation. La principale raison en est évidemment temporelle. Un processus évolutif se déroule sur plusieurs milliers de générations et, en évolution, une "expérience" nécessite plusieurs mois voire plusieurs années. La durée des expériences n'est cependant pas rédhibitoire et les projets d'évolution expérimentale sont aujourd'hui relativement courants (Elena et Lenski, 2003; Sanjuan *et al.*, 2004).

Les principales difficultés sont en fait liées à l'impossibilité d'associer un caractère à une cause évolutive précise. Face à un caractère biologique particulier, comment savoir en effet s'il a été sélectionné, s'il est apparu de façon purement contingente ou s'il est la conséquence d'un biais dans les mécanismes de mutation? Pour répondre à une telle question, il faudrait pouvoir étudier des organismes issus de processus évolutifs différents et indépendants¹. Il faudrait aussi être certain de ne pas perturber les équilibres fragiles qui ont conduit à l'émergence de cette structure. Il faudrait enfin connaître parfaitement l'organisme pour pouvoir remonter aux causes à partir des effets.

La modélisation individu-centrée permet de contourner ces difficultés en réalisant des expériences idéalisées dans lesquelles les conditions expérimentales sont parfaitement maîtrisées (Adami, 1998, 2006). Elle permet de répéter ces expériences, de contrôler les paramètres et de connaître le déroulement exact de l'évolution², le tout en un temps raisonnable, pouvant aller de quelques minutes (pour un modèle mathématique) à quelques jours (pour un modèle computationnel).

3.2 Problématique

Lorsqu'en 1973, Theodosius Dobzhansky écrit, dans *The American Biology Teacher* : "Nothing in biology makes sense except in the light of evolution" (Dobzhansky, 1973), il nous lance indirectement un défi scientifique : retrouver, derrière tout système biologique, une cause évolutive. On suppose souvent que celle-ci est nécessairement liée à un avantage sélectif et qu'un individu possède une fonction ou une structure parce que celle-ci lui a permis de devancer ses concurrents dans la "course à la survie". L'évolution est pourtant plus subtile : un avantage sélectif n'est pas toujours nécessaire et un caractère peut être fixé par simple dérive, pour peu que la taille de la population soit suffisamment faible. En outre, le processus évolutif est aussi soumis à des contraintes de robustesse : pour qu'un système biologique évolue, il faut déjà qu'il perdure. Or, le processus de variation indispensable à l'évolution n'a de cesse de détruire les systèmes qu'il a lui-même créés (Eigen, 1971). Ce "paradoxe d'Eigen" a ouvert la voie à une réflexion sur les relations entre robustesse et variabilité ; réflexion qui se prolonge encore aujourd'hui (Lenski *et al.*, 2006).

¹"So far, we have been able to study only one evolving system, and we cannot wait for interstellar flight to provide us with a second. If we want to discover generalizations about evolving systems, we will have to look at artificial ones." (Smith, 1992).

²"There are no missing links in the digital world.", R. Lenski – cité dans (O'Neill, 2003).

La sélection indirecte

L'évolution d'un organisme dépend de l'apparition de mutations, c'est-à-dire de variations sur lesquelles la sélection va pouvoir porter. Cela implique que les variations, pour être "visibles" par la sélection, doivent avoir un impact suffisant sur la valeur adaptative des individus : toute variation neutre (relativement à la taille de la population) va être soumise à la seule dérive génétique. Ce schéma suppose donc que tous les individus partageant la même valeur adaptative sont équivalents au regard des processus évolutifs.

Pourtant, à fitness égale, deux individus peuvent avoir des destins évolutifs différents selon la fidélité avec laquelle ils transmettront leurs génomes à leurs descendants (variation mutationnelle). Ainsi, si un individu reproduit son génotype plus fidèlement qu'un autre, ses descendants auront plus de chances de conserver les avantages acquis et la lignée aura plus de chances de perdurer (contrainte de robustesse). En revanche, il est alors peu susceptible de subir une mutation avantageuse et risque donc de disparaître au profit d'une lignée plus variable si cette dernière a la chance de subir une telle mutation (contrainte de variabilité). Pour qu'une lignée puisse perdurer, les individus qui la composent doivent donc parvenir à un équilibre entre robustesse et variabilité (Lenski *et al.*, 2006). Ces deux contraintes permettent à des organismes d'être sélectionnés sur la base de critères qui, en première instance, seraient neutres, d'où l'appellation de "sélection indirecte" (Tenaillon *et al.*, 1999; Ancel et Fontana, 2000).

En outre, si les organismes peuvent être sélectionnés sur la base de leur variabilité, il peut sembler logique que l'évolution sélectionne, à fitness égale, les organismes susceptibles d'évoluer "au mieux" (en supposant que ce sont ceux qui, par le passé, ont le mieux évolué). La sélection indirecte conduirait alors à l'apparition d'une "évolution de second ordre" (Worden, 1995) correspondant à la sélection, par l'évolution, des mécanismes évolutifs les plus efficaces (contrainte d'évolutivité¹).

Les mécanismes susceptibles d'influencer la robustesse, la variabilité et l'évolutivité d'un organisme sont extrêmement divers. Parmi les plus classiques, on peut citer les mécanismes de réparation de l'ADN qui déterminent en partie les taux de mutation (Baer *et al.*, 2007), ou l'apparition d'organismes hyper-mutants dans des conditions de stress (Taddei *et al.*, 1997). En fait, les variations étant sélectionnées en fonction de leur impact sur le phénotype, c'est l'ensemble de l'organisme qui est concerné par ces contraintes et il est particulièrement difficile de déterminer, parmi les différents éléments susceptibles d'intervenir, ceux qui vont effectivement avoir un impact sur la sélection indirecte et lequel.

En raison de la complexité des mécanismes mis en œuvre dans la sélection indirecte, c'est surtout par la modélisation que ces questions sont abordées. Ainsi, les questions liées à l'équilibre entre la variabilité et la robustesse peuvent être abordées via des modèles mathématiques. Cependant, dès qu'on s'intéresse aux causes ou aux conséquences de ces équilibres et non plus seulement à la valeur d'équilibre, ces modèles deviennent difficiles à manipuler. En effet, l'organisme ne peut alors plus être décrit par une ou plusieurs va-

¹La langue française n'a pas de vocable clair pour l'anglais "evolvability" (qui a d'ailleurs lui-même plusieurs acceptions). Nous le traduirons ici par "évolutivité". Alors que la variabilité correspond à la probabilité de subir une mutation, l'évolutivité correspond à la probabilité de subir une mutation avantageuse. S'il est relativement admis que la robustesse et la variabilité sont susceptibles d'être sélectionnées (Ancel Meyers *et al.*, 2005; Félix et Wagner, 2008), la question de la sélection de l'évolutivité reste une question ouverte en biologie évolutive (Kirschner et Gerhart, 1998; Earl et Deem, 2004).

riables phénoménologiques (sa “valeur de robustesse” ou sa “valeur de variabilité”) puisque ces variables sont liées à la structure de l’organisme lui-même. Il devient alors intéressant de se tourner vers des modèles individu-centrés (Adami, 2006).

Sélection indirecte et structure génétique

Nous nous intéressons ici à une caractéristique particulière des individus, susceptible de modifier leur robustesse et leur variabilité (et donc susceptible d’être indirectement sélectionnée), à savoir leur structure génétique. Par structure génétique, nous entendons le nombre, la taille et la position relative des gènes et des séquences intergéniques ainsi, éventuellement, que la présence de structures particulières telles que les opérons. En effet, avec la disponibilité d’un nombre toujours croissant de séquences génétiques, des régularités (ou au contraire des différences) interspécifiques apparaissent de plus en plus clairement dans la structure de ces génomes. Parce qu’elles ne modifient pas la fitness des individus (du moins en première approximation), certaines de ces régularités ne sont pas directement soumises à un processus de sélection. Or, il semble relativement évident qu’à fitness égale, la structure génétique d’un individu va influencer sur sa robustesse et sur sa variabilité. La duplication d’un gène, par exemple, peut entraîner un accroissement de la robustesse et ce, d’autant plus que les deux copies sont très éloignées sur le chromosome (ce qui limite le risque de perte simultanée). Malheureusement, les mécanismes par lesquels la structure d’un génome participe à la robustesse (et à la variabilité) sont extrêmement variés, mettant en jeu des équilibres complexes et jouant sur de nombreuses caractéristiques structurelles. Ainsi, dans le cas de la duplication d’un gène, on pourrait argumenter que la duplication conduit à l’apparition d’une protéine nouvelle (augmentation de la variabilité). On doit alors supposer que cette nouvelle protéine ne va pas avoir d’effet négatif – sans quoi la duplication du gène diminuerait la robustesse ! On le voit, un programme de recherche construit sur la base d’une telle intuition (et qui consisterait, par exemple, à rechercher, dans un grand nombre d’organismes, les positions relatives des gènes dupliqués) aurait toutes les chances de se heurter de front à la complexité des mécanismes impliqués dans la robustesse, la variabilité et l’évolutivité. C’est pourquoi il nous semble nécessaire d’intercaler, entre l’expérience de pensée et l’expérimentation *in vivo*, la modélisation.

3.3 Modèles et expériences

Afin de comprendre comment les mécanismes évolutifs ont pu structurer le génome des organismes, nous avons proposé le modèle d’évolution expérimentale *in silico* “aevol” (pour “Artificial EVOLution”). Dans cette section, nous allons décrire les principes généraux qui nous ont conduits à développer ce modèle. Pour une description exhaustive, nous renvoyons le lecteur vers le manuscrit de thèse de Carole Knibbe (Knibbe, 2006) qui décrit le modèle mieux que nous ne saurions le faire ici.

Aevol, un modèle individu-centré de l’évolution des génomes bactériens

En termes de modèles, aevol occupe une position intermédiaire entre les modèles moléculaires (décrivant des séquences génétiques réelles) et les modèles d’évolution artificielle (issus des algorithmes évolutionnaires) : comme les premiers, il comporte un génome “réaliste” (moins réaliste cependant que les modèles destinés à étudier les phénomènes d’accu-

mulation de mutations) mais comme chez les seconds, une fonction de fitness permet de calculer une valeur d'adaptation à partir du génome d'un individu. Ces deux ingrédients sont en effet indispensables pour notre étude puisque, à l'évidence, celle-ci demande que soient simultanément présents une structure génétique (suffisamment réaliste pour être rapprochée de la structure des génomes réels¹) et un mécanisme de sélection directe (la sélection indirecte ne peut en effet se manifester que sur la base d'une sélection directe qui viendra *in fine* sélectionner les descendants des individus).

La mise en œuvre simultanée d'un génome structuré et d'une fonction de fitness n'est pas sans poser problème. En effet, le génome devra pouvoir (i) être de taille variable, (ii) comporter un nombre de gènes variable, (iii) dans un ordre variable et (iv) séparés par des séquences non-codantes de taille variable. Le problème que nous avons à résoudre est alors de proposer un "mapping" génotype-phénotype qui autorise le génome à exprimer ces degrés de liberté tout en nous laissant la possibilité de calculer une valeur d'adaptation. L'idée qui a prévalu à la construction du modèle *aevol* est la recherche d'une proximité avec le "modèle biologique" là où cette proximité est nécessaire et, inversement, une exonération des contraintes de réalisme dès lors que le modèle s'éloigne de la question biologique. Dans notre cas, nous savons que le génome doit être suffisamment réaliste pour pouvoir être interprété en termes structurels. En revanche, nous ne souhaitons pas étudier spécifiquement le phénotype des organismes. Nous pourrions donc modéliser des organismes abstraits sans perte d'expressivité pour les questions qui nous intéressent. Il nous importe donc de déterminer, dans le mécanisme de traduction du génotype au phénotype (qui va être aussi une transition du réalisme à l'abstraction), à quelle(s) étape(s) nous perdrons le réalisme biologique. Il faut cependant garder à l'esprit que ce choix va en retour entraîner des contraintes sur le génome puisque seuls seront acceptables les génomes effectivement traduisibles. Ainsi, dans le cas des algorithmes évolutionnaires, l'absence de réalisme à l'échelle du génome n'est pas un choix *a priori* mais la conséquence d'un mécanisme de traduction bijectif (Lefort, 2007).

Un protéome dans un algorithme évolutionnaire.

Il ressort de ce qui vient d'être énoncé que le mécanisme de traduction devra être plus réaliste qu'une simple bijection génotype-phénotype. C'est pourquoi nous avons choisi d'introduire un niveau structurel intermédiaire, bio-inspiré : le protéome. Dans le modèle, son rôle principal sera d'être une structure de donnée indépendante du nombre d'éléments en présence et de l'ordre ces éléments (en termes algorithmiques il s'agit donc d'une "collection"). Le mapping génotype-phénotype devient alors un mapping génotype-*protéome*-phénotype dans lequel une première fonction transforme un gène en une protéine avant qu'une deuxième transforme une *collection* de protéines en un phénotype. Cette deuxième fonction assure donc la calculabilité du phénotype, quelle que soit la structure du génome. Il suffira pour cela qu'elle repose sur une opération de calcul commutative et associative. Le résultat est alors calculable quel que soit le nombre de gènes et indépendant de leur ordre. L'addition ou l'union, par exemple, présentent ces propriétés.

¹Cette contrainte nous interdit l'usage de modèles issus de l'algorithmie évolutionnaire. En effet, dans ces modèles, les génomes sont très contraints, ce qui entraîne de forts risques d'artefacts de simulation (on peut citer le cas de Burke *et al.* (1998) et Zeldovich *et al.* (2007) qui, à partir de modèles proches, obtiennent des résultats contradictoires mais tous deux artefactuels.).

À l'autre extrémité du processus de traduction, nous avons vu que nous pouvions manipuler des phénotypes abstraits sans perte d'expressivité du modèle. Nous avons choisi d'exprimer les individus dans un espace mathématique abstrait que, par facilité de communication, nous appelons "espace des fonctions biologiques". Un individu sera donc représenté par une fonction mathématique de $[0, 1]$ dans $[0, 1]$ exprimant sa capacité à réaliser une fonction "biologique" dans cet espace¹. Dans le modèle une protéine sera donc une entité, traduite d'un gène, capable de réaliser une (ou d'interdire la réalisation d'une) fonction biologique particulière².

Description du modèle aevol

Aevol est donc un algorithme évolutionnaire organisé en trois niveaux (quand les algorithmes évolutionnaires "classiques" n'en ont que deux), auxquels il faut rajouter le niveau "population" pour permettre la sélection. Dans la suite de cette section, nous présenterons les principes généraux sur lesquels reposent ces quatre niveaux.

Le génome. Le génome d'aevol est une séquence binaire circulaire double-brin. L'existence de deux brins complémentaires est indispensable car le positionnement des gènes sur les deux brins est une caractéristique génétique importante. La circularité permet quant à elle d'éviter les effets de bords.

Transcription/traduction. Dans aevol, les mécanismes de transcription puis de traduction reposent sur la recherche des séquences transcrites (basée sur un promoteur consensus) puis des séquences géniques situées sur le transcrit. Chaque gène est traduit en une "séquence protéique" qui est elle-même convertie en trois valeurs numériques (m , w et h) déterminant les caractéristiques fonctionnelles de la protéine³ (figure II.7).

Le protéome. Chaque protéine de aevol correspond à une fonction "biologique" particulière. Plus précisément, elle correspondra à une distribution de possibilités dans l'espace des fonctions biologiques. Cette distribution est un triangle isocèle défini par trois paramètres traduits de la séquence du gène (m la moyenne de la distribution, w la demi-largeur de la base du triangle et h sa hauteur⁴). Le protéome est donc une collection de triangles, éventuellement chevauchants, répartis sur l'axe des fonctions biologiques (voir figure II.9).

La "biochimie" du modèle. Des opérations ensemblistes permettent de calculer la distribution de possibilités du phénotype à partir du protéome : les fonctions d'un organisme correspondent à l'ensemble P des fonctions réalisées par les protéines de cet organisme, mais non inhibées par elles⁵.

¹Le formalisme mathématique utilisé est la théorie des possibilités, qui se rattache à la théorie des ensembles flous (Zadeh, 1978). Les opérateurs utilisés pour combiner l'ensemble des protéines et calculer le phénotype seront donc des opérateurs ensemblistes flous.

²Dans notre transition du réalisme du génome à l'abstraction du phénotype, nous avons donc un processus de traduction phénotype-protéome réaliste mais qui aboutit à un ensemble de protéines abstraites.

³La fonction de la protéine ne dépend donc que de la séquence du gène. À un même gène correspondra donc une même protéine qui réalisera une même fonction, quelle que soit la position du gène sur le génome. Le processus de transcription/traduction garantit donc à notre génome ses degrés de liberté structurels.

⁴ h pouvant être négative dans le cas de protéines inhibitrices réprimant une fonction biologique.

⁵Plus formellement, si les A_i sont les distributions de possibilités des protéines activatrices et les I_j celles des protéines inhibitrices, alors : $P = (\cup A_i) \cap (\cup \overline{I_j})$. Le formalisme utilisé étant la théorie des

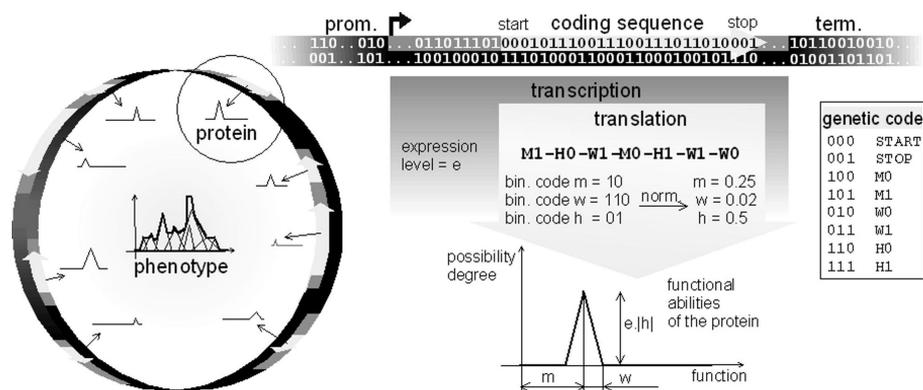


FIG. II.7 – Les étapes de transcription/traduction dans *aeol*. À partir de la séquence génétique circulaire (à gauche), les séquences transcrites puis les séquences traduites sont extraites (en haut). La séquence traduite est convertie en une séquence protéique grâce à un code génétique artificiel (à droite). La séquence protéique permet alors de calculer les paramètres de la distribution de possibilités fonctionnelles de la protéine (en bas). Une fois toutes les protéines extraites, elles sont combinées pour obtenir l'organisme complet (à gauche au centre du génome). D'après (Knibbe *et al.*, 2007a).

Le phénotype et la valeur sélective. L'ensemble P va permettre de calculer la valeur adaptative de l'individu. Pour cela, on fixe, pour une situation biologique donnée, l'ensemble des fonctions optimales E (correspondant à la contrainte posée par l'environnement sur l'évolution des organismes). L'adaptation d'un individu sera alors inversement proportionnelle à la différence entre P et E (appelée le "gap", g , de l'individu¹).

La sélection. À partir d'une population d'individus², chacun possédant sa valeur adaptative propre (en fait son gap g), la sélection permet de déterminer la probabilité de reproduction de chaque individu. *Aevol* permet :

- La sélection sur le gap. La probabilité de reproduction d'un individu dépend de son gap relativement à la somme des gaps. Cette sélection est biologiquement "réaliste" (elle modélise des ressources limitées) mais elle ne permet pas de maîtriser la pression de sélection (puisque celle-ci dépend de l'homogénéité de la population et que celle-ci varie au cours de l'évolution).
- La sélection sur le rang. La fitness d'un individu dépend linéairement de son rang dans la liste des individus (ordonnée par gaps décroissants). L'ensemble des rangs étant invariant, la pression de sélection est constante (mais faible et non modifiable).
- La sélection rang-exponentielle. Similaire à la précédente, elle est basée sur l'exponentielle du rang ce qui permet de faire varier l'intensité de la sélection (Whitley, 1989).

La reproduction. Dans *aeol* un individu se reproduira par copie de son patrimoine génétique. Le point clé est alors l'introduction de mécanismes de variation génétique.

possibilités, nous utiliserons les opérateurs de Lukasiewicz.

¹Aux ensembles P et E sont associées des distributions de possibilités f_P et f_E . On définit alors le "gap" : $g = \int_0^1 |f_E(x) - f_P(x)| dx$. Ce choix est probablement la plus grande des libertés que nous ayons prises avec la biologie évolutive. Dans *aeol*, l'évolution est en effet dirigée vers un objectif fixé *a priori*.

²Dans *aeol*, les populations sont de taille fixe et la durée de vie d'un individu est de une génération.

La structure génétique n'étant pas fixée *a priori*, nous pouvons utiliser des mécanismes mutationnels beaucoup plus variés que ceux qui sont généralement à l'œuvre dans les algorithmes génétiques. Nous disposons de trois types d'opérateurs de variation : les mutations ponctuelles (insertion, suppression, mutation d'une base à une position quelconque) qui n'ont qu'un impact local, les opérateurs de réarrangement qui affectent de larges segments chromosomiques (délétion, translocation, duplication et inversion) et la recombinaison qui permet d'échanger de l'ADN entre deux individus différents¹.

Une évolution typique dans *aevol*

Aevol permet de réaliser des campagnes d'évolution expérimentale sur plusieurs milliers de générations et d'analyser les structures génétiques obtenues en fonction des paramètres (taux de mutation, taux de sélection, taille de la population, etc.). Si ces structures peuvent être très variables, le processus évolutif est relativement stable d'une expérience à l'autre. On observe ainsi une amélioration rapide de la fitness des individus dans les premières générations puis un ralentissement progressif. On notera que la fitness n'est jamais totalement stable et que des mutations avantageuses se produisent régulièrement. Même à des stades avancés de l'évolution, la sélection reste donc directionnelle.

L'évolution des individus s'accompagne de profondes mutations de la structure de leur génome (figure II.8). Dans un premier temps, la taille du génome augmente fortement pour passer des 5000 paires de bases initiales (initialisation par défaut dans *aevol*) à plusieurs dizaines de milliers, voire plusieurs centaines de milliers de paires de bases en quelques générations. Dans le même temps, tous les paramètres structurels (taille des génomes, nombre de gènes, taille des séquences non codantes, ...) augmentent fortement. La deuxième phase se caractérise par une décroissance rapide de la taille du génome et du nombre de gènes, tandis que la taille des gènes, elle, continue de croître. Enfin, au cours de la troisième phase, la taille des génomes est stable. Par contre, l'organisme recommence à acquérir des gènes (mais plus modérément) tandis que la taille des séquences codantes augmente continûment.

Dans *aevol*, la fitness des organismes dépend de l'adaptation de leur phénotype à l'environnement. Pour augmenter leur adaptation, les organismes peuvent acquérir de nouveaux gènes et/ou améliorer ceux dont ils disposent déjà. Le déroulement de l'évolution nous montre que ces deux mécanismes ne sont pas exploités simultanément. Dans un premier temps les organismes augmentent rapidement la taille de leur répertoire génique, le plus souvent par duplication-divergence de gènes pré-existants. Ils sélectionnent ensuite les gènes les plus adaptés avant d'affiner leur répertoire génique en améliorant progressivement chacune de leurs séquences codantes².

3.4 Résultats

Nous ne présentons pas ici dans le détail les résultats obtenus grâce à *aevol* et nous invitons pour cela le lecteur à se référer aux travaux publiés dans le cadre de la thèse

¹La recombinaison est possible dans *aevol*, soit par échange de segments chromosomiques, soit par insertion d'un segment exogène. Cependant, le choix des points d'insertion étant aléatoire, elle est peu réaliste et n'a pas été utilisée dans les expériences menées jusqu'à maintenant.

²Dans *aevol*, la taille d'un gène est directement corrélée avec la précision des valeurs m , w et h ; l'augmentation de la taille des gènes permet donc d'affiner la position et/ou la forme des "protéines".

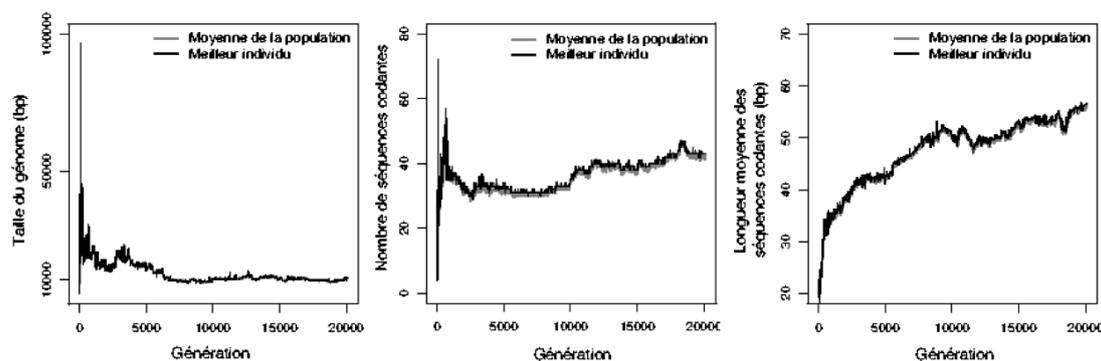


FIG. II.8 – Évolution des paramètres structuraux des génomes pour une exécution “typique” de *aevol*. **À gauche** : évolution de la taille du génome. **Au centre** : évolution du nombre de séquences codantes (nombre de gènes). **À droite** : évolution de la taille moyenne des séquences codantes. D’après (Knibbe, 2006).

de Carole Knibbe (Knibbe, 2006; Knibbe *et al.*, 2007b,a; Knibbe *et al.*, 2008). Nous nous contenterons de décrire les grandes lignes des conclusions auxquelles ce modèle nous a conduit et de montrer en quoi ces conclusions apportent des éléments de réponse à certaines questions ouvertes en évolution.

Régulation du nombre de descendants neutres

L’expérience la plus complète que nous ayons effectuée avec *aevol* est l’évolution, sur 20000 générations, de 72 populations indépendantes correspondant à trois répétitions pour six taux de mutation différents (de $u = 5.10^{-6}$ à $u = 2.10^{-4}$) et quatre intensités de sélection (en sélection rang-exponentielle). En observant les structures génétiques des 72 meilleurs organismes à la génération 20000, on constate qu’elles sont très variables, en particulier en fonction des taux de mutation (figure II.9). Ces individus présentent cependant tous une caractéristique commune : leur probabilité de se reproduire à l’identique $F_\nu W$ (produit de la probabilité de reproduction neutre¹ F_ν et du nombre de descendants W) est constante, quel que soit le taux de mutation (qui influe fortement sur la valeurs de F_ν) ou l’intensité de la sélection (qui fait varier W entre 1.25 environ pour une sélection faible et 10.0 pour sélection forte).

La régulation du nombre de descendants neutres est un résultat majeur du modèle. En effet, elle montre que, dans un système en évolution auquel on a laissé un certain nombre de degrés de liberté (ici la structure du génome), l’évolution va “exploiter” ces degrés de liberté pour ajuster la variabilité des individus². Nous sommes donc bien face à un mécanisme de sélection indirecte. En outre, la valeur de stabilisation est en elle-même remarquable puisque le nombre moyen de descendants neutres est très proche de 1, ce qui correspond à la valeur optimale pour conserver le phénotype courant et, dans le même temps, explorer au mieux l’espace de recherche de façon à découvrir des phénotypes plus adaptés. On peut

¹Cette probabilité est obtenue expérimentalement en effectuant 1000 reproductions de l’individu.

²La figure II.9 montre en outre que ces degrés de liberté peuvent, si les taux de mutation sont trop élevés, être exploités aux dépends de la fitness des organismes. On retrouve ici un résultat classique sur la sélection de la robustesse (Wilke *et al.*, 2001).

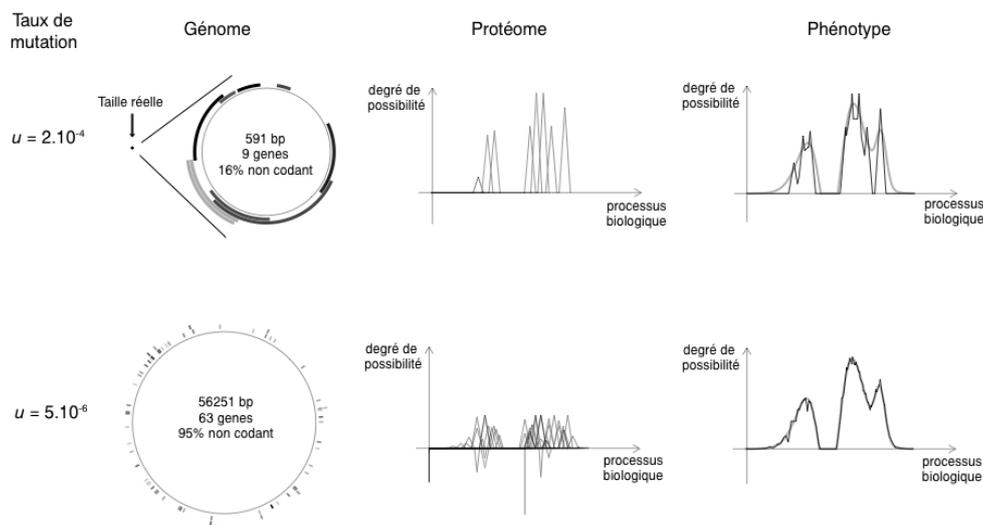


FIG. II.9 – Selon les taux de mutation, la structure des organismes après évolution est très différente. Sous un fort taux de mutation ($u = 2.10^{-4}$), les génomes sont très courts et très codants (en haut à gauche). Sous un faible taux de mutation ($u = 5.10^{-6}$), les génomes sont beaucoup plus longs et ils comportent de grandes zones non-codantes (en bas à gauche). Ces différences structurelles se répercutent sur les niveaux supérieurs : la complexité des protéomes est très différente (au milieu) et les phénotypes obtenus avec de forts taux de mutation (en haut à droite) sont plus éloignés de l'optimum (courbe continue à droite) que les phénotypes obtenus sous de faibles taux (en bas à droite). D'après (Knibbe, 2006)

en déduire que, non seulement la sélection indirecte a permis de sélectionner une robustesse minimale, mais qu'elle a aussi permis de sélectionner un compromis entre robustesse et variabilité (la sélection de la robustesse seule aurait conduit à maximiser $F_v W$).

Régulation de la taille des séquences non-codantes et du nombre de gènes

Les variations de structure chez les organismes ayant évolué dans des conditions différentes nous montrent que, pour maintenir un niveau de variabilité constant, l'évolution a sélectionné (indirectement) des génomes plus ou moins compacts. La figure II.10 nous montre ainsi que, dans aeol, une loi de puissance a émergé entre le nombre de bases non-codantes (ou le nombre de gènes) et le taux de mutation spontané. Cette loi de puissance est intéressante à plus d'un titre. D'une part elle permet d'expliquer comment les organismes arrivent à maintenir un taux de reproduction neutre constant, d'autre part, elle reproduit une observation réalisée par John Drake en 1991 et encore mal expliquée (Drake, 1991). De plus, le modèle permet d'explorer les causes de cette régulation.

Influence des réarrangements chromosomiques

Dans aeol le code génétique est non-redondant. Toute mutation affectant une région codante va donc avoir un impact sur la fitness des individus. Cette simplification nous

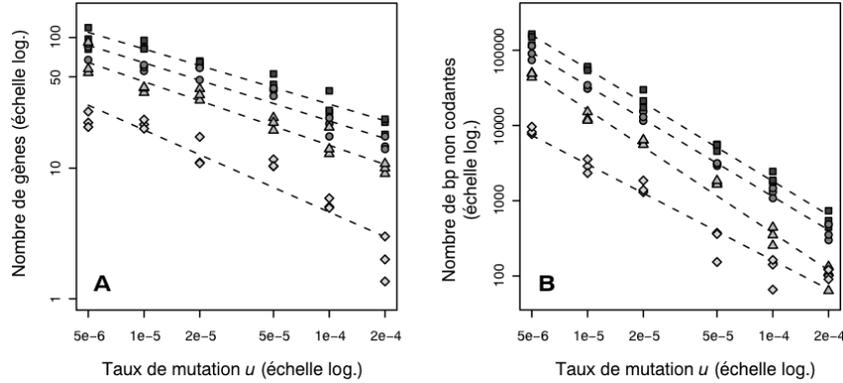


FIG. II.10 – Valeur finale du nombre de gènes (à gauche) et du nombre de bases non-codantes (à droite) en fonction du taux de mutation. Les symboles représentent des taux de sélection plus ou moins forts (les carrés correspondant à une sélection forte tandis que les losanges correspondent à une sélection faible). La loi de puissance est clairement visible sur les deux indicateurs. D'après (Knibbe, 2006).

permet de calculer la probabilité $\tilde{\nu}_i$ qu'une mutation de type i (mutation ponctuelle, insertion, délétion...) soit neutre (équation II.5) :

$$\left\{ \begin{array}{l} \tilde{\nu}_{\text{punct}} = \tilde{\nu}_{\text{ins}} = \tilde{\nu}_{\text{del}} = 1 - \frac{l}{L} \\ \tilde{\nu}_{\text{inv}} = \left(1 - \frac{l}{L}\right)^2 \\ \tilde{\nu}_{\text{transloc}} = \left(1 - \frac{l}{L}\right)^3 \end{array} \right\} \left\{ \begin{array}{l} \tilde{\nu}_{\text{gdel}} = \frac{1}{2L^2} \sum_{j=1}^{N_G} \lambda_j (\lambda_j + 1) \\ \tilde{\nu}_{\text{dup}} = \frac{1}{2L^2} \left(1 - \frac{l}{L}\right) \sum_{j=1}^{N_G} \lambda_j (\lambda_j + 1) \end{array} \right. \quad (\text{II.5})$$

où L est la longueur du génome, N_G le nombre de régions codantes, l leur longueur totale, et λ_j la distance en bp séparant les régions codantes j et $j+1$. On en déduit la probabilité qu'une reproduction soit neutre¹ (équation II.6) :

$$\tilde{F}_\nu = \prod_i (1 - u(1 - \tilde{\nu}_i))^L \quad (\text{II.6})$$

où i désigne le type de mutation et u le taux de mutation par bp.

L'analyse de ces équations montre que, du fait des grandes délétions (gdel) et des duplications (dup), la probabilité qu'une reproduction soit neutre dépend des valeurs λ_j (équation II.5), donc de la taille des séquences non-codantes. L'augmentation du nombre de bases non-codantes, qui provoque une augmentation du nombre de mutations, ne va pas être compensée – contrairement à l'intuition première – par une augmentation de la neutralité de ces mutations. Le modèle nous permet alors d'énoncer une règle relativement simple :

Si (i) il se produit de grandes délétions et de grandes duplications, si (ii) leur nombre augmente avec la taille du génome et si (iii) leur taille moyenne augmente avec la taille

¹On fait ici l'approximation qu'une mutation affectant une région non codante est toujours neutre.

moyenne du génome, alors l'accroissement des zones non-codantes du génome va conduire à l'augmentation du nombre de mutations délétères et, in fine, à une diminution du nombre de descendants neutres.

Cette règle simple permet d'expliquer les différents résultats énoncés jusqu'à maintenant : la régulation du nombre de descendants neutres serait due à un mécanisme de sélection indirecte de la taille des séquences codantes. En outre, cette règle est suffisamment générale pour pouvoir s'appliquer aux systèmes biologiques réels.

Tout est lié ...

Dans *aevol*, l'évolution sélectionne donc les organismes sur la base de leur capacité reproductive directe mais aussi sur la base de critères indirects tels que leur robustesse et leur variabilité. Or, nous avons vu que, pour un organisme donné, ces critères indirects sont susceptibles d'être influencés par un grand nombre de facteurs. Ainsi, il a été proposé que la structure du réseau métabolique influence la robustesse des organismes (Jeong *et al.*, 2001; Barabasi, 2004).

Même si cette relation est moins claire que la relation qui lie le nombre de bases non-codantes au taux de mutation, on observe dans *aevol* une tendance significative à la diminution de la taille du génome lorsque les mutations dans les gènes ont un impact plus fort (Knibbe *et al.*, 2007b). Cette tendance s'explique à nouveau par un mécanisme de sélection indirecte : plus l'impact des mutations dans les gènes est fort, moins les génomes sont robustes (les pertes de fitness d'une génération à la suivante étant plus importantes). La sélection indirecte va alors augmenter la robustesse en réduisant la taille des génomes. Ce résultat illustre une des caractéristiques importantes de la sélection indirecte. Parce qu'elle lie entre eux tous les éléments contribuant à la robustesse ou à la variabilité d'un génome ou d'un individu, la sélection indirecte entraîne des interdépendances entre toutes les échelles et tous les niveaux d'organisation d'un individu. Dès lors, la variation de taille d'un génome (par exemple) pourra être "liée" plus ou moins directement à la variation des taux de mutation mais aussi, éventuellement, à la variation structurelle d'un des nombreux réseaux contribuant au métabolisme cellulaire¹. Ces interactions entre niveaux d'organisation pourraient expliquer pourquoi on ne retrouve pas la loi de puissance "taux de mutation-taille du génome" lorsqu'on compare des organismes très différents (Drake, 1999; Lynch et Conery, 2003).

3.5 De Aevol à RAevol : évolution des réseaux de régulation génétiques

Une des limites majeures de *aevol* est l'absence de régulation génétique. Or, la régulation génétique est au cœur du fonctionnement cellulaire et de nombreuses questions sont ouvertes quant à la dynamique de ces réseaux, à leur structure ou à leur évolution. Les

¹Un exemple de telles interactions, telles qu'elles se manifestent, au moins, dans *aevol*, est la stabilité plus ou moins importante du génome en fonction de l'essentialité des gènes. Nous avons en effet constaté que, plus il existe de gènes essentiels, plus l'ordre des gènes tend à être conservé. Cet effet pourrait être dû à la sélection indirecte d'une structure génétique ayant optimisé l'impact moyen des réarrangements (Knibbe *et al.*, 2008).

premiers succès obtenus avec *aevol* nous ont donc tout naturellement conduits à étudier les possibilités d'extension du modèle dans cette direction en proposant le modèle "Regulatory-Aevol" : RAevol¹.

Un modèle pour quelle(s) question(s) ?

Les questions ouvertes sur les réseaux de régulation sont nombreuses et concernent souvent des problématiques très diverses. De plus, la construction de RAevol entre pleinement dans le cadre itératif dont nous avons dit (chapitre I, section 4.4) qu'il était incontournable mais dangereux. Il est donc crucial d'explicitier clairement la problématique afin de ne pas tomber dans le piège du "détournement de modèle" (rappelons qu'un modèle n'est un succès que relativement à une question particulière. Détourner le modèle avec pour seul motif son succès n'a donc aucun sens).

RAevol est destiné à étudier la dynamique évolutive des réseaux de régulation en lien avec les caractéristiques environnementales. Ainsi, nous souhaitons par exemple étudier la structure des réseaux en fonction de la stabilité de l'environnement. Il a en effet été proposé que les variations des conditions environnementales soient à l'origine de la modularité de réseaux (Kashtan et Alon, 2005; Wagner *et al.*, 2007). Nous rechercherons en particulier l'impact de la structure des réseaux sur la robustesse et l'évolutivité des organismes. Cela nous permettra de relier clairement les résultats obtenus avec RAevol et ceux déjà publiés dans le cadre de *aevol* (Knibbe *et al.*, 2007a; Knibbe *et al.*, 2008). Par ailleurs, RAevol permettrait d'isoler les motifs minimaux évolutivement plausibles et les comparer avec les motifs observés *in vivo* (Alon, 2007) ou avec ceux créés par la biologie synthétique (switch, oscillateur, etc. (Simpson, 2004)). Il permettrait alors de déterminer la part de déterminisme et de contingence dans l'apparition de ces réseaux et d'observer si, dans des conditions équivalentes, les réseaux évoluent systématiquement vers des structures similaires et pourquoi (Barabasi, 2004; Fox Keller, 2005).

De la régulation dans *aevol*.

Même privé de régulation, *aevol* comporte déjà une grande partie des structures nécessaires à la construction de RAevol. On peut citer, par exemple, les promoteurs, les ARNs ou les protéines. En outre, dans ce modèle, les protéines sont caractérisées par une séquence primaire (la séquence d'"acides aminés" issue de la séquence de codons du gène) et une fonction "métabolique". La mise en place d'un réseau de régulation demande donc "simplement" que trois nouveaux éléments soient introduits dans le modèle :

Des régions régulatrices Ces régions correspondent à des séquences d'ADN sur lesquelles les protéines régulatrices vont pouvoir se fixer temporairement et modifier le taux de transcription d'un promoteur. Dans RAevol, les régions régulatrices sont celles situées immédiatement en amont ou en aval des promoteurs².

¹RAevol est un modèle "en devenir". Il est développé dans le cadre de la thèse de Mlle Yolanda Sanchez-Dehesa. Nous considérons qu'il est aujourd'hui dans sa phase "deux" (exploration).

²Ce qui est une simplification grossière : on sait que des régions lointaines peuvent contribuer à l'activité de certains promoteurs. Néanmoins, dans le cas des procaryotes, la régulation est généralement plus simple et plus directe que chez les eukaryotes. Elle impliquerait, en particulier, moins de mécanismes globaux (Struhl, 1999).

Des protéines régulatrices Nous avons introduit, dans RAevol un “code de régulation” (en plus du code génétique qui convertit la séquence génétique en une séquence d’“acides aminés” et du code “fonctionnel” qui permet de calculer la fonction métabolique de la protéine) qui permet d’évaluer la capacité de régulation d’une protéine sur les différents promoteurs en fonction de sa séquence d’acides aminés et de la séquence de la région régulatrice¹.

Une dynamique cellulaire Une fois connues toutes les relations entre les protéines et les régions régulatrices, nous sommes en mesure de construire le réseau et de le simuler. Pour cela, on modélise mathématiquement la dynamique des protéines. La littérature est ici riche puisque c’est généralement à ce niveau que sont modélisés les réseaux de régulation (Mendes *et al.*, 2003). Nous nous basons pour l’instant sur un modèle inspiré de la fonction de Hill mais ce choix relève selon nous plus d’un paramétrage du modèle (donc à explorer) que d’une caractéristique figée.

Les premières expérimentations sont en cours et les premiers résultats sont encourageants puisque des mécanismes de régulation apparaissent de façon relativement robuste. Nous avons ainsi constaté l’apparition de “facteurs de transcription”, c’est-à-dire de protéines qui ne contribuent pas au métabolisme des individus mais qui, en revanche, participent activement au réseau de régulation (Sanchez-Dehesa *et al.*, 2007). En l’état, la transition vers la phase d’exploitation du modèle nous semble imminente et cette phase promet d’être relativement courte le paramétrage du modèle est aujourd’hui bien maîtrisés².

3.6 Conclusion et perspectives

Les résultats obtenus avec aevol ont largement dépassé nos espérances : nous cherchions à mettre en évidence des mécanismes de structuration des génomes et nous avons abouti à une pression évolutive nouvelle, agissant sur des structures jusque là supposées essentiellement soumises à la dérive. Qui plus est, ces résultats sont suffisamment génériques pour dépasser le seul cadre du modèle (Knibbe *et al.*, 2007a). En ce sens, de tous nos travaux, aevol est le plus avancé (au sens du cycle de vie du modèle³). L’histoire d’aevol n’est cependant pas terminée. En effet, la mise en évidence d’une nouvelle pression évolutive ouvre de nombreuses pistes.

Perspectives (1) : une pression évolutive parmi d’autres

Une perspective immédiate est d’étudier les interactions de cette pression nouvelle avec d’autres, plus directes. Ainsi, une question particulièrement intéressante serait l’étude des interactions entre la pression pour la robustesse (et/ou la variabilité) et les des biais

¹Sans entrer ici dans les détails de RAevol, le code de régulation est basé sur une matrice qui associe capacité d’interaction à chaque couple acide aminé/séquences génétiques. La capacité de régulation est alors calculée en alignant toutes les séquences d’acides aminés sur toutes les séquences régulatrices

²Il n’est cependant pas certain que le modèle RAevol soit réellement exploitable en raison des temps de calculs qui nous semblent être à la limite du raisonnable.

³Plusieurs des arguments développés dans le chapitre I sont d’ailleurs directement issus des difficultés que nous avons eu à faire reconnaître ce résultat. Ainsi, la possibilité de conduire des expériences complémentaires à la demande des biologistes (par exemple sur des mécanismes de sélection différents) mais aussi à exprimer le résultat du modèle indépendamment du modèle lui-même ont largement contribué à le rendre acceptable.

mutationnels directs. Ces biais sont en effet susceptibles d'influer fortement sur la quantité de non-codant (Petrov *et al.*, 2000; Petrov, 2002; Gregory, 2003, 2004). On peut alors s'interroger sur les influences relatives des pressions indirectes et des biais¹.

Cette première question nous renvoie naturellement au cas des organismes endosymbiotiques (*Buchnera*, *Rickettsia*, *Mycoplasma*, ...) dont les génomes sont particulièrement compacts (Gil *et al.*, 2002). En effet, les causes de la perte génique massive qu'ont subie ces organismes sont très mal connues (Andersson et Kurland, 1998; Gomez-Valero *et al.*, 2004; Lynch, 2006b). Outre le fait que nos résultats permettent de proposer un nouveau scénario évolutif, aevol pourrait être utilisé pour comparer différents scénarii. Ainsi, il a été proposé que les bottlenecks reproductifs soient à l'origine de la perte de gènes (Nilsson *et al.*, 2005; Ochman, 2005). Nous avons testé cette hypothèse dans aevol mais nous avons constaté un accroissement des génomes²! Dans une perspective voisine, nous projetons d'utiliser aevol pour tester différents scénarii susceptibles d'expliquer les différences de compacité des génomes mitochondriaux chez les animaux et chez les végétaux³ (Bullerwell et Gray, 2004; Bazin *et al.*, 2006; Lynch, 2006a,b).

Nous envisageons aussi d'utiliser aevol pour étudier les interactions entre la sélection indirecte due aux réarrangements et d'autres mécanismes de sélection indirecte. Des interactions complexes peuvent en particulier relier la taille des génomes avec l'évolution des taux de mutation ou l'évolution de mécanismes de variabilité phénotypique (Kussell et Leibler, 2005). Or, on sait que les taux de mutation sont susceptibles d'être sélectionnés (Bedau et Packard, 2003) ou de varier temporellement (Taddei *et al.*, 1997) ou spatialement sur le génome (Moxon *et al.*, 2006). De même, on sait que, à génome identique, la variabilité phénotypique peut constituer un avantage sélectif (Elowitz *et al.*, 2002; Veening *et al.*, 2008). Cependant, les interactions entre ces différents mécanismes sont actuellement totalement inexplorées. Nous avons développé plusieurs prototypes permettant d'étudier ces différents mécanismes (sélection des taux de mutation, mécanismes d'hyper-mutation, sélection de la variabilité) mais les campagnes d'expérimentation restent à conduire.

Enfin, il reste à déterminer toutes les conséquences de "notre" pression indirecte. La régulation de la taille des séquences non-codantes s'accompagne en effet de nombreux effets secondaires. Ainsi, nous avons observé une relation non triviale entre les taux de mutation spontanés ("pedigree rate") et les taux de mutation fixés ("phylogenetic rate"), les seconds étant inversement proportionnels aux premiers, au moins pour certains types de mutations. Si cette relation, se vérifie les conséquences en seraient surprenantes puisque les génomes les plus variables *a posteriori* seraient ceux qui subissent le moins de mutations⁴.

¹Des tests préliminaires conduits avec aevol nous ont ainsi montré que, dans certains cas, l'augmentation des seuls taux de duplication pouvait provoquer une réduction de la taille des génomes!

²Surprenante de prime abord, cette observation est cohérente avec nos résultats précédents puisque les bottlenecks entraînent une réduction de la taille efficace de la population et donc un besoin moindre en robustesse. Ces résultats sont cependant contraires à ceux de Nilsson *et al.* (2005) qui a constaté *in vivo* une réduction des génomes en présence de bottlenecks reproductifs.

³Étude en cours dans le cadre du stage de master de M. Gaël Kaneko.

⁴Ce résultat viendrait résonner étonnamment avec le "red-queen effect" mis en évidence par Van Valen (1973) qui énonce qu'"il faut courir de toute la vitesse de ses jambes pour simplement rester là où l'on est [et que] si on veut aller quelque part, ailleurs, il faut courir au moins deux fois plus vite ...". Dans notre cas, plus on court vite et plus on reste sur place!

Perspectives (2) : au delà de aevol

L'extension du modèle est une perspective évidente, même si nous nous refusons à parler "d'amélioration" (le modèle ne pouvant être évalué que relativement à une question). Nous avons déjà évoqué les travaux initiés avec RAevol. Deux autres extensions sont envisagées :

Une recombinaison plus réaliste. Telle qu'elle est codée dans le modèle, la recombinaison est trop artificielle pour être exploitable (les séquences génétiques étant choisies aléatoirement). Nous avons cependant toujours repoussé la mise en place de mécanismes de recombinaison basés sur des alignements de séquences en raison de leur coût algorithmique. Récemment, nous avons proposé une approche probabiliste de l'alignement et celle-ci s'est révélée relativement rapide. Disposer de mécanismes de recombinaison réalistes ouvrirait la voie à de nombreuses expériences, par exemple à propos de l'influence de la recombinaison sur le compromis robustesse/variabilité ou de l'effet de la recombinaison sur la structure des génomes¹.

Vers une "écologie" dans aevol. Dans le modèle actuel, les interactions entre les organismes sont limitées à la seule comparaison de gaps g calculés individuellement. Une modification du formalisme qui permettrait aux organismes d'évoluer dans un environnement partagé et éventuellement spatialisé. Cela nous ouvrirait l'accès à des questions particulièrement intéressantes telles que la spéciation, le parasitisme ou la symbiose. Les difficultés conceptuelles liées au partage de l'environnement sont cependant nombreuses et nous n'avons pas encore tous les éléments pour démarrer ne serait-ce que la phase 2 du cycle de vie (l'exploration).

Perspectives (3) : retour vers l'informatique

Les caractéristiques évolutives constatées dans avol peuvent-elles être exploitées dans le cadre des méta-heuristiques d'optimisation évolutionnaires ? La question mérite au moins d'être explorée (Banzhaf *et al.*, 2006). Aevol est, nous l'avons vu, capable d'adapter sa structure génétique. Exploitée dans une méta-heuristique, cette propriété permettrait d'envisager des algorithmes évolutionnaires capables d'adapter la complexité de leur réponse aux conditions "environnementales". Nous avons d'ores et déjà exploré cette possibilité dans le cadre de la thèse de Virginie Lefort (Lefort, 2007) avec le développement de l'algorithme "RBF-Gene". Nous avons ainsi pu montrer que cet algorithme permet une évolution itérative dans laquelle la solution au problème d'optimisation posé s'affine au cours des générations.

Il est important de noter ici que l'exploitation de mécanismes de robustesse dans le cadre de méta-heuristiques évolutives n'est pas nécessairement bénéfique. En effet, les mécanismes de robustesse identifiés jusqu'ici sont des mécanismes de robustesse mutationnelle. En d'autres termes, il s'agit d'une robustesse liée au processus évolutif lui-même. Dans le cadre d'une méta-heuristique, une telle propriété n'est pas toujours souhaitable puisqu'on cherche avant tout à optimiser la fitness des "organismes". Une réflexion nous semble donc aujourd'hui nécessaire pour déterminer les avantages et/ou les inconvénients que l'optimisation de la robustesse pourrait avoir dans ce cadre.

¹Ce travail a été confié à M. David Parsons dans le cadre de son stage de Master.

4 Vers des modèles cellulaires 4D

Quelque changeant que soit le contour de sa crête, cette vague [au pied des chutes du Niagara] est visible, approximativement au même endroit, et avec la même forme générale, depuis des siècles. Vue d'une distance d'un mille, elle semble être un monticule d'eau stationnaire. Vue de près, elle est une expression caractéristique des impulsions en tous sens nées d'un courant rapide de particules matérielles. Or, avec tous nos instruments, nous ne pouvons nous approcher, en quelque sorte, qu'à un bon nombre de milles de l'écrevisse. Si nous pouvions l'approcher de plus près, nous verrions qu'elle n'est que la forme constante d'un tourbillon de molécules matérielles, qui pénètrent constamment dans l'animal par un côté, et s'écoulent par l'autre.

Thomas Huxley "Traité sur l'écrevisse"¹

Résumé

Jusqu'à une période récente, l'interface entre la biologie cellulaire et la biologie moléculaire n'était réalisée qu'au prix d'une hypothèse d'homogénéité des constituants moléculaires au sein du système cellulaire (ou au sein d'un compartiment cellulaire). Cependant, la progression des outils d'observation montre aujourd'hui que cette hypothèse est loin d'être vérifiée et, surtout, que la structuration spatiale de la cellule est beaucoup plus complexe que prévu, y compris dans des compartiments cellulaires jusque là supposés relativement homogènes. Une des découvertes les plus déroutantes, pour le cadre de lecture "classique", est la très forte dynamique de certains compartiments cellulaires puisqu'ils semblent échanger en permanence, et de façon très rapide, leurs constituants avec le milieu "extérieur". Afin de permettre la compréhension de la dynamique de ces structures cellulaires, et d'approcher ainsi leurs fonctions, nous nous proposons de développer des modèles cellulaires individu-centrés, dans lesquels les molécules (ou, du moins, certaines d'entre elles) sont modélisées explicitement mais de manière très simplifiée afin de permettre le passage à l'échelle. L'objectif est de permettre à ces modèles d'approcher les phénomènes de structuration spatiale à l'échelle mésoscopique. En effet, notre hypothèse est que les mécanismes de structuration cellulaire (ou, du moins, certains d'entre eux) dépendent à la fois des propriétés locales des constituants moléculaires et de propriétés globales telles que l'agitation thermique dans le milieu cellulaire. Nous appliquons ces modèles à l'étude de structures cellulaires caractéristiques, les corps nucléaires et les domaines membranaires, ainsi qu'à l'étude des complexes de transcription. L'utilisation de modèles mésoscopiques nous permet de proposer des hypothèses d'auto-organisation pour expliquer la dynamique de ces structures, mais aussi pour mieux cerner leur rôle dans la cellule.

¹Cité dans : J.-M. Lévy-Leblond, *Aux Contraires, l'exercice de la pensée et la pratique de la science*, Gallimard, 1996.

4.1 Introduction

Le projet de la biologie des systèmes contient en filigrane l'idée de comprendre comment la cellule émerge, avec son individualité propre (son "apparence d'autonomie"), de l'ensemble de ses constituants moléculaires (Di Ventura *et al.*, 2006). Devant les difficultés d'un tel projet, plusieurs approches se sont structurées, allant de la collecte exhaustive (si tant est qu'elle soit possible) de l'ensemble des propriétés des constituants (identification de tous les composants moléculaires de la cellule et de leurs interactions, en général deux à deux) à la reconstruction *in silico* d'une cellule virtuelle. En pratique d'ailleurs ces deux approches extrêmes se rejoignent puisque, devant l'immensité et l'hétérogénéité des données récoltées, la cellule virtuelle peut devenir un moyen d'intégration de l'ensemble des données collectées.

Depuis la fin des années quatre-vingt dix, ces deux approches ont connu un essor considérable, soit par la mise en place de campagnes de collecte de données à très grande échelle (sur des organismes "modèles"), soit par le lancement de projets pharaoniques visant à simuler une cellule complète, généralement *Escherichia coli* ou *Mycoplasma genitalium* (Tomita, 2001). Selon nous, ces deux approches souffrent d'un même manque de modestie. Elles supposent en effet que sont connus tous les principes organisationnels par lesquels les constituants vont "produire" la cellule. En d'autres termes, elles supposent que nous ne manquons "que" des informations, voire "que" d'un moyen de les intégrer dans un modèle unifié¹ (la cellule "virtuelle"). Or, même si de grands progrès ont été réalisés ces dernières années (grâce, en particulier, aux nouvelles techniques d'imagerie fluorescente), nous en sommes encore réduits à émettre des hypothèses très générales quant aux mécanismes structurants permettant de passer de l'échelle moléculaire à l'échelle cellulaire. Force est de constater en effet que nous n'avons quasiment aucune idée de ce qui permet de passer du "tourbillon de molécules matérielles" que constitue la cellule à l'échelle microscopique à la totalité organisée, apparemment stable et régulée, que nous observons à l'échelle macroscopique.

Notre objectif est ici de développer des modèles qui permettent, par l'exploration, de comprendre les mécanismes structurant *de et dans* la cellule. Nous verrons que ce projet diffère radicalement d'un projet de "cellule virtuelle". En effet, il ne s'agira pas ici de construire des objets dont la complexité approche celle de la cellule mais, au contraire, de construire des objets simples – ou du moins le plus simple possible – qui s'organisent selon des principes biologiquement réalistes. À nouveau, il s'agira donc ici de "créer du lien" ; lien entre les différents niveaux d'organisation de la cellule, bien sûr, mais aussi lien entre différentes disciplines. Le niveau "micro" (la molécule) relève en effet clairement de la science physique (ou biophysique), tandis que le niveau "macro" (la cellule ou l'organite) est un objet biologique qui, lui, échappe aux outils de la physique.

4.2 Problématique

Avec le développement de la biologie des systèmes, un paradigme s'est largement imposé qui englobe aujourd'hui l'ensemble de la modélisation cellulaire : le "réseau". Il est en effet

¹En quelques sorte, et quoi qu'elles s'en défendent, ces approches sont donc particulièrement réductionnistes puisqu'elles supposent qu'il suffit d'intégrer l'ensemble des constituants sur la base des principes organisationnels connus à l'échelle des constituants (donc, ici, la biochimie), pour expliquer la cellule.

aujourd'hui commun de parler d'une cellule en termes de "réseau métabolique", de "réseau génétique", de "réseau de protéines", quand ce n'est pas en termes de "réseau de réseaux" supposé unifier l'ensemble des niveaux de description de la cellule. Fort logiquement, ce paradigme est accompagné de langages de description et d'outils de modélisation adaptés, qu'il s'agissent d'outils mathématiques (le plus souvent des équations différentielles couplées ou des modèles stochastiques) ou d'outils informatiques tels que les réseaux de pétri ou les réseaux booléens¹. Si ces approches "en réseau" permettent de développer une vision systémique de la cellule et, ainsi, de se libérer du carcan réductionniste consistant à étudier les constituants moléculaires individuellement, elles ne permettent pas de saisir une des caractéristiques majeures de la cellule biologique, à savoir son organisation spatiale. Or, si la biologie moléculaire a longtemps reposé sur une hypothèse d'homogénéité spatiale, celle-ci est aujourd'hui battue en brèche, que ce soit par le développement de nouvelles méthodes d'observation (Lippincott-Schwartz *et al.*, 2001), par le fait qu'un grand nombre de pathologies (en particulier les maladies dégénératives) trouvent leur source dans la perturbation de la structure spatiale de la cellule, ou par la découverte relativement récente que certains constituants moléculaires de la cellule sont présents en une très faible concentration². Ce dernier point a rapidement été pris en compte et plusieurs modèles ont ainsi vu le jour qui incluent les effets stochastiques imposés par le faible nombre de composants (Becskei *et al.*, 2005). Cependant, ces modèles conservent l'hypothèse d'homogénéité la plus forte, à savoir que la probabilité de présence d'une molécule ne dépend pas du point de l'espace considéré (c'est-à-dire que l'environnement est considéré "parfaitement mixé"). Or, dans de nombreux cas, cette hypothèse ne permet pas de décrire correctement le fonctionnement cellulaire qui apparaît fortement compartimenté. Il est évidemment toujours possible d'inclure cette compartimentation dans les modèles (l'homogénéité n'étant alors supposée qu'au sein d'un compartiment particulier) mais dans ce cas l'organisation cellulaire est plaquée sur le modèle et il n'est pas possible d'étudier son émergence. C'est pourquoi l'étude des mécanismes structurant à l'échelle cellulaire demande que soient développés des modèles "4D" (c'est-à-dire des modèles spatiaux *et* dynamiques) à l'échelle microscopique (c'est-à-dire, ici, à l'échelle moléculaire). De tels modèles sont en effet les seuls à permettre l'étude des mécanismes de structuration cellulaire, de leur dynamique et de leurs interactions avec le niveau fonctionnel (Bork et Serrano, 2005).

Nous avons évoqué, au chapitre I, les phénomènes émergents et nous avons supposé implicitement que les systèmes biologiques sont, de fait, de tels systèmes émergents. Il importe cependant de préciser ici que ce n'est pas cette caractéristique qui, à elle seule, implique l'utilisation de modèles dynamiques "4D" pour l'étude de la cellule³. Ce qui, ici, impose l'utilisation de modèles microscopiques c'est le couplage dynamique entre le niveau microscopique et le niveau macroscopique émergent, c'est-à-dire le fait que les temps caractéristiques des deux niveaux se chevauchent. En effet, dans un système global (par exemple le réseau routier) émergeant d'un ensemble de constituants locaux (le trafic automobile⁴),

¹Pour une introduction à ces différents outils et à leur application à la biologie cellulaire, on se reportera à (Roux, 2007) et, plus particulièrement à (Kepes, 2007)

²Il n'est alors plus possible de supposer que ces constituants sont répartis de façon homogène dans la cellule, ni même dans un compartiment cellulaire.

³D'autant plus que, comme l'énonce Morin (1977), "tout état global présente des émergences". Il est évident – et heureux – que tout état global ne nécessite pas, pour être modélisé, un modèle microscopique émergent.

⁴Nous réalisons ici l'hypothèse simplificatrice que les aménagements routiers sont décidés en fonction

les deux niveaux peuvent être modélisés indépendamment si leurs dynamiques sont clairement découplées : dans le cas du réseau routier le temps caractéristique du trafic est de l'ordre de l'heure tandis que le temps caractéristique de modification du réseau est de l'ordre de l'année. On peut alors modéliser le trafic (micro) en supposant le réseau (macro) stable, ce dernier pouvant, lui, être modélisé indépendamment de la dynamique des acteurs locaux (on n'utilise alors que des données agrégées telles que le trafic moyen ou le trafic maximum). Au contraire, si la dynamique globale ne peut pas être découplée de la dynamique locale (ce qui, dans notre exemple, sera le cas si l'objet d'étude est l'embouteillage), il devient impossible de modéliser indépendamment les deux niveaux (ou du moins, cela ne permet pas de comprendre la dynamique de formation du système ; il est évidemment toujours possible de modéliser phénoménologiquement l'embouteillage indépendamment des véhicules qui le composent).

Le cadre de notre vie quotidienne est celui d'un comportement découplé, c'est-à-dire d'un monde dans lequel les structures peuvent être considérées comme stables au regard de nos actions. En l'absence d'outils d'observation et/ou de mesure des dynamiques moléculaires et cellulaires, notre *interprétation naturelle* (Feyerabend, 1975) a logiquement plaqué ce cadre sur les systèmes biologiques en général et cellulaires en particulier, supposant l'existence de "points fixes", structurants, indépendamment de la dynamique générale du système¹. Or, le développement d'outils d'observation "en temps réel" de la cellule à l'échelle moléculaire a fait s'écrouler ce présupposé en montrant que la dynamique des compartiments cellulaires est du même ordre de grandeur (en temps) que la dynamique de leurs composants moléculaires. Se pose alors ouvertement la question de l'émergence de ces compartiments, de leur dynamique structurelle et de leurs interactions avec la dynamique fonctionnelle à laquelle ils participent.

Traditionnellement, la question de l'apparition de structures organisées à partir d'un ensemble de composants désorganisés relève de la physique statistique. Pourtant, celle-ci, parce qu'elle considère essentiellement des ensembles homogènes, peine à décrire les systèmes biologiques qui, eux, sont composés d'entités très hétérogènes. La modélisation individu-centrée offre alors un angle d'attaque qui peut intégrer cette hétérogénéité et ainsi mettre au jour les mécanismes structurant à l'œuvre dans un système biologique particulier (Zemirline *et al.*, 2002; Troisi *et al.*, 2005). Cependant, il n'est pas aujourd'hui envisageable de simuler, à l'échelle de la protéine, une cellule complète ni même un compartiment cellulaire et rien ne permet d'espérer parvenir à une telle capacité à moyen terme². C'est pourquoi nous nous plaçons ici dans une perspective différente : il ne s'agit pas de modéliser une entité biologique dans sa totalité mais de modéliser cette entité dans le cadre d'un processus particulier (on retrouve ici le principe énoncé au chapitre I qui veut que le processus de modélisation ne soit pas centré sur un *objet* mais bien sur un *ques-*

du trafic constaté sur les différents axes.

¹"L'idée que certaines caractéristiques décisives des organismes dépendent de systèmes matériels stables inaccessibles à des modifications de l'environnement est devenue maintenant partie intégrante d'un système si complexe d'hypothèses que sa critique directe n'a pratiquement aucune chance d'être entendue." (Feyerabend, 1975), p. 40.

²Si on compare la progression des outils de simulation et des outils de modélisation au cours des vingt dernières années, il semble clair que l'on sera capable d'observer une cellule réelle à l'échelle moléculaire *avant* d'avoir la capacité de simuler une cellule à la même résolution. Ainsi, il est aujourd'hui possible de suivre le déplacement de molécules uniques dans une cellule vivante, ce qu'aucun modèle ne permet d'espérer (Siebrasse *et al.*, 2007).

tionnement). En d'autres termes, nous nous autoriserons toutes simplifications de l'objet tant que celles-ci ne perturbent pas le processus étudié (du moins pas au point de rendre son étude impossible). En outre, il est évident que nous ne pourrions pas aborder de front les mécanismes structurant la cellule dans son ensemble. Aussi nous intéresserons-nous à des sous-systèmes cellulaires, supposés caractéristiques ou dans lesquels les processus de structuration spatio-dynamiques que nous souhaitons étudier nous semblent avoir une place centrale. Trois de ces sous-systèmes ont pour l'instant plus particulièrement retenu notre attention.

Dynamique et fonction des corps nucléaires

Le noyau des cellules eukaryotes est une structure connue de longue date puisque facilement discernable en microscopie optique. Par contre, en dehors des quelques éléments identifiables par les mêmes moyens (essentiellement les nucléoles et les chromosomes), il a longtemps été considéré comme relativement homogène, réceptacle d'une information génétique essentiellement traitée dans le cytoplasme. Les années 90 ont été fatales à cette répartition des rôles (Gorski et Misteli, 2005). D'une part, bien que le noyau cellulaire apparaisse relativement homogène en microscopie optique (ou, du moins, constitué de grandes zones homogènes), les nouvelles techniques de microscopie à fluorescence ont montré qu'il est fortement hétérogène en termes de composition moléculaire. De nombreuses structures nucléaires ont alors été identifiées (Spector, 2001; Dundr *et al.*, 2004; Misteli, 2005), structures qui, contrairement à la plupart des compartiments cytoplasmiques, ne sont pas délimitées par des membranes lipidiques mais "simplement" par leur composition protéique¹. D'autre part, le développement des programmes de séquençage (et en premier lieu du programme de séquençage du génome humain) et, d'une façon générale, les techniques d'acquisition à haut débit (génomique, transcriptomique) ont montré toute l'importance de la régulation des fonctions génétiques. Initialement considérés comme simples sources d'information (le "programme génétique"), le génome et le noyau ont alors acquis le statut d'élément central de traitement dynamique de cette information.

La découverte des corps nucléaires et, plus généralement, de la forte structuration du noyau pose de nombreuses questions quant au rôle de ces structures, à leur dynamique de formation ou à leur organisation spatiale (Misteli, 2001, 2005). Or, l'absence d'élément structurant (ou de frontière, c'est-à-dire ici de membrane) interdit de poser ces questions indépendamment les unes des autres. En outre, les caractéristiques structurales du noyau apparaissent comme extraordinairement régulières (par exemple en nombre, en forme, en répartition ...) au regard de la stochasticité des mécanismes moléculaires, supposée prévaloir dans le nucléoplasme (Pederson, 2001). Il est difficile de penser la structure nucléaire car elle repose à la fois sur des mécanismes macroscopiques (l'agitation thermique) et microscopiques (la structure et les interactions des protéines impliquées dans les méca-

¹Bien que cette caractéristique puisse apparaître accessoire, elle a de fortes implications. En effet, l'impossibilité de séparer ontologiquement l'intérieur de la frontière interdit la dissociation classique entre le processus de création de la structure (correspondant schématiquement à la dynamique de la membrane) et la fonction métabolique de la structure elle-même. L'absence de membrane interdit alors de supposer qu'une causalité externe prévaut à la formation de la structure avant de "passer le relais" à la fonction de la structure proprement dite. C'est alors toute la problématique de la morphogenèse qui se retrouve transcrite à l'échelle nucléaire. De fait, on observe des lignes de fracture épistémologiques qui ne sont pas sans rappeler celles qui prévalent dans le domaine de l'embryogenèse (Bourguin et Lesne, 2006).

nismes de formation des corps nucléaires). Or, si l'interface entre ces deux échelles est connue pour être le siège de mécanismes structurants (on peut ainsi citer les travaux sur la DLA – Diffusion Limited Aggregation – ou les mécanismes de réaction-diffusion), ceux-ci ne sont bien caractérisés que dans des situations simples, très loin de la réalité du noyau cellulaire. C'est pourquoi une démarche de modélisation/simulation nous semble indispensable pour appréhender les mécanismes à l'œuvre dans le noyau cellulaire.

Au delà des corps nucléaires et du noyau cellulaire, la question de l'auto-organisation issue du couplage entre processus de diffusion et d'agrégation peut se poser pour de nombreux systèmes biologiques. On peut ainsi évoquer le cas de l'agrégation pathologique des protéines dans les maladies neurodégénératives ou l'agrégation de résidus protéiques qui pourrait être impliquée dans le vieillissement cellulaire, par exemple chez la bactérie *Escherichia coli* (Stewart *et al.*, 2005). Si ces systèmes ne sont pas directement abordés dans le cadre de notre étude, ils sont autant d'applications potentielles des modèles développés. En effet, comme dans le noyau, ils posent la question de la structuration dynamique d'un sous-système moléculaire en lien avec des variations fonctionnelles à l'échelle cellulaire.

Structures membranaires

Tout comme le noyau cellulaire, les membranes ont longtemps été considérées comme des structures relativement homogènes. Cependant, le modèle de la “mosaïque fluide” (Singer et Nicholson, 1972; Saman et Delbruck, 1975), qui suppose les composants membranaires répartis de manière homogène du fait de la diffusion brownienne, a été battu en brèche lorsqu'ont été identifiés les *domaine membranaires*. Ces domaines sont des zones identifiables par leur composition lipidique, différente du reste de la membrane, ou par la surconcentration de protéines membranaires. En outre, suivant leur nature, ils peuvent être plus ou moins mobiles et dériver sur la membrane. On parle alors de “lipid rafts” (Jacobson *et al.*, 1995; Simons et Ikonen, 1997; Pike, 2004).

L'existence d'hétérogénéités membranaires étant *a priori* incompatible avec le principe de diffusion homogène, plusieurs explications ont été proposées sans qu'aucune ne se détache particulièrement. Ainsi, il a été suggéré que la diffusion puisse être contrainte par des “barrières membranaires” (Fujiwara *et al.*, 2002), limitant le déplacement des molécules et provoquant des surconcentrations de protéines membranaires. Plus récemment, des hypothèses basées sur des phénomènes d'auto-organisation ont été proposées (Daumas *et al.*, 2003a,b; Chen *et al.*, 2004). Celles-ci sont cependant difficiles à vérifier expérimentalement et il est souvent fait appel à des modèles de diffusion pour explorer ce type de mécanisme.

Dans tous les modèles cités ci-dessus, il est fait l'hypothèse que la diffusion membranaire est homogène ou, plus exactement, qu'elle est indépendante de la constitution des domaines membranaires. Pourtant, les processus de diffusion sont dépendants de la composition du substrat, c'est-à-dire, ici, de la composition lipidique de la membrane. Or, dans le cas des domaines membranaires, la surconcentration protéique est accompagnée d'une (ou accompagne une) modification de la composition lipidique de la membrane. Dès lors, nous pouvons identifier une question suffisamment précise pour amorcer un processus de modélisation conforme au cycle de vie énoncé au chapitre I : quel est l'impact de la constitution lipidique des domaines membranaires sur la répartition des protéines membranaires ? À cette première question pourront ensuite se rattacher plusieurs questions corollaires, par exemple sur la diffusion des domaines membranaires eux-mêmes ou sur

les mécanismes de coopération/compétition induits par les différences de concentration des protéines membranaires. Il a en effet été proposé que les variations de concentration des récepteurs membranaires permettent à la cellule de moduler la transmission du signal et ainsi d'adapter la sensibilité de ses voies de signalisation à son environnement (Bray *et al.*, 1998; Bray, 2002; Mello *et al.*, 2004).

Dynamique et stochasticité de la transcription

Les années 90 ont placé la régulation génique au cœur des processus cellulaires et de nombreux programmes de recherche ont vu le jour qui visent à caractériser le processus de transcription, à en mesurer la dynamique et à inférer les “réseaux de gènes” supposés réguler bon nombre de processus cellulaires. Dans la plupart de ces travaux, les mécanismes de transcription sont supposés stables et déterministes. En d'autres termes, pour un organisme donné (donc pour un réseau génétique donné), la dynamique de transcription est considérée homogène d'une cellule à l'autre (faible variabilité phénotypique) et d'un instant à l'autre (faible variabilité temporelle). Cependant, il est aujourd'hui clair que deux cellules identiques génétiquement (cellules isogéniques) et placées dans un même environnement peuvent présenter des caractéristiques phénotypiques très différentes et que ces caractéristiques sont en grande partie dues au caractère stochastique des événements moléculaires (Elowitz *et al.*, 2002). En outre, alors que ces variations étaient supposées négatives pour la cellule, plusieurs auteurs ont montré qu'elles pouvaient au contraire être un atout, par exemple en permettant à une population cellulaire d'affronter des milieux hétérogènes souvent plus efficacement que par un mécanisme de régulation génétique¹ (Steuer *et al.*, 2003; Thattai et Oudenaarden, 2004; Kussell et Leibler, 2005; Wolf *et al.*, 2005; Blake *et al.*, 2006).

Considérés globalement, ces résultats concourent à revoir notre position épistémologique vis-à-vis du “bruit” dans les systèmes cellulaires. En effet, alors que le bruit est traditionnellement perçu comme un phénomène de dégradation thermodynamique inéluctable contre lequel la cellule doit lutter pour maintenir son intégrité, il devient *aussi* un levier sur lequel la cellule a la possibilité de jouer pour assurer sa pérennité. Dès lors, il semble légitime de supposer que la cellule régule le caractère stochastique des processus de transcription, même si les mécanismes sous-jacents restent en grande partie à identifier. En effet, il est souvent supposé que la principale source de stochasticité réside dans le caractère discret des événements moléculaires de transcription qui, si les espèces moléculaires mises en jeu sont en faible concentration, entraîne des fluctuations d'autant plus fortes que les événements subséquents (c'est-à-dire, ici, les événements de traduction) se produisent par *bursts* (bouffées), amplifiant ainsi les variations initiales (Cai *et al.*, 2006).

Les mécanismes moléculaires sous-jacents de la stochasticité de l'expression génique sont proches de ceux que nous avons déjà évoqués dans le cadre de la modélisation des corps nucléaires. En effet, outre qu'ils partagent le même environnement cellulaire, du moins dans le cas des cellules eukaryotes, tous deux sont fortement dépendants des processus de diffusion moléculaires. Ainsi, van Zon *et al.* (2006) montrent que le bruit de transcription

¹Cette stratégie, dite de “bet-hedging”, a été initialement proposée par Cohen (1966) dans le cadre de la germination des plantes. On notera qu'on retrouve ici partiellement les thèmes abordés section II.3 puisque les questions de bruit phénotypique peuvent être approchées dans un contexte évolutif en utilisant le modèle *aevol*.

est fortement dépendant du temps de retour des facteurs de transcription diffusant dans le noyau (ou dans le cytoplasme dans le cas des cellules prokaryotes). L'utilisation de modèles cellulaires 4D apparaît alors comme particulièrement intéressante puisqu'elle permet de réaliser une première étude par simulation avant, lorsque cela est possible, d'approfondir les connaissances acquises sur le modèle par l'utilisation couplée de modèles mathématiques. On notera par ailleurs que certaines questions abordées section II.2 dans le cadre des réseaux de neurones peuvent trouver ici un équivalent (Swain et Longtin, 2006). Ainsi, l'influence du bruit sur la transmission du signal à travers le réseau (ou l'influence du réseau sur la propagation du bruit) est une question transversale qui concerne aussi bien les réseaux de neurones (Gerstner, 1999; Mar *et al.*, 1999) que les réseaux génétiques (Chen et Wang, 2006).

4.3 Modèles et expériences

Les deux sections précédentes (II.2 et II.3) nous ont vu présenter des modèles finalisés, relativement bien établis (bien que pouvant toujours faire l'objet de variantes). Dans le cas de la modélisation cellulaire 4D, la question des modèles est beaucoup plus ouverte. En effet, cette approche de la modélisation cellulaire n'en est encore qu'à ses balbutiements et aucun consensus ne se dégage clairement pour définir le niveau de modélisation le plus adapté entre l'échelle atomique propre à la dynamique moléculaire et les équations différentielles utilisées dans les modèles métaboliques.

La méthodologie proposée au chapitre I a mis en avant l'importance d'une définition claire des questions scientifiques préalablement à la construction d'un modèle. Cependant, nous avons aussi insisté sur la nécessité d'inclure, dans le cycle de vie du modèle, une phase que nous avons qualifiée d'"exploratoire" au cours de laquelle le modélisateur doit manipuler plusieurs modèles. Dans le cas des modèles cellulaires, il est nécessaire, parallèlement, d'identifier les niveaux de description pertinents et calculables. Il nous semble que, même si des modèles cellulaires 4D ont déjà pu permettre de produire des résultats pertinents – on citera par exemple les travaux pionniers réalisés par l'équipe de Dennis Bray à Cambridge (Shimizu *et al.*, 2003; Andrews et Bray, 2004; Lipkow *et al.*, 2005), ceux-ci en sont encore largement au stade exploratoire comme en témoigne le foisonnement de modèles souvent très différents les uns des autres (Uhrmacher *et al.*, 2005).

L'échelle mésoscopique, interface entre biologie moléculaire et cellulaire ?

Au sein d'un système cellulaire les événements d'intérêt se produisent à des échelles très variées, allant de la molécule – voire de l'ion – à la cellule considérée dans sa globalité. Or, loin d'être déconnectées, ces échelles sont au contraire intimement imbriquées. Ainsi, le déroulement des événements cellulaires repose-t-il naturellement sur les événements moléculaires tandis que ceux-ci subissent le jeu des contraintes imposées par les structures cellulaires. C'est cette imbrication étroite des différents niveaux d'organisation qui rend difficile l'identification des niveaux de description pertinents pour la modélisation cellulaire. Ainsi, la modélisation du système cellulaire fait généralement intervenir des variables continues représentant le comportement moyen d'un très grand nombre de molécules (température, concentration, taux de réaction ...) et néglige l'influence individuelle des molécules. À l'inverse, à l'échelle moléculaire (par exemple dans les modèles

“all-atoms” utilisés en dynamique moléculaire) ces comportements moyens perdent leur sens puisque les atomes sont alors considérés comme des entités individuelles, discrètes.

Nous avons vu que les systèmes cellulaires abordés ici sont caractérisés par un fort couplage entre le niveau macroscopique et le niveau microscopique. Ainsi, dans les trois sous-systèmes évoqués précédemment, les comportements émergents sont-ils dépendants des caractéristiques des constituants moléculaires impliqués (échelle microscopique) mais aussi des propriétés de diffusion dans le noyau cellulaire, le cytoplasme ou sur la membrane cellulaire. Or, la diffusion est un processus lié à l’agitation thermique, dépendant de caractéristiques macroscopiques (la température et la viscosité du milieu) et qu’il est aujourd’hui inimaginable de modéliser à l’échelle microscopique¹. Il est donc clair que les modèles devront intégrer les caractéristiques microscopiques des constituants moléculaires tout en conservant partiellement les descriptions moyennées du niveau macroscopique. En d’autres termes, les modèles cellulaires devront se situer au niveau *mésoscopique*².

La notion même d’échelle mésoscopique ne peut être définie que corrélativement à un niveau “micro” et à un niveau “macro” donnés (ainsi, on peut parler d’échelle mésoscopique pour les phénomènes météorologiques – il s’agit alors d’une échelle intermédiaire entre l’échelle climatique, le déplacement des masses d’air à l’échelle planétaire, et l’échelle géographique où les déplacements d’air s’individualisent. C’est à cette échelle que peuvent se comprendre/prévoir les phénomènes météorologiques tels que les tempêtes ou les orages). Dans notre cas, nous savons déjà que le niveau microscopique correspondra à l’échelle moléculaire et que le macro correspondra à l’échelle cellulaire (avec cependant une certaine variabilité : nous considérerons ainsi que les organites peuvent être décrites macroscopiquement aux mêmes échelles que les cellules³). Cependant, les questions que nous avons soulevées dans la section précédente permettent de préciser notre propos. Nous avons vu que nous concentrons nos études sur l’émergence, la fonction et la dynamique de structures multi-moléculaires. Cela signifie, d’une part, que nos modèles devront être spatialisés, et d’autre part que la description microscopique des entités moléculaires devra tenir compte – éventuellement de manière très simplifiée – de leurs caractéristiques spatiales (incluant donc des notions d’encombrement, de loi de déplacement et d’interactions entre molécules). En revanche, dans le cas général, la description biochimique des molécules pourra être simplifiée, de même que les réactions dans lesquelles elles sont impliquées. De façon similaire, à l’échelle macroscopique, nous pourrions négliger les caractéristiques biochimiques des constituants cellulaires pour nous concentrer sur la description des contraintes qu’ils imposent au déplacement moléculaire.

¹Plus précisément, si une telle modélisation est possible, elle ne permet en aucun cas le passage à l’échelle qui consisterait à modéliser, à l’échelle microscopique, les processus de diffusion dans l’ensemble d’une cellule.

²Du grec “mésos” (“entre deux”), l’échelle mésoscopique correspond à une échelle intermédiaire à laquelle les approximations macroscopiques sont *encore* valides (les notions de température, de pression, de viscosité ... sont donc définies) mais où il est *déjà* nécessaire de considérer le comportement d’entités moléculaires individuelles discrètes.

³Compte-tenu de la très grande diversité cellulaire – en taille comme en structure – cette “approximation” n’est pas moins valide que de considérer une échelle cellulaire “standard”.

Contre la cellule virtuelle

Avant de détailler les modèles que nous avons mis en œuvre, nous voudrions revenir brièvement sur la distinction que nous avons opérée ci-dessus entre modèles cellulaires mésoscopiques et “cellule virtuelle”. L’idée de cellule virtuelle correspond à la volonté d’intégrer, au sein d’un modèle unique, *toutes* les caractéristiques microscopiques d’une espèce cellulaire donnée afin d’en comprendre la dynamique globale. Elle soulève, selon-nous, trois questions. (i) Comment de tels modèles s’intègrent-ils au corpus de connaissance de la biologie moléculaire et cellulaire ? (ii) Quelles connaissances nouvelles est-on en droit d’en attendre ? et (iii) dans quel cadre épistémologique ces modèles se situent-ils ?

La biologie moderne se veut de plus en plus *intégrative* (Ideker *et al.*, 2006), signifiant ainsi la volonté de dépasser le cadre de lecture moléculaire auquel la biologie était parvenue et de remonter dans les échelles en intégrant les connaissances moléculaires. Dans ce contexte, l’informatique et la modélisation ont acquis un rôle clé en permettant d’exprimer cette intégration. Le développement de cellules virtuelles – modèle *intégré* par nature – peut alors apparaître comme l’aboutissement de ce projet. Cependant, le développement de modèles intégrés ne doit pas être confondu avec le développement de modèles “totaux” et il nous semble qu’il y a là un glissement sémantique non maîtrisé. En effet, l’objectif de la biologie intégrative n’est pas l’intégration pour elle-même. Au contraire, l’objectif est (doit être) de découvrir ce qui *permet* l’intégration, c’est-à-dire la constitution d’une entité cellulaire à partir d’individualités moléculaires. En d’autres termes, les modèles de cellules virtuelles posent la cellule *a priori* et y intègrent ensuite les constituants moléculaires (voire, plus couramment, les voies métaboliques). Or, étant donnée l’extraordinaire quantité de paramètres de ces modèles, étant données les caractéristiques physiques qu’ils doivent le plus souvent négliger (par exemple la spatialisation), il nous semble fort peu probable que ces modèles permettent d’élaborer de nouvelles connaissances¹. En fait, selon nous, les modèles de cellule virtuelle souffrent d’un défaut majeur inscrit dans leur nature même. En effet, ils se définissent comme modèles d’un objet, indépendamment de toute question sur cet objet. Dès lors, il n’est plus possible de définir une frontière opérationnelle au modèle cellulaire : toute connaissance sur l’objet a vocation à intégrer le modèle, ce qui conduit à une croissance non-maîtrisée et à la production d’un modèle dont le seul sens est d’être (sensément) complet. Pour conclure, il nous semble que les modèles de cellule virtuelle sont avant tout un fantasme démiurgique comme l’étaient avant eux les “modèles” d’hommes artificiels ou de consciences artificielles et comme le sont aujourd’hui les tentatives de biologie synthétique. Faisant appel plus à notre inconscient qu’à notre conscience scientifique, ils posent des questions qui sont en fait très éloignées de la biologie et contribuent avant tout, de part leur extraordinaire complexité, à l’avancée des sciences

¹Il serait cependant prétentieux de l’affirmer : comme nous l’avons énoncé au chapitre I, l’utilité *a priori* d’un modèle est impossible à affirmer/infirmer. Cependant, les modèles de cellule virtuelle sont d’une complexité telle qu’ils progressent moins vite que les outils d’observation et, nous l’avons vu, il semble très probable que la biologie cellulaire disposera d’outils d’observation temps-réel à l’échelle moléculaire avant de disposer de modèles cellulaires complets à la même échelle : “*We physicists often look at [biologists] and say, “You know the reason you fellows are making so little progress ?” (Actually I don’t know any field where they are making more rapid progress than they are in biology today.) “You should use more mathematics, like we do.” They could answer us—but they’re polite, so I’ll answer for them : “What you should do in order for us to make more rapid progress is to make the electron microscope 100 times better”* (Richard Feynmann, discours devant le congrès annuel de la American Physical Society, Caltech, 1959).

de l'ingénieur et des sciences de l'information.

Est-ce à dire que les modèles cellulaires sont à rejeter ? Certainement pas, à condition de revenir à l'essence même des modèles et d'accepter notre méconnaissance des règles de structuration à l'échelle cellulaire : il y a une place (un besoin) pour des modèles qualitatifs de complexité limitée, définis à l'échelle moléculaires, et destinés à identifier les mécanismes structurants la "société moléculaire" pour aboutir à une action coordonnée (la cellule). En outre, plutôt que de déployer toute notre énergie pour définir un modèle unique, nous préférons multiplier les modèles pour mieux cerner les phénomènes étudiés et développer progressivement une connaissance qui ne sera, de toutes façons, pas directement présente dans le(s) modèle(s) mais bien dans notre esprit (dans un sens élargi à l'ensemble de la communauté scientifique).

3DSPI : simulation 3D d'interactions dans de grands ensembles de protéines

Les questions d'intérêt – pour nous – se situant dans les règles de structuration des entités multi-moléculaires, leur dynamique, et, *in fine*, leur fonction, nous avons défini un modèle individu-centré, 3DSPI (pour "3D Simulation of Protein Interactions"), permettant de simuler les interactions dans de grands ensembles protéiques. Ce modèle est basé sur une description très simplifiée des entités moléculaires d'une part et du milieu cellulaire d'autre part. On notera que ces simplifications sont imposées (le modèle doit rester paramétrable et calculable), mais aussi choisies (pour que les résultats du modèle soient analysables en termes de causalité). Deux versions de 3DSPI ont été développées à ce jour et la troisième version est actuellement en cours de développement. Nous décrivons ci-après les choix principaux tels qu'ils ont prévalu au développement de ce modèle. Pour une description complète, nous renvoyons le lecteur aux travaux publiés par Soula *et al.* (2005b) pour la version 1 et par Coulon *et al.* (2007); Coulon *et al.* (2008) pour la version 2¹.

Modélisation du milieu cellulaire Dans 3DSPI, le milieu cellulaire est tridimensionnel, considéré comme a priori homogène, et sujet à une agitation thermique permanente. Il transmet cette agitation thermique aux entités moléculaires qui sont donc soumises à un mouvement brownien. Suivant les simulations, le milieu pourra être considéré clos (c'est-à-dire cerné par une membrane hermétique, ce qui est toujours le cas dans les versions 1 et 2 de 3DSPI) ou non (espace torique) et peuplé uniquement de molécules, même s'il peut contenir des éléments macroscopiques supposés contraindre les mouvements moléculaires.

Modélisation des entités moléculaires Les entités moléculaires constituent le cœur du modèle. Dans la version 1, les molécules étaient modélisées comme des sphères dures, caractérisées par leur rayon, et interagissent selon des lois probabilistes d'association ou de dissociation. Dans les versions 2 et 3, en revanche, elles sont modélisées comme des entités ponctuelles (éventuellement des regroupements d'entités ponctuelles²) et elles interagissent suivant un champ de potentiel leur conférant un volume d'exclusion stérique.

¹Les choix principaux de modélisation de la version 3 sont similaires à ceux de la version 2. La version 3 est surtout destinée à rendre la réalisation des expériences *in silico* plus simple.

²Un prototype de modèle "multi-domaines" a d'ores et déjà été développé sur la base de la version 1. Cette possibilité sera pleinement intégrée à la version 3.

Ce modèle permet d'unifier les différentes interactions moléculaires et de les décrire par une seule équation (équation II.7).

$$V_{ij} = \epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma}{r_{ij}} \right)^6 \right] + \frac{Q}{r_{ij}} e^{-\frac{r_{ij}}{r^*}} \quad (\text{II.7})$$

où r_{ij} représente la distance séparant la molécule i de la molécule j ; ϵ , σ et Q sont les paramètres de l'interaction (et dépendent donc des espèces moléculaires en présence) et r^* est une caractéristique du milieu.

L'équation II.7 permet de décrire les interactions moléculaires issues des forces de van der Waals (potentiel de Lennard-Jones) et des interactions coulombiennes écrantées par le solvant. À partir des champs de potentiels, les forces d'interaction entre deux molécules sont obtenues par dérivation et correspondent donc à des forces successivement répulsives, attractives et répulsives au fur et à mesure qu'augmente la valeur de r_{ij} (figure II.11). On notera par ailleurs que la fonction de potentiel utilisée permet de s'affranchir du modèle des sphères dures puisque ce sont désormais les forces qui interdisent l'interpénétration des molécules. Il est alors possible de modéliser toutes les interactions sans supposer de discontinuité aux points de "contact".

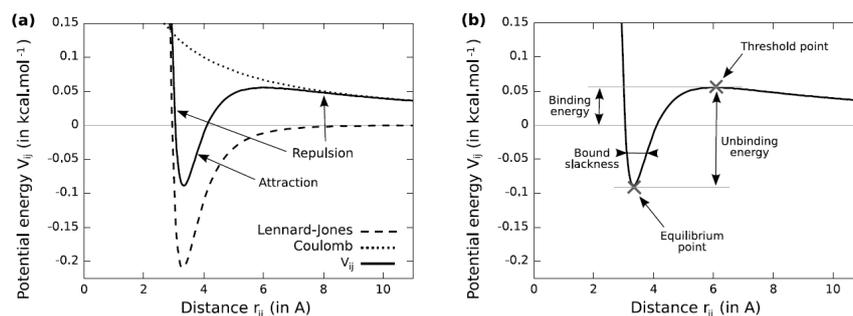


FIG. II.11 – Potentiel d'interaction moléculaire dans le modèle atomique. (a) Le potentiel d'interaction entre deux molécules résulte de la combinaison de deux forces principales (à gauche) : la force de Coulomb, répulsive ici, et les interactions de Van der Waals. Celles-ci sont attractives à moyenne distance mais fortement répulsives à courte distance (la force appliquée au couple moléculaire est égale à la dérivée du champs de potentiel). Les interactions moléculaires sont donc fortement répulsives à courte distance, attractives au delà d'un seuil de minima d'énergie (point d'équilibre) puis, dans le cas d'une force de Coulomb répulsive, faiblement répulsives au delà d'une barrière d'énergie d'interaction. (b) La forme du champs de potentiel permet de décrire les principales caractéristiques de l'interaction (stabilité, souplesse ...). D'après (Coulon *et al.*, 2008).

Dans la version 2 du logiciel, toutes les interactions moléculaires sont décrites par une équation similaire à l'équation II.7. La version 3 permettra d'utiliser une plus grande variété de champs de potentiels, y compris des potentiels nuls. Nous rejoignons en cela les principes définis au chapitre I : le réalisme voudrait que toutes les interactions soient modélisées sous la forme d'un champ de potentiel (comme dans la version 2). Cependant, ce réalisme interdit la réalisation d'un grand nombre d'expériences indispensables pour comprendre le modèle. En effet, il interdit d'étudier indépendamment l'influence des différents paramètres et rend donc difficile l'analyse du système modélisé.

Au delà de 3DSPI, un foisonnement de modèles complémentaires

3DSPI est un modèle complexe dans lequel un très grand nombre de composants, éventuellement hétérogènes, interagissent suivant des lois non triviales (voir équation II.7 et figure II.11). Si elles permettent de mettre en évidence des phénomènes d’auto-organisation à l’échelle mésoscopique (Soula *et al.*, 2005b; Coulon *et al.*, 2007), les simulations réalisées avec un tel modèle sont difficiles à analyser et à interpréter. L’analyse du comportement du modèle demande donc l’utilisation conjointe de modèles plus simples et/ou plus généraux afin de dégager quelles sont, dans 3DSPI, les caractéristiques qui conduisent à l’organisation observée. Parmi les modèles complémentaires, nous pouvons ainsi citer les modèles mathématiques de la diffusion¹ et, surtout, les systèmes de particules (dans lesquels les particules sont supposées ponctuelles) qui permettent de s’affranchir de l’encombrement moléculaire. En pratique, la démarche suivie consiste à utiliser 3DSPI pour forger une première intuition des phénomènes puis de modéliser ceux-ci spécifiquement au moyen d’outils mathématiques ou informatiques dédiés. Cette deuxième série de modèles est alors construite spécifiquement pour comprendre et/ou préciser l’intuition première plutôt que pour modéliser un système biologique particulier.

Par ailleurs, si l’objectif de 3DSPI est d’appréhender les mécanismes de régulation spatio-temporels, la modélisation de l’espace n’est heureusement pas systématiquement nécessaire à la compréhension de phénomènes cellulaires. Dans de nombreux cas, le système peut être considéré homogène, par exemple lorsque l’effet des concentrations moléculaires dépasse l’effet des hétérogénéités spatiales.

4.4 Résultats

L’utilisation des différentes versions de 3DSPI nous permet de mieux comprendre les mécanismes de structuration cellulaire. Plutôt que de “résultats”, nous préférons cependant parler ici de “résultats préliminaires”. En effet, les résultats acquis jusqu’ici demandent à être précisés, soit par des simulations complémentaires, soit par des modèles mathématiques, soit par des discussions avec des biologistes moléculaires et cellulaires. Ils doivent donc être considérés comme des “pistes de recherche et de réflexion” plus que comme des acquis solides et stabilisés. Nous considérons donc que les modèles présentés ici sont encore au stade 2 (exploratoire) du cycle de vie énoncé au chapitre I (section I.4.4).

Auto-organisation des corps nucléaires

Dès les premières expérimentations *in-silico* conduites avec 3DSPI-V1, il est clairement apparu que la formation des corps nucléaires pouvait être issue d’un processus d’auto-organisation (Soula *et al.*, 2005b). En d’autres termes, la structure et la dynamique des corps nucléaires pourraient résulter d’un mécanisme de transition de phase dans un système de particules en interaction (Carrero *et al.*, 2006; Misteli, 2007). Cette hypothèse permet de rendre compte de plusieurs des caractéristiques des corps nucléaires et, en

¹Bien qu’ils soient indispensables à la compréhension des mécanismes d’auto-organisation, ces modèles mathématiques ne seront pas abordés ici. Ils sont en effet réalisés dans un cadre collaboratif et ne relèvent donc pas directement de travaux personnels. Pour une excellente introduction à la modélisation mathématique des phénomènes de diffusion brownienne, le lecteur pourra se référer à (Berg, 1993; Schweitzer, 2003).

particulier, de leur stabilité locale et globale dans un contexte très dynamique. En effet, plusieurs études *in vivo* ont montré que les temps de résidence des protéines composant les corps nucléaires y sont très courts (de l'ordre de quelques secondes) alors que les structures elles-mêmes sont conservées pendant des temps beaucoup plus longs (stabilité locale) et que leur nombre total est relativement constant (stabilité globale). Nous avons ainsi pu montrer que la taille moyenne des structures ainsi que leur nombre dépendent des paramètres d'affinité entre les protéines qui les composent (Soula *et al.*, 2005b).

Au delà des mécanismes de structuration des corps nucléaires, cette hypothèse permet de déduire un certain nombre de propriétés quant à la formation (ou, inversement, à la disruption) et à la fonction des corps nucléaires. Ainsi, si la formation des corps nucléaires est liée à une transition de phase dans le nucléoplasme, alors la concentration des protéines composant le corps nucléaire doit être constante (aux fluctuations thermiques près) dans le noyau et dépendre directement des caractéristiques du potentiel d'interaction (figure II.12). On peut en déduire que la dynamique des corps nucléaires va être directement liée à la dynamique globale du noyau. Ainsi, une simple augmentation de la taille du noyau pourrait conduire à la disruption de certains corps nucléaires : si l'augmentation du volume nucléaire entraîne la variation des concentrations protéiques, alors la simple régulation de la concentration dans le nucléoplasme peut provoquer la perte de matière au niveau des corps nucléaires, perte pouvant aller jusqu'à la disruption totale (on notera que, si cette hypothèse est valide, la disruption ne doit pas s'accompagner d'une augmentation de la concentration de ces protéines dans le nucléoplasme – ce qui pourrait permettre de vérifier cette hypothèse *in-vivo*). Enfin, la régulation de la concentration de ces protéines, si elle peut apparaître comme un effet secondaire de la formation des corps nucléaires, pourrait aussi bien en être une fonction. Les corps nucléaires (ou du moins certains d'entre eux) auraient alors un rôle de régulateur de la concentration de certaines protéines dans le noyau, régulant ainsi indirectement leur activité biochimique. Cette hypothèse, formulée par Misteli (2001), est confirmée par nos modèles (figure II.12).

Structuration des domaines membranaires : l'hypothèse cholestérol

Les travaux préliminaires conduits sur les corps nucléaires ont attirés notre attention sur les liens étroits entre les mécanismes de diffusion des protéines et les mécanismes de structuration spatiale. Dans ce cadre, nous avons proposé que de tels liens puissent être à l'œuvre dans d'autres systèmes biologiques et, plus particulièrement, dans les membranes cellulaires. En effet, la membrane n'est pas un fluide homogène. Elle est constituée de différentes familles lipidiques (essentiellement les phospholipides, les glycolipides et le cholestérol qui composent respectivement environ 50%, 25% et 25% de la membrane). Or, les propriétés de diffusion peuvent être localement modifiées par la présence de certains types de lipides. C'est pourquoi nous avons étudié l'impact de l'hétérogénéité de la composition de la membrane cellulaire sur la répartition des protéines membranaires¹.

Nous avons ainsi pu montrer, à l'aide d'un modèle préliminaire (modélisation de particules ponctuelles animées par un mouvement brownien) que des variations spatiales des coefficients de diffusion liées aux différences de composition lipidique de la membrane peuvent provoquer des variations de concentrations protéiques. En outre, ce résultat a pu être confirmé par un modèle mathématique de la diffusion en milieu hétérogène, ce qui nous

¹En collaboration avec le laboratoire de Régulation Métabolique, Nutrition et Diabète à Lyon.

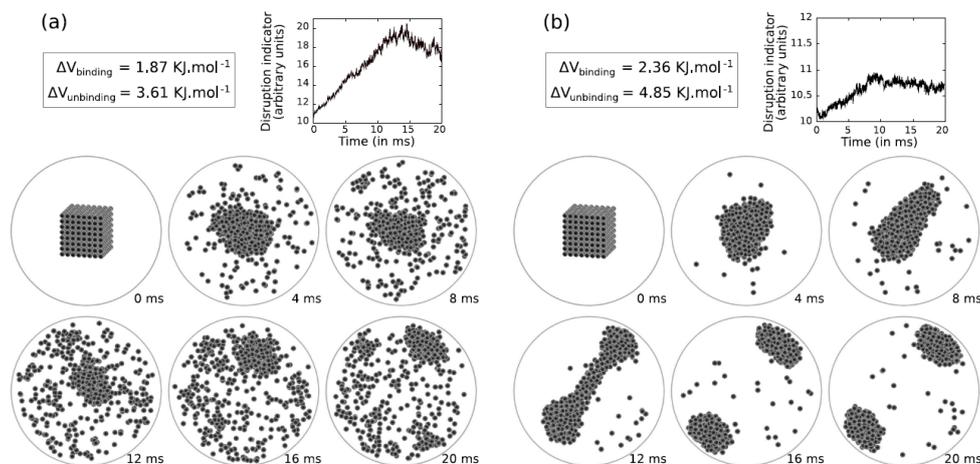


FIG. II.12 – Influence de l’interaction protéine-protéine sur la structure des agrégats multi-protéiques. Les valeurs d’énergie d’association (binding) et de dissociation (unbinding) correspondent aux seuils d’énergie nécessaires pour passer d’un état à l’autre (voir figure II.11). À partir d’un état initial totalement agrégé (un cube), les valeurs d’interaction déterminent le comportement du système (observé ici toutes les 4 ms et mesuré en temps réel par un “indice de disruption”). En outre, on constate que les valeurs d’interaction (faibles à gauche, fortes à droite) déterminent indirectement la morphologie des structures ainsi que la concentration moyenne de protéines dans l’environnement (hors agrégats). D’après (Coulon *et al.*, 2008).

a aussi permis de quantifier les variations de concentration en fonction des coefficients de diffusion. Or, on sait que, dans la membrane, la présence de cholestérol induit un ralentissement de la diffusion (Niemelä *et al.*, 2007), les protéines diffusant environ dix fois plus lentement à proximité du cholestérol membranaire. Un tel ralentissement provoque donc, selon notre modèle, l’apparition de surconcentrations locales de protéines, même si celles-ci diffusent toujours librement. Nous avons donc pu proposer que les domaines membranaires se forment à proximité d’îlots de cholestérol, du simple fait du ralentissement induit par les variations de la composition lipidique (publication en préparation).

Plusieurs observations viennent appuyer notre hypothèse. En particulier, Goodwin *et al.* (2005) montrent que la déplétion de la membrane en cholestérol modifie la diffusion des protéines membranaires et provoque la disruption des micro-domaines. Fujita *et al.* (2007) montrent qu’il en est de même en cas de refroidissement. Or, contrairement aux phospholipides, la diffusion au voisinage du cholestérol est inversement proportionnelle à la température. Dans ce cas, notre modèle prédit que l’uniformisation de la diffusion va effectivement conduire à une répartition uniforme des protéines. On notera enfin qu’il n’y a pas nécessairement opposition entre notre “hypothèse cholestérol” et d’autres mécanismes de structuration des domaines membranaires. En effet, plusieurs mécanismes peuvent être actifs simultanément et une transition de phase due à une sur-concentration protéique locale, par exemple, aura plus de chances de se former en une zone de diffusion lente (donc, selon notre hypothèse, au voisinage des cholestérols membranaires). Une de nos perspectives est d’ailleurs de tester, dans un modèle similaire à 3DSPI mais en dimension 2, l’influence couplée du ralentissement de la diffusion et des interactions protéine-protéine.

Contrôle de la stochasticité de l'expression génique

Alors que la stochasticité de l'expression génique est souvent considérée comme directement liée à la concentration des espèces moléculaires impliquées, la modélisation nous a permis de montrer que les interactions entre ces différentes espèces peuvent très largement compliquer ce processus. Ainsi, en modélisant les interactions coopératives ou compétitives entre facteurs de transcription¹, nous avons pu montrer qu'en présence de deux facteurs de transcription non-indépendants, il est possible de réguler le bruit indépendamment des concentrations moléculaires (voire d'accroître le bruit en augmentant les concentrations moléculaires ce qui était *a priori* inattendu). En outre, suivant les interactions entre facteurs de transcription (et suivant leur nombre), une très grande variété de situations sont envisageables, telles que la bistabilité (la transcription oscillant donc aléatoirement entre deux états et non autour d'un état moyen) ou le découplage entre les échelles de temps des association/dissociation des facteurs de transcription sur le site de fixation et les variations du taux d'initiation de la transcription (les premières pouvant être beaucoup plus rapide que les secondes).

Plus encore que pour les deux systèmes cellulaires étudiés ci-dessus, nous n'en sommes ici qu'à la formulation de résultats préliminaires et le modèle demande encore à être analysé et caractérisé pour aboutir à des résultats généraux. Par ailleurs, dans son état actuel, il ne tient aucun compte des mécanismes de diffusion alors même qu'il a été montré que ceux-ci peuvent avoir une forte influence sur la stochasticité de la transcription (van Zon *et al.*, 2006). C'est pourquoi nous envisageons aujourd'hui de coupler ce modèle de transcription avec nos modèles 3D afin de mieux comprendre comment les différentes sources de stochasticité interagissent et quelles sont leurs influences respectives sur le processus de transcription.

4.5 Conclusion et perspectives

La modélisation individu-centrée des systèmes cellulaires est un domaine encore jeune et de nombreux modèles sont proposés, à différentes échelles et basés sur des formalismes très différents (de l'automate cellulaire aux "modèles tout atomes" en passant par les graphes). Nous avons vu que, pour notre part, nous pensons souhaitable de multiplier les modèles pour développer une réelle maîtrise de l'objet d'étude. En effet, il est clair que les modèles individu-centrés ne pourront pas, à moyen terme, approcher la complexité réelle des systèmes cellulaires. Un effort d'abstraction est donc particulièrement nécessaire, effort qui, selon nous, demande que soient manipulés des modèles complémentaires les uns des autres. Une fois encore, l'objectif doit être ici de *comprendre* et non de *reproduire*.

Plus encore que dans les deux sections précédentes, nous avons présenté ici des travaux exploratoires. Nos perspectives sont donc immédiates : il s'agira, pour chacun des modèles présentés ici, d'avancer dans le cycle de vie des modèles, c'est-à-dire de passer de la phase exploratoire (phase 2) à la phase de stabilisation du modèle (phase 3) avant de conduire de réelles expérimentations *in silico* (phase 4).

¹Ce travail est réalisé dans le cadre de la thèse de M. Antoine Coulon. Les modèles développés sont des modèles individu-centrés basés sur l'algorithme Gillespie SSA (Gillespie, 1976) et des modèles mathématiques.

Conclusion

En rédigeant ce mémoire, nous avons un double objectif : présenter l'état de nos travaux en modélisation de systèmes biologiques et formaliser la "méthodologie pragmatique" qui s'est imposée à nous au cours de nos interactions avec la biologie et avec les biologistes. Ce double objectif pourra avoir donné au lecteur une impression d'hétérogénéité, d'autant plus que les domaines d'application présentés dans le chapitre II peuvent eux-mêmes sembler très hétérogènes. De fait, sur un plan strictement biologique, ils le sont mais nous aimerions, pour conclure, montrer que malgré cela les passerelles sont nombreuses pour peu qu'on prenne le temps de s'attacher à la description des mécanismes généraux étudiés et non à leur "implémentation" biologique ou biochimique.

La première passerelle est évidemment celle de l'approche de modélisation et c'est celle que nous avons cherché à exprimer dans le premier chapitre. En effet, même si les applications peuvent être diverses et même si les modèles eux-mêmes peuvent être très différents, la démarche que nous avons suivie est la même pour les différentes questions abordées. Ainsi, nous sommes tout particulièrement attachés à une démarche de modélisation guidée par les questions et non par les objets. C'est cette démarche qui nous a guidé dans le développement de nos modèles (du moins à partir du moment où nous en avons pris conscience). En outre, dans tous les cas, notre approche de modélisation a été, techniquement et épistémologiquement, individu-centrée. Techniquement, cela signifie que tous les modèles développés intègrent, nous l'avons dit, plusieurs niveaux d'organisation mais que seul le niveau microscopique est explicitement modélisé. Les niveaux supérieurs (macroscopique et, surtout, mésoscopique), eux, sont "simplement" observés. Épistémologiquement, cette approche de la modélisation est issue de notre intérêt pour les phénomènes émergents et du sentiment que les systèmes biologiques reposent majoritairement sur de tels mécanismes ou, du moins, que les phénomènes émergents ont un rôle important dans la plupart des phénomènes biologiques.

La seconde passerelle est donc celle de l'étude des phénomènes émergents en biologie. Dans ce cadre, les systèmes étudiés ici sont tous basés sur un antagonisme entre une force de diffusion – tendant à disperser les éléments du système dans l'espace d'états – et une force d'agrégation tendant à les recentrer ou à les regrouper dans une ou plusieurs zones déterminées par la position et les caractéristiques des agents eux-mêmes. Dans le cas de la modélisation des corps nucléaires (section II.4), ces deux forces sont clairement identifiées – il s'agit ici réellement d'une diffusion et d'une agrégation¹. De même, dans le cas de l'évolution, les deux forces en présence sont la mutation et la sélection. La

¹L'utilisation de ces termes ne doit pas conduire le lecteur à confondre les travaux présentés ici et la "DLA" (Diffusion-Limited Aggregation). En effet, si les mécanismes de diffusion peuvent être similaires, les mécanismes d'agrégation sont ici relativement différents. Ainsi, dans le cadre de nos modèles, l'agrégation autorise le relargage, ce qui n'est pas le cas de la DLA "classique".

première va disperser les génomes dans l'espace des séquences tandis que la seconde va conduire à leur convergence autour des meilleurs individus de la population. Dans cette vision c'est l'évolution darwinienne elle-même qui est le phénomène émergent. La sélection pour un équilibre adéquat entre robustesse et variabilité (section II.3) est alors une micro-émergence – une rétroaction du tout sur les caractères individuels – puisque ce sont les agents eux-mêmes qui s'adaptent à leur environnement (celui-ci étant façonné par le phénomène émergent lui-même). Considérée dans une perspective d'équilibre entre diffusion et agrégation, la sélection pour la robustesse et la variabilité correspond donc à une auto-régulation du processus de diffusion en fonction de l'intensité de l'agrégation. Enfin, dans le cas de l'apprentissage de comportement (section II.2), il est intéressant de constater que l'alternance STDP/ANTI-STDP correspond bien à un mécanisme de dispersion/agrégation (dans l'espace des poids synaptiques) mais pour lequel c'est l'alternance des deux lois qui provoque l'équilibre dynamique du comportement.

La troisième passerelle est l'existence de questions transversales aux trois domaines d'application. Deux questions transversales nous semblent particulièrement importantes. La première concerne l'évolution des réseaux. Même si nous n'avons fait que l'évoquer, la notion de réseau est en effet devenue transverse à un grand nombre de domaines scientifiques. Elle est en particulier centrale en biologie des systèmes où elle offre un formalisme intégratif relativement simple d'accès. Ici, la question clé est le processus de construction des réseaux (que ce soit par l'évolution ou par l'adaptation), leur dynamique et leur fonction. Les réseaux biologiques, en particulier, présentent des structures particulières puisqu'ils sont généralement modulaires, invariants d'échelle et "petit monde" (Barabasi, 2004). Il serait particulièrement intéressant de déterminer l'origine de ces structures : sont-elles neutres, auquel cas elles seraient simplement issues de leur dynamique de construction ? directement sélectionnées parce qu'elles présentent un avantage sélectif ? ou indirectement sélectionnées, par exemple en fonction de leur robustesse ? La deuxième question transversale est celle du bruit. Nous avons vu (section II.4) que les systèmes biologiques sont soumis à un fort bruit intrinsèque. La question de l'intérêt du bruit pour le système est apparue dans chacun de nos domaines d'application, depuis les réseaux neuronaux (pour lesquels nous avons vu que le bruit peut améliorer le codage de l'information) jusqu'à l'évolution où nous avons vu que le bruit, héritable ou non, peut être régulé par l'évolution pour s'adapter aux caractéristiques environnementales. L'introduction d'un bruit de transcription dans le modèle Rævol constitue d'ailleurs selon nous une perspective particulièrement intéressante.

Enfin, et cela constitue selon nous une quatrième passerelle, il est possible de développer des modèles transverses, réunissant deux, voire trois des modèles présentés au chapitre II. Dans cet état d'esprit, une perspective immédiate serait d'utiliser l'évolution digitale pour faire évoluer les paramètres de nos modèles cellulaires. Il serait ainsi possible de déterminer les points d'équilibre dans les structures moléculaires ou dans les réseaux de régulation. De même, il est possible de faire évoluer des structures neuronales afin d'étudier la façon dont les comportements des animats vont se structurer au sein d'une collectivité. Ici, la perspective la plus intéressante serait d'étudier l'émergence de structures de communication entre animats en fonction des caractéristiques environnementales mais aussi de la structure génétique de la population (Floreano *et al.*, 2007).

Toutes ces passerelles constituent autant de perspectives pour les travaux présentés dans ce mémoire et viennent s'ajouter aux perspectives déjà évoquées dans chacun des domaines

applicatifs. Dès lors, une difficulté majeure est de ne pas se laisser dépasser par ses propres modèles et de savoir sélectionner, parmi toutes ces perspectives, celles qui sont susceptibles de conduire aux résultats les plus pertinents. Pour cela, nous renvoyons le lecteur aux conclusions que nous avons énoncées dans le chapitre I quant au caractère itératif des modèles : le risque est grand de se laisser entraîner par le “paraître” des modèles et de glisser involontairement vers la construction de modèles purement représentatifs. Une fois encore, nous voudrions insister sur la nécessité de soumettre le développement d’un modèle à la formulation préalable d’un questionnement scientifique clair, même si cela peut nous donner le sentiment de passer à côté de bon nombre de modèles potentiels. Il importe en effet de toujours garder à l’esprit que ce n’est pas la ressemblance avec un objet A qui fait de A^* un modèle. Il n’acquerra ce statut que s’il nous permet effectivement de répondre à une question d’intérêt sur l’objet A .

Enfin, avant de refermer ce mémoire, j’aimerais apporter un éclairage plus personnel sur ce parcours, sur ces aller-retours entre informatique et biologie. J’aimerais dire ici tout l’enthousiasme que m’inspire cette approche du vivant par la modélisation et ces liens qui se tissent progressivement entre disciplines. J’aimerais dire aussi le respect et l’humilité dont doit faire preuve le modélisateur lorsqu’il s’approche ainsi d’une discipline. Cette humilité est celle de celui qui apprivoise lentement – il faut être très patient, répondit le renard – une autre discipline. C’est l’humilité nécessaire à celui qui souhaite accéder à la pluridisciplinarité et qui doit prendre le temps nécessaire. C’est l’humilité de celui qui a *appris* qu’il doit prendre le temps nécessaire ...

Tu t’assoiras d’abord un peu loin de moi, comme ça, dans l’herbe. Je te regarderai du coin de l’œil et tu ne diras rien. Le langage est source de malentendus. Mais, chaque jour, tu pourras t’asseoir un peu plus près ...

Le lendemain revint le petit prince ...

Bibliographie

- ABBOTT, L. F. et NELSON, S. B. (2000). Synaptic plasticity : taming the beast. *Nature Neuroscience*, 3 (suppl.)(11):1178–1183.
- ADAMI, C. (1998). *Introduction to Artificial Life*. Springer-Verlag, New York. 374 pages.
- ADAMI, C. (2002). What is complexity? *BioEssays*, 24(12):1085–1094.
- ADAMI, C. (2006). Digital genetics : unravelling the genetic basis of evolution. *Nature Review Genetics*, 7(2):109–118.
- ADEREM, A. (2005). Systems biology : Its practice and challenges. *Cell*, 121(4):511–513.
- AGGARWAL, K. et LEE, K. (2003). Functional genomics and proteomics as a foundation for systems biology. *Briefings in functional genomics and proteomics*, 2(3):175–184.
- ALON, U. (2007). Network motifs : theory and experimental approaches. *Nature Review Genetics*, 8(6):450–461.
- AMAR, P., COMET, J.-P., KÉPÈS, F. et NORRIS, V. (2004). *Modelling Complex Biological Systems in the Context of Genomics (proceedings of the Évry spring school)*, volume 3. Platypus Press, Paris.
- AMAR, P., KÉPÈS, F., MOLINA, F. et NORRIS, V. (2005). *Modelling Complex Biological Systems in the Context of Genomics (proceedings of the Montpellier spring school)*, volume 4. EDP Sciences, Paris.
- AMAR, P., KÉPÈS, F., NORRIS, V. et BERNOT, G. (2007). *Modelling Complex Biological Systems in the Context of Genomics (proceedings of the Évry spring school)*, volume 6. EDP Sciences, Paris.
- AMAR, P., KÉPÈS, F., NORRIS, V., BEURTON-AIMAR, M. et MAZAT, J.-P. (2006). *Modelling Complex Biological Systems in the Context of Genomics (proceedings of the Bordeaux spring school)*, volume 5. EDP Sciences, Paris.
- AMAR, P., KÉPÈS, F., NORRIS, V. et TRACQUI, P. (2002). *Modelling Complex Biological Systems in the Context of Genomics (proceedings of the Autrans spring school)*, volume 1. Génopole d'Évry, Évry.
- AMAR, P., KÉPÈS, F., NORRIS, V. et TRACQUI, P. (2003). *Modelling Complex Biological Systems in the Context of Genomics (proceedings of the Dieppe spring school)*, volume 2. Platypus Press, Paris.

- AMBLARD, F. et PHAN, D. (2006). *Modélisation et simulation multi-agents : applications pour les Sciences de l'Homme et de la Société*. Hermès Science Publications, Paris. 446 pages.
- ANCEL, L. W. et FONTANA, W. (2000). Plasticity, evolvability, and modularity in rna. *Journal of Experimental Zoology*, 288(3):242–283.
- ANCEL MEYERS, L., ANCEL, F. D. et LACHMANN, M. (2005). Evolution of genetic potential. *PLoS Computational Biology*, 1(3):236–243.
- ANDERSSON, S. G. E. et KURLAND, C. G. (1998). Reductive evolution of resident genomes. *Trends in Microbiology*, 6(7):263–268.
- ANDLER, D., FAGOT-LARGEAULT, A. et SAINT-SERNIN, B. (2002). *Philosophie des sciences (vol. I et II)*. Gallimard, Paris. 1334 pages.
- ANDREWS, S. S. et BRAY, D. (2004). Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Physiological Biology*, 1(3-4):137–151.
- BACHELARD, G. (1947). *La formation de l'esprit scientifique*. Vrin, Paris. 256 pages.
- BAER, C. F., MIYAMOTO, M. M. et DENVER, D. R. (2007). Mutation rate variation in multicellular eukaryotes : causes and consequences. *Nature Review Genetics*, 8(8):619–631.
- BANKES, S. (2002). Agent-based modeling : A revolution? *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 99:7199–7200.
- BANZHAF, W., BESLON, G., CHRISTENSEN, S., FOSTER, J. A., KÉPÈS, F., LEFORT, V., MILLER, J. F., RADMAN, M. et RAMSDEN, J. J. (2006). Guidelines : From artificial evolution to computational evolution : a research agenda. *Nature Review Genetics*, 7(9):729–735.
- BARABASI, A.-L. (2004). network biology. understanding the cell's functional organization. *Nature Review Genetics*, 5(2):101–113.
- BARBARAS, R. (2002). Francisco varela : A new idea of perception and life. *Phenomenology and the Cognitive Sciences*.
- BATT, G., ROPERS, D., de JONG, H., GEISELMANN, J., MATEESCU, R., PAGE, M. et SCHNEIDER, D. (2005). Validation of qualitative models of genetic regulatory networks by model checking : analysis of the nutritional stress response in escherichia coli. *Bioinformatics*, 21 Suppl 1:i19–28.
- BAZIN, E., GLÉMIN, S. et GALTIER, N. (2006). Population size does not influence mitochondrial genetic diversity in animals. *Science*, 312(5773):570–2.
- BECSKEI, A., KAUFMANN, B. et van OUDENAARDEN, A. (2005). Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nature Genetics*.

- BEDAU, M. et PACKARD, N. (2003). Evolution of evolvability via adaptation of mutation rates. *BioSystems*, 69(2-3):143–162.
- BENKIRANE, R. (2002). *La complexité, vertiges et promesses*. Le Pommier, Paris. 419 pages.
- BERG, H. (1993). *random walk in biology*. Princeton Press, Princeton (USA), 2ième édition. pages.
- BERNOT, G., COMET, J.-P., RICHARD, A. et GUESPIN, J. (2004). Application of formal methods to biological regulatory networks : extending thomas' asynchronous logical approach with temporal logic. *Journal of Theoretical Biology*, 229(3):339–347.
- BESLON, G. (1995). *Contrôle Sensori-Moteur par Réseaux Neuromimétiques Modulaires*. Thèse de doctorat, INSA-Lyon. 333 pages.
- BESLON, G. (2002). Coopération implicite et émergence dans des groupes d'agents. In GUILLOT, A. et DAUCÉ, E., éditeurs : *Approche Dynamique de la Cognition Artificielle*, pages 267–283. Eds. Hermès-Lavoisier, Paris.
- BESLON, G., BIENNIER, F. et FAVREL, J. (2000). Towards neuroreactivity and neural assemblies emergence : The asynchronous spike propagation model. In *Proceedings of Neural Computation (NC'2000), Berlin (Germany)*, pages 456–462.
- BESLON, G., BIENNIER, F. et HIRSBRUNNER, B. (1998). Multi-robots path planning based on implicit cooperation in robotic swarm. In *Proceeding of Autonomous Agents (AA'98), Minneapolis (USA)*, pages 39–46.
- BESLON, G. et DAUCÉ, E. (2002). Modularité et apprentissage dans les réseaux de neurones récurrents. In GUILLOT, A. et DAUCÉ, E., éditeurs : *Approche Dynamique de la Cognition Artificielle*, pages 61–80. Eds. Hermès-Lavoisier, Paris.
- BESLON, G. et SOULA, H. (2003). Laissez les vivre ! La codépendance temporelle comme clé de l'apprentissage de comportements chez un animat. *Technique et Science Informatique*, 22(2):145–170.
- BI, G. et POO, M. (1998). Synaptic modifications in cultured hippocampal neurons : Dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience*, 18(24):10464–10472.
- BI, G. et POO, M. (1999). Distributed synaptic modification in neural networks induced by patterned stimulation. *Nature*, 401:792–796.
- BLAKE, W., BALAZSI, G., KOHANSKI, M., ISAACS, F., MURPHY, K., KUANG, Y., CANTOR, C., WALT, D. et COLLINS, J. (2006). Phenotypic consequences of promoter-mediated transcriptional noise. *Molecular Cell*, 24(6):853–865.
- BONABEAU, E. (2002). Agent-based modeling : Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 99(suppl. 3):7280–7287.

- BORK, P. et SERRANO, L. (2005). Towards cellular systems in 4D. *Cell*, 121(4):507–509.
- BOTHWELL, J. H. F. (2006). The long past of systems biology. *New Phytologist*, 170(1):6–10.
- BOURGINE, P. et LESNE, A. (2006). *Morphogénèse, l'origine des formes*. Belin, échelles, Paris. 352 pages.
- BOURNEZ, C. (2001). *Une Architecture Multi-Agents Réflexive pour le Contrôle des Systèmes de Production Distribués et Hétérogènes*. Thèse de doctorat, INSA-Lyon. 192 pages.
- BOURNEZ, C., BESLON, G. et FAVREL, J. (2001). Ocean : A multi-agent market for distributed system control. In *Proceedings of the European Simulation Symposium (ESS'2001)*, Marseille, pages 934–940.
- BRAITENBERG, V. (1986). *Vehicles : Experiments in Synthetic Psychology*. The MIT Press, Cambridge (USA). 152 pages.
- BRAY, D. (2002). Bacterial chemotaxis and the question of gain. *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 99(1):7–9.
- BRAY, D., LEVIN, M. D. et MORTON-FIRTH, C. J. (1998). Receptor clustering as a cellular mechanism to control sensitivity. *Nature*, 393(6680):85–88.
- BRICMONT, J. et FRANCK, J., éditeurs (2007). *Chomsky*. Éditions de l'Herne, Paris. 356 pages.
- BROOKS, R. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2(1):14–23.
- BROOKS, R. (1991). Intelligence without representation. *Artificial Intelligence*, 47(1-3):139–159.
- BULLERWELL, C. E. et GRAY, M. W. (2004). Evolution of the mitochondrial genome : protist connections to animals, fungi and plants. *Current Opinion in Microbiology*, 7(5):528–534.
- BURKE, D. S., DE JONG, K. A., GREFENSTETTE, J. J., RAMSEY, C. L. et WU, A. S. (1998). Putting more genetics into genetic algorithms. *Evolutionary Computation*, 6:387–410.
- CAI, L., FRIEDMAN, N. et XIE, X. (2006). Stochastic protein expression in individual cells at the single molecule level. *Nature*, 440(7082):358–362.
- CARRERO, G., HENDZEL, M. J. et de VRIES, G. (2006). Modelling the compartmentalization of splicing factors. *Journal of Theoretical Biology*, 239(3):298–312.
- CHALMERS, A. F. (1998). *Qu'est-ce que la science ?* LGF/Livre de poche, Paris. 287 pages.

- CHEN, B.-S. et WANG, Y.-C. (2006). On the attenuation and amplification of molecular noise in genetic regulatory networks. *BMC Bioinformatics*, 7(1):52.
- CHEN, Y., YANG, B. et JACOBSON, K. (2004). Transient confinement zones : a type of lipid raft ? *Lipids*, 39(11):1115–1119.
- COHEN, D. (1966). Optimizing reproduction in a randomly varying environment. *Journal of Theoretical Biology*, 12(1):119–129.
- COHEN, J. E. (2004). Mathematics is biology’s next microscope, only better ; biology is mathematics’ next physics, only better. *PLoS Biology*, 2(12):2017–2023.
- COLLINOT, A. et DROGOUL, A. (1998). Using the cassiopeia method to design a soccer robot team. *Applied Artificial Intelligence (AAI) Journal*, 12(2-3):127–147.
- COULON, A., BESLON, G. et GANDRILLON, O. (2008). Large multi-protein structure modeling and simulation : the need for mesoscopic models. *Methods in Molecular Biology*, 484:537–558.
- COULON, A., SOULA, H., MAZET, O., GANDRILLON, O. et BESLON, G. (2007). Modélisation cellulaire pour l’émergence de structures multiprotéiques auto-organisées. *Technique et Science Informatiques*, 26(1-2):123–148.
- DAUMAS, F., DESTAINVILLE, N., MILLOT, C. et LOPEZ, A. (2003a). Confined diffusion without fences of a g-protein-coupled receptor as revealed by single particle tracking. *Biophysical Journal*, 84(1):356–366.
- DAUMAS, F., DESTAINVILLE, N., MILLOT, C. et LOPEZ, A. (2003b). Interprotein interactions are responsible for the confined diffusion of a g-protein-coupled receptor at the cell surface. *Biochemical Society Transactions*, 31(pt 5):1001–1005.
- DEFFUANT, G., AMBLARD, F., DUBOZ, R. et RAMAT, E. (2003). Une démarche expérimentale pour la simulation individu-centrée. In *Actes des 9ièmes Journées de Rochebrune : Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels*, pages 45–64.
- DELAHAYE, J.-P. (1999). *Information, complexité et hasard*. Hermès Sciences, Paris. 178 pages.
- DELAHAYE, J.-P. (2006). *Vertiges de la complexité*. Belin, Paris. 256 pages.
- DEMAZEAU, Y. (1995). From interactions to collective behaviour in agent-based systems. In *European Conference on Cognitive Science*, pages 117–132.
- DI PAOLO, E. A. (2003). Evolving spike-timing-dependent plasticity for single-trial learning in robots. *Philosophical transactions Series A, Mathematical, physical, and engineering sciences*, 361(1811):2299–319.
- DI PAOLO, E. A. et IIZUKA, H. (2008). How (not) to model autonomous behaviour. *BioSystems*, 91(2):409–23.

- DI VENTURA, B., LEMERLE, C., MICHALODIMITRAKIS, K. et SERRANO, L. (2006). From in vivo to in silico biology and back. *Nature*, 443(7111):527–533.
- DOBZHANSKY, T. (1973). Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher*, 35(3):125–129.
- DRAKE, J. W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 88:7160–7164.
- DRAKE, J. W. (1999). The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. *Annals of the New-York Academy of Science.*, 870:100–107.
- DROGOUL, A., VANBERGUE, D. et MEURISSE, T. (2002). Multi-agent based simulation : Where are the agents ? In *Multi-Agent-Based Simulation II, Proceedings of MABS 2002*, pages 89–104.
- DUHEM, P. (1908). *Sauver les apparences. Essai sur la notion de physique théorique de Platon à Galilée*. Réédition Vrin (2004), Paris. 156 pages.
- DUNDR, M., HEBERT, M. D., KARPOVA, T. S., STANEK, D., XU, H., SHPARGEL, K. B., MEIER, U. T., NEUGEBAUER, K. M., MATERA, A. G. et MISTELI, T. (2004). In vivo kinetics of Cajal body components. *Journal of Cellular Biology*, 164(6):831–842.
- EARL, D. J. et DEEM, M. W. (2004). Evolvability is a selectable trait. *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 101(32):11531–6.
- EIGEN, M. (1971). Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523.
- ELENA, S. et LENSKI, R. (2003). Microbial genetics : Evolution experiments with microorganisms : the dynamics and genetic bases of adaptation. *Nature Reviews Genetics*, 4:457–469.
- ELOWITZ, M., LEVINE, A., SIGGIA, E. et SWAIN, P. (2002). Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186.
- FÉLIX, M.-A. et WAGNER, A. (2008). Robustness and evolution : concepts, insights and challenges from a developmental model system. *Heredity*, 100(2):132–140.
- FERBER, J. (1995). *Les systèmes multi-agents, vers une intelligence collective*. InterEdition, Paris. 522 pages.
- FERBER, J. et GUTKNECHT, O. (1998). A meta-model for the analysis and design of organizations in multi-agent systems. In *Proc. of 3rd Int. Conf on Multi-Agent Systems*, pages 128–135.
- FEYERABEND, P. (1975). *Contre la méthode*. Seuil, Paris. 349 pages.
- FINO, E., DENIAU, J.-M. et VENANCE, L. (2008). Cell-specific spike-timing-dependent plasticity in gabaergic and cholinergic interneurons in corticostriatal rat brain slices. *The Journal of Physiology*, 586(1):265–282.

- FLOREANO, D. et MATTIUSI, C. (2001). Evolution of spiking neural controllers for autonomous vision-based robots. *Lecture Notes in Computer Sciences (LNCS)*, 2217:38–61.
- FLOREANO, D., MITRI, S., MAGNENAT, S. et KELLER, L. (2007). Evolutionary conditions for the emergence of communication in robots. *Current Biology*, 17(6):514–519.
- FOX-KELLER, E. (1999). *Le rôle des métaphores dans les progrès de la biologie*. Les empêcheurs de penser en rond, Paris. 158 pages.
- FOX KELLER, E. (2003). *Le siècle du gène*. Gallimard, Paris. 173 pages.
- FOX KELLER, E. (2004). *Expliquer la vie, modèles, métaphores et machines en biologie du développement*. Gallimard, Paris. 418 pages.
- FOX KELLER, E. (2005). Revisiting "scale-free" networks. *Bioessays*, 27(10):1060–1068.
- FREEMAN, W. (1991). The physiology of perception. *Scientific American*, 264(2):78–85.
- FREEMAN, W. (2000). Brains create macroscopic order from microscopic disorder by neurodynamics in perception. In ARHEM, P., BLOMBERG, C. et LILJENSTRÖM, H., éditeurs : *Disorder versus Order in Brain Function : Essays in Theoretical Neurobiology*, pages 205–219. WorldScientific, Singapour.
- FUJITA, A., CHENG, J., HIRAKAWA, M., FURUKAWA, K., KUSUNOKI, S. et FUJIMOTO, T. (2007). Gangliosides GM1 and GM3 in the living cell membrane form clusters susceptible to cholesterol depletion and chilling. *Molecular Biology of the Cell*, 18(6): 2112–2122.
- FUJIWARA, T., RITCHIE, K., MURAKOSHI, H., JACOBSON, K. et KUSUMI, A. (2002). Phospholipids undergo hop diffusion in compartmentalized cell membrane. *The Journal of Cell Biology*, 157(6):1071–1081.
- GARBAY, C. (2003). Les sciences du traitement de l'information comme pivot de l'interdisciplinarité : une vision systémique. *Information-Interaction-Intelligence*, 2(2). 21 pages.
- GERSTNER, W. (1999). Rapid signal transmission by populations of spiking neurons. In *Proceedings of the Ninth International Conference on Artificial Neural Networks (ICANN'99), Edinburgh (Scotland)*, pages 7–12.
- GERSTNER, W. et KISTLER, W. M. (2002). Mathematical formulations of hebbian learning. *Biological cybernetics*, 87(5-6):404–15.
- GIL, R., SABATER-MUNOZ, B., LATORRE, A., SILVA, F. J. et MOYA, A. (2002). Extreme genome reduction in buchnera spp. : toward the minimal genome needed for symbiotic life. *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 99(7):4454–4458.
- GILLESPIE, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434.

- GOLDENFELD, N. et KADANOFF, L. P. (1999). Simple lessons from complexity. *Science*, 284(5411):87–89.
- GOMEZ-VALERO, L., LATORRE, A. et SILVA, F. J. (2004). The evolutionary fate of non-functional DNA in the bacterial endosymbiont *buchnera aphidicola*. *Molecular Biology and Evolution*, 21(11):2172–2181.
- GOODWIN, J. S., DRAKE, K. R., REMMERT, C. L. et KENWORTHY, A. K. (2005). Ras diffusion is sensitive to plasma membrane viscosity. *Biophysical Journal*, 89(2):1398–1410.
- GORSKI, S. et MISTELI, T. (2005). Systems biology in the cell nucleus. *Journal of Cell Science*, 118(Pt 18):4083–4092.
- GOUYON, P.-H., HENRY, J.-P. et ARNOULD, J. (1997). *Les avatars du gène*. Belin, Paris. 335 pages.
- GRAY, C., KÖNING, P., ENGEL, A. et SINGER, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *nature*, 338(6213):334–337.
- GREGORY, T. R. (2003). Is small indel bias a determinant of genome size? *Trends in Genetics*, 19(9):485–488.
- GREGORY, T. R. (2004). Insertion-deletion biases and the evolution of genome size. *Gene*, 324:15–34.
- GRIBBIN, J. (2006). *Simplicité profonde, le chaos, la complexité et l'émergence de la vie*. Flammarion, Paris. 357 pages.
- GRIMM, V. (1999). Ten years of individual-based modelling in ecology : what have we learned and what could we learn in the future? *Ecological Modelling*, 115:129–148.
- GRIMM, V. (2002). Visual debugging : a way of analyzing, understanding, and communicating bottom-up simulation models in ecology. *Natural Resource Modeling*, 15(1):23–38.
- GRIMM, V. et RAILSBACK, S. F. (2005). *Individual-Based Modeling and Ecology*. Princeton University Press, Princeton. 480 pages.
- GRUENE-YANOFF, T. (2006). Agent-based simulation, generative science, and its explanatory claims. In *Proceedings of Models and Simulations conference, Paris (France)*. (Available on Internet : <http://philsci-archive.pitt.edu/archive/00002784>). 19 pages.
- GRÉCO, P. (2001). Jean Piaget. In *Dictionnaire des philosophes*, pages 1210–1219. Albin Michel/Encyclopædia Universalis, Paris (France), 2ième édition.
- GUILLOT, A. et DAUCÉ, E. (2002). *Approche dynamique de la cognition*. Traité des sciences cognitives. Hermes Science Publications, Paris. 316 pages.
- HEBB, D. (1949). *The Organization of Behavior*. J. Wiley and Sons, New York.

- HUERTA, R., VARONA, P., RABINOVICH, M. et ABARBANEL, H. (2001). Topology selection by chaotic neurons of a pyloric central pattern generator. *Biological Cybernetics*, 84(1):L1–8.
- IDEKER, T., GALITSKI, T. et HOOD, L. (2001). A new approach to decoding life : systems biology. *Annual Review of Genomics and Human Genetics*, 2:343–372.
- IDEKER, T. et LAUFFENBURGER, D. (2003). Building with a scaffold : emerging strategies for high- to low-level cellular modeling. *Trends in biotechnology*, 21(6):255–262.
- IDEKER, T., WINSLOW, L. R. et LAUFFENBURGER, D. (2006). Bioengineering and systems biology. *Annals of Biomedical Engineering*, 34(2):257–264.
- JACOBSON, K., SHEETS, E. D. et SIMSON, R. (1995). Revisiting the fluid mosaic model of membranes. *Science*, 268(5216):1441–1442.
- JEONG, H., MASON, S. P., BARABASI, A.-L. et OLTVAI, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, 411(6833):41–42.
- KASHTAN, N. et ALON, U. (2005). Spontaneous evolution of modularity and network motifs. *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 102(39):13773–13778.
- KEPES, F. (2007). Formalismes mathématiques : des standards de fait. In ROUX, M., éditeur : *Biologie systémique, standards et modèles*, pages 199–214. Omniscience, Paris.
- KIRSCHNER, M. et GERHART, J. (1998). Evolvability. *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 95(15):8420–8427.
- KIRSHNER, M. W. (2005). The meaning of systems biology. *Cell*, 121(4):503–504.
- KITANO, H. (2000). Perspectives on systems biology. *New Generation Computing*, 18(3):199–216.
- KITANO, H. (2001). *Foundations of Systems Biology*. MIT Press, Cambridge (USA). 297 pages.
- KITANO, H. (2002). Systems biology : a brief overview. *Science*, 295(5560):1662–1664.
- KNIBBE, C. (2006). *Structuration des génomes par sélection indirecte de la variabilité mutationnelle, une approche de modélisation et de simulation*. Thèse de doctorat, INSA-Lyon. 174 pages.
- KNIBBE, C., COULON, A., FAYARD, J.-M. et BESLON, G. (2007a). A long term evolutionary pressure on the amount of noncoding DNA. *Molecular Biology and Evolution*, 24(10):2344–2353.
- KNIBBE, C., FAYARD, J.-M. et BESLON, G. (2008). The topology of the protein network influences the dynamics of gene order : From systems biology to a systemic understanding of evolution. *Artificial Life*, 14(1):149–156.

- KNIBBE, C., MAZET, O., CHAUDIER, F., FAYARD, J.-M. et BESLON, G. (2007b). Evolutionary coupling between the deleteriousness of gene mutations and the amount of non-coding sequences. *Journal of Theoretical Biology*, 244(4):621–630.
- KUHN, T. S. (1983). *La structure des révolutions scientifiques*. Flammarion, Paris. 284 pages.
- KUSSELL, E. et LEIBLER, S. (2005). Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309(5743):2075–2078.
- KUTTLER, C. (2007). *Modélisation de l'expression génétique bactérienne dans un calcul stochastique à objets concurrents*. Thèse de doctorat, Université des Sciences et Technologies de Lille.
- LE MOIGNE, J.-L. (1977). *La théorie du système général. Théorie de la modélisation*. Presses Universitaires de France, Paris.
- LEFORT, V. (2007). *Evolution de second ordre et algorithmes évolutionnaires : l'algorithme RBF-Gene*. Thèse de doctorat, INSA-Lyon. 154 pages.
- LEGAY, J.-M. (1997). *L'expérience et le modèle : un discours sur la méthode*. INRA Editions, Paris. 112 pages.
- LEGAY, J.-M. et LEBRETON, J.-D. (1973). *La méthode des modèles, état actuel de la méthode expérimentale. Introduction aux modèles mathématiques de la dynamique des populations. Introduction à l'étude des modèles à compartiments*. Informatique et biosphère, Paris.
- LENSKI, R. E., BARRICK, J. E. et OFRIA, C. (2006). Balancing robustness and evolvability. *PLoS Biology*, 4(12):2190–2192.
- LIPKOW, K., ANDREWS, S. S. et BRAY, D. (2005). Simulated diffusion of phosphorylated chey through the cytoplasm of escherichia coli. *Journal of Bacteriology*, 187(1):45–53.
- LIPPINCOTT-SCHWARTZ, J., SNAPP, E. et KENWORTHY, A. (2001). Studying protein dynamics in living cells. *Nature Reviews Molecular Cell Biology*, 2(6):444–456.
- LIU, E. T. (2005). Systems biology, integrative biology, predictive biology. *Cell*, 121(4):505–506.
- LOEW, L. et SCHAFF, J. (2001). The virtual cell : a software environment for computational cell biology. *Trends in Biotechnology*, 19(10):401–406.
- LYNCH, M. (2006a). Mutation pressure and the evolution of organelle genomic architecture. *Science*, 311(5768):1727–1730.
- LYNCH, M. (2006b). Streamlining and simplification of microbial genome architecture. *Annual Review Microbiology*, 60:327–349.
- LYNCH, M. et CONERY, J. S. (2003). The origins of genome complexity. *Science*, 302:1401–1404.

- MAAS, W. et BISHOP, C., éditeurs (2001). *Pulsed Neural Networks*. MIT Press, Cambridge (USA). 377 pages.
- MACAL, C. M. et NORTH, M. J. (2005). Tutorial on agent-based modeling and simulation. *In WSC05 : Proceedings of the 37th conference on Winter simulation, Orlando (USA)*, pages 2–15.
- MACAL, C. M. et NORTH, M. J. (2006). Tutorial on agent-based modeling and simulation part 2 : how to model with agents. *In WSC06 : Proceedings of the 38th conference on Winter simulation, Monterey (USA)*, pages 73–83.
- MAR, D., CHOW, C., GERSTNER, W., ADAMS, R. et COLLINS, J. (1999). Noise shaping in populations of coupled model neurons. *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 96(18):10450–10455.
- MELLO, B. A., SHAW, L. et TUZ, Y. (2004). Effects of receptor interaction in bacterial chemotaxis. *Biophysical journal*, 87(3):1578–1595.
- MENDES, P., SHA, W. et YE, K. (2003). Artificial gene networks for objective comparison of analysis algorithms. *Bioinformatics*, 19(Suppl 2):122–129.
- MERLEAU-PONTY, M. (1942). *La structure du comportement*. Presses Universitaires de France, Paris. 248 pages.
- MERLEAU-PONTY, M. (1945). *Phénoménologie de la perception*. Gallimard, Paris. 531 pages.
- MEYRAND, P., SIMMERS, J. et MOULIN, M. (1991). Construction of a pattern generating circuit with neurons of different networks. *Nature*, 351(6321):60–63.
- MICHEL, F. (2004). *Formalisme, outils et éléments méthodologiques pour la modélisation et la simulation multi-agents*. Thèse de doctorat, Université Montpellier II. 273 pages.
- MILLER, G. (1995). Artificial life as theoretical biology : How to do real science with computer simulation. Rapport technique 378, School of Cognitive and Computing Sciences, University of Sussex, Brighton, UK.
- MINSKY, M. (1995). Matter, mind and models. Rev. version of the essay in "Semantic Information Processing", ed. Marvin Minsky. Cambridge MA : MIT Press.
- MISTELI, T. (2001). Protein dynamics : implication for nuclear architecture and gene expression. *Science*, 291(5):843–847.
- MISTELI, T. (2005). Concepts in nuclear architecture. *BioEssays*, 27(5):477–487.
- MISTELI, T. (2007). Beyond the sequence : cellular organization of genome function. *Cell*, 128(4):787–800.
- MORANGE, M. (2005). *Les secrets du vivant : contre la pensée unique en biologie*. La Découverte, Paris. 228 pages.

- MORIN, E. (1977). *La méthode, Tome 1 : La nature de la nature*. Seuil, Paris. 399 pages.
- MOXON, R., BAYLISS, C. et HOOD, D. (2006). Bacterial contingency loci : the role of simple sequence dna repeats in bacterial adaptation. *Annual Review of Genetetics*, 40:307–333.
- MULLER, J., WOOLDRIDGE, M. et JENNINGS, N., éditeurs (1997). *Intelligent agents III : agents theories, architectures and languages*. Springer, Berlin. 401 pages.
- MURPHY, R. R. (2000). *Introduction to AI robotics*. MIT Press, Cambridge, USA. 400 pages.
- NAGEL, T. (1981). What is it like to be a bat? *In DENNETT, D. C. et HOFSTADTER, D. R., éditeurs : The Mind's I : Fantasies and Reflections on Self and Soul*. Basic Books.
- NICOLESCU, B. (1996). *La transdisciplinarité, Manifeste*. Éditions du Rocher, Monaco. 232 pages.
- NIEMELÄ, P. S., OLLILA, S., HYVÖNEN, M. T., KARTTUNEN, M. et VATTULAINEN, I. (2007). Assessing the nature of lipid raft membranes. *PLoS Computational Biology*, 3(2):304–312.
- NILSSON, A. I., KOSKINIEMI, S., ERIKSSON, S., KUGELBERG, E., HINTON, J. C. D. et ANDERSSON, D. I. (2005). Bacterial genome size reduction by experimental evolution. *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 102(34):12112–12116.
- NOLFI, S. (1997). Using emergent modularity to develop control systems for mobile robots. *Adaptive Behavior*, 5(3-4):343–363.
- NOUVEL, P. (2002). *Enquête sur le concept de modèle*. Presses Universitaires de France, Paris. 246 pages.
- OCHMAN, H. (2005). Genomes on the shrink. *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 102(34):11959–60.
- OFTA (2004). *Systèmes Multi-Agents*. Editions Tec&Doc, Paris. 289 pages.
- O'NEILL, B. (2003). Digital evolution. *PLoS Biology*, 1(1):11–14.
- PALSSON, B. (2000). The challenges of in silico biology. *Nature Biotechnology*, 18(11): 1147–1150.
- PARROCHIA, D. (2000). L'expérience dans les sciences : modèles et simulations. *In* MICHAUD, Y., éditeur : *Qu'est-ce que la vie*, pages 193–203. Odile Jacob.
- PAVÉ, A. (1994). *La modélisation en biologie et en écologie*. Aléa, Lyon. 459 pages.
- PAYET, D., MEDOC, J., RALAMBONDRAIN, T., GUERRIN, F. et COURDIER, R. (2005). Outils d'observation et d'analyse de simulations multi-agents : l'expérience de la plateforme Geamas/Biomass. *In Conference on Multi-agent modelling for environmental management, Bourg Saint Maurice - Les Arcs, France*.

- PEDERSON, T. (2001). Protein mobility within the nucleus – what are the right moves? *Cell*, 104(5):635–638.
- PETROV, D. A. (2002). Mutational equilibrium model of genome size evolution. *Theoretical Population Biology*, 61(4):531–44.
- PETROV, D. A., SANGSTER, T. A., JOHNSTON, J. S., HARTL, D. L. et SHAW, K. L. (2000). Evidence for DNA loss as a determinant of genome size. *Science*, 287(5455):1060–1062.
- PFEIFER, R. et BONGARD, J. (2006). *How the Body Shapes the Way We Think*. MIT Press, Cambridge (USA). 394 pages.
- PIAGET, J. (1968). *Le structuralisme*. Que sais-je. PUF, Paris. 128 pages.
- PIKE, L. J. (2004). Lipid rafts : heterogeneity on the high seas. *Biochemical Journal*, 378(2):281–292.
- POMERLEAU, D. (1993). *Neural Network Perception for Mobile Robot Guidance*. Kluwer Academic Publisher, Boston (USA). 191 pages.
- RESNICK, M. (1995). *Turtles, termites and traffic jams*. MIT Press, Cambridge (USA). 181 pages.
- REY, A., TOMI, M., HORÉ, T. et TANET, C. (1983). *Dictionnaire historique de la langue française*. Le Robert, Paris. 4304 pages.
- REYNOLDS, C. (1987). Flocks, herds, and schools : A distributed behavioral model. *Computer Graphics*, 21(4):25–34.
- RONALD, E. et SIPPER, M. (2001). Surprise versus unsurprise : Implications of emergence in robotics. *Robotics and Autonomous Systems*, 37:19–24.
- RONCO, E. et GAWTHROP, P. (1995). Modular neural networks : a state of the art. Rapport technique CSC-95026, Center for System and Control, University of Glasgow (UK). 22 pages.
- ROUX, M. (2007). *Biologie systémique, standards et modèles*. Omniscience/Ecrin, Paris. 288 pages.
- SAMAN, P. et DELBRUCK, M. (1975). Brownian motion in biological membranes. *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 72(8):3111–3113.
- SANCHEZ-DEHESA, Y., CERF, L., PENA, J. M., BOULICAUT, J.-F. et BESLON, G. (2007). Artificial Regulatory Networks Evolution. In *Proceedings of the first Int. Workshop on Machine Learning in Systems Biology (MLSB'2007), Evry (France)*, pages 47–52.
- SANJUAN, R., MOYA, A. et ELENA, S. (2004). The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 101(22):8396–8401.
- SCHWEITZER, F. (2003). *Brownian Agents and Active Particles, Collective Dynamics in the Natural and Social Sciences*. Springer, Berlin. 420 pages.

- SHARKEY, A. (1996). On combining artificial neural nets. *Connection Science*, 8:299–313.
- SHIMIZU, T., AKSENOV, S. et BRAY, D. (2003). A spatially extended stochastic model of the bacterial chemotaxis signalling pathway. *Journal of Molecular Biology*, 329(2):291–309.
- SIEBRASSE, J. P., GRÜNWARD, D. et KUBITSCHECK, U. (2007). Single-molecule tracking in eukaryotic cell nuclei. *Analytical and Bioanalytical Chemistry*, 387(1):41–44.
- SIMON, H. A. (2004). *Les sciences de l'artificiel*. Folio essais, Paris, 3ième édition. 464 pages.
- SIMONS, K. et IKONEN, E. (1997). Functional rafts in cell membranes. *Nature*, 387(6633):569–572.
- SIMPSON, M. L. (2004). Rewiring the cell : synthetic biology moves towards higher functional complexity. *Trends in Biotechnology*, 22(11):555–7.
- SINGER, S. et NICHOLSON, G. (1972). The fluid mosaic model of the structure of cell membranes. *Science*, 175(23):720–731.
- SINGER, W. (1993). Synchronization of cortical activity and its putative role in information processing and learning. *Annual Review of Physiology*, 55:349–374.
- SKARDA, C. et FREEMAN, W. (1987). How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences*, 10:161–195.
- SMITH, J. (1992). Byte-sized evolution. *Nature*, 355(6363):772–773.
- SOULA, H. (2005). *Dynamique et plasticité dans les réseaux de neurones à impulsion ; étude du couplage temporel réseau/agent/environnement*. Thèse de doctorat, INSA-Lyon. 180 pages.
- SOULA, H., ALWAN, A. et BESLON, G. (2005a). Learning at the edge of chaos : Temporal coupling of spiking neuron controller for autonomous robotics. *In AAAI Spring symposia on developmental robotics, Stanford (USA)*.
- SOULA, H. et BESLON, G. (2006). Spike-timing dependent plasticity learning for visual-based obstacles avoidance. *Lecture Notes in Artificial Intelligence*, 4095:335–345.
- SOULA, H., BESLON, G. et FAVREL, J. (2001). Controlling an animat with a self-organized modular neural model. *In Proceedings of the European Workshop on Learning Robots (EWLR'2001), Prague (Czech Republic)*, pages 39–46.
- SOULA, H., BESLON, G. et FAVREL, J. (2003). Evolving spiking neurons nets to control an animat. *In Proceedings of the International Conference on Artificial Neural Nets and Genetic Algorithms (ICANNGA'2003), Roannes (France)*, pages 193–197.
- SOULA, H., BESLON, G. et MAZET, O. (2006). Spontaneous dynamics of asymmetric random recurrent spiking neural networks. *Neural Computation*, 18(1):60–79.

- SOULA, H., ROBARDET, C., PERRIN, F., GRIPON, S., BESLON, G. et GANDRILLON, O. (2005b). Modeling the emergence of multi-protein dynamic structures by principles of self-organization through the use of 3DSpi, a multi-agent-based software. *BMC Bioinformatics*, 6(228).
- SPECTOR, D. L. (2001). Nuclear domains. *Journal of Cell Science*, 114(Pt 16):2891–2893.
- SPORNS, O., HONEY, C. J. et KÖTTER, R. (2007). Identification and classification of hubs in brain networks. *PLoS ONE*, 2(10):e1049.
- SPORNS, O., TONONI, G. et KÖTTER, R. (2005). The human connectome : A structural description of the human brain. *PLoS Computational Biology*, 1(4):e42.
- STENGERS, I. (1997). *Cosmopolitique*, volume 2. La Découverte, Paris. 402 pages.
- STENGERS, I. et BENSUADE-VINCENT, B. (2003). *100 mots pour commencer à penser les sciences*. Les Empêcheurs de penser en rond, Paris. 412 pages.
- STEUER, R., ZHOU, C. et KURTHS, J. (2003). Constructive effects of fluctuations in genetic and biochemical regulatory systems. *Biosystems*, 72(3):241–251.
- STEWART, E., MADDEN, R., PAUL, G. et TADDEI, F. (2005). Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biology*, 3(2):0295–0300.
- STEWART, J. (2004). *La vie existe-t-elle ? Réconcilier génétique et biologie*. Vuibert, Paris. 152 pages.
- STRUHL, K. (1999). Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell*, 9(1):1–4.
- SUTTON, R. et BARTO, A. (1998). *Reinforcement Learning : An Introduction*. MIT Press, Cambridge (USA). 322 pages.
- SWAIN, P. S. et LONGTIN, A. (2006). Noise in genetic and neural networks. *Chaos (Woodbury, NY)*, 16(2):026101.1–6.
- TADDEI, F., RADMAN, M., MAYNARD-SMITH, J., TOUPANCE, B., GOUYON, P.-H. et GODELLE, B. (1997). Role of mutator alleles in adaptive evolution. *Nature*, 387(6634):700–702.
- TENAILLON, O., TOUPANGE, B., LE NAGARD, H., TADDEI, F. et GODELLE, B. (1999). Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. *Genetics*, 152:485–493.
- THATTAI, M. et OUDENAARDEN, A. V. (2004). Stochastic gene expression in fluctuating environments. *Genetics*, 167(1):523–530.
- THORPE, S., FIZE, D. et MARLOT, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582):520–522.

- TOMITA, M. (2001). Whole-cell simulation : a grand challenge of the 21st century. *Trends in biotechnology*, 19(6):205–210.
- TONONI, G. et EDELMAN, G. M. (1998). Consciousness and complexity. *Science*, 282: 1846–1851.
- TROISI, A., WONG, V. et RATNER, M. A. (2005). An agent-based approach for modeling molecular self-organization. *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 102(2):255–60.
- UHRMACHER, A., DEGENRING, D. et ZEIGLER, B. (2005). Discrete event multi-level models for systems biology. *Lecture Notes in Computer Science*, 3380:66–89.
- VAN VALEN, L. (1973). A new evolutionary law. *Evolutionary Theory*, 1(1):1–30.
- van ZON, J. S., MORELLI, M. J., TĂNASE-NICOLA, S. et ten WOLDE, P. R. (2006). Diffusion of transcription factors can drastically enhance the noise in gene expression. *Biophysical Journal*, 91(12):4350–67.
- VANRULLEN, R. et THORPE, S. (2001). The time course of visual processing : from early perception to decision-making. *Journal of Cognitive Neuroscience*, 13(4):454–461.
- VARELA, F. (1996). *Invitation aux sciences cognitives*. Éditions du Seuil, Paris, 2ième édition. 126 pages.
- VARELA, F., LACHAUX, J.-P., RODRIGUEZ, E. et MARTINERIE, J. (2001). The brainweb : Phase synchronization an large-scale integration. *Nature Review Neuroscience*, 2(4): 229–239.
- VARELA, F., ROCH, E. et THOMPSON, E. (1992). *L'inscription corporelle de l'esprit*. Seuil, Paris. 377 pages.
- VARENNE, F. (2001). What does a computer simulation prove ? The case of plant modeling at CIRAD (France). In *Proceedings of the European Simulation Symposium (ESS'2001)*, Marseille, pages 549–554.
- VARENNE, F. (2003). La simulation conçue comme expérience concrète. In *Actes des 9ièmes Journées de Rochebrune : Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels*, pages 299–313.
- VARENNE, F. (2004). *Le destin des formalismes : à propos de la forme des plantes. Pratiques et épistémologie des modèles face à l'ordinateur*. Thèse de doctorat, Université Lumière - Lyon II. 836 pages.
- VARENNE, F. (2006). *Les notions de métaphore et d'analogie dans les épistémologies des modèles et des simulations*. Petra éditions, Paris. 97 pages.
- VARENNE, F. (2007). *Du modèle à la simulation informatique*. Vrin, Paris. 256 pages.

- VEENING, J.-W., STEWART, E. J., BERNGRUBER, T. W., TADDEI, F., KUIPERS, O. P. et HAMOEN, L. W. (2008). Bet-hedging and epigenetic inheritance in bacterial cell development. *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 105(11):4393–4398.
- VERGNIUX, A. (2003). *L'explication dans les sciences*. De Boeck, Bruxelles. 240 pages.
- WAGNER, G. P., PAVLICEV, M. et CHEVERUD, J. M. (2007). The road to modularity. *Nature Review Genetics*, 8(12):921–31.
- WALLISER, B. (1977). *Systèmes et modèles. Introduction critique à l'analyse des systèmes*. Seuil, Paris. 248 pages.
- WATTS, J. (1998). Animats : computer-simulated animals in behavioral research. *Journal of Animal Science*, 76(10):2596–2604.
- WEBB, B. et CONSI, T., éditeurs (2001). *Biorobotics*. MIT Press, Cambridge (USA). 208 pages.
- WHITESIDES, G. M. et ISMAGILOV, R. F. (1999). Complexity in chemistry. *Science*, 284(5411):89–91.
- WHITLEY, D. (1989). The GENITOR algorithm and selection pressure : Why rank-based allocation of reproductive trials is best. In *Proceedings of the 3rd International Conference on Genetic Algorithms, Fairfax (USA)*, pages 116–121.
- WILKE, C. O., WANG, J. L., OFRIA, C., LENSKI, R. E. et ADAMI, C. (2001). Evolution of digital organisms at high mutation rates leads to the survival of the flattest. *Nature*, 412(6844):331–333.
- WITTEN, T. et SANDER, L. (1983). Diffusion-limited aggregation. *Physical Review B*, 27(9):5686–5697.
- WOLF, D., VAZIRANI, V. et ARKIN, A. (2005). Diversity in times of adversity : probabilistic strategies in microbial survival games. *Journal of Theoretical Biology*, 234(2):227–253.
- WOODIN, M., GANGULY, K. et POO, M.-m. (2003). Coincident pre- and postsynaptic activity modifies gabaergic synapses by postsynaptic changes in Cl⁻ transporter activity. *Neuron*, 39(5):807–820.
- WORDEN, R. (1995). A speed limit for evolution. *Journal of Theoretical Biology*, 176(1):137–152.
- ZADEH, L. (1978). Fuzzy sets as a basis for theory of possibility. *Fuzzy Sets and Systems*, 1:3–28.
- ZELDOVICH, K., CHEN, P., SHAKHNOVITCH, B. et SHAKHNOVICH, E. (2007). A first-principle model of early evolution : Emergence of gene families, species, and preferred protein folding. *PLoS Computational Biology*, 3(7):0001–0015.

ZEMIRLINE, A., BALLETT, P., MARCÉ, L., AMAR, P., BERNOT, G., DELAPLACE, F., GIAVITTO, J.-L., MICHEL, O., DELOSME, J.-M., INCITTI, R., BOURGINE, P., GODIN, C., KÉPÈS, F., TRACQUI, P., NORRIS, V., GUESPIN, J., DEMARTY, M. et RIPOLL, C. (2002). Cellular automata, reaction-diffusion and multiagents systems for artificial cell modelling. *In Modelling and simulation of biological processes in the context of genomics, Autrans, France*, pages 257–280.

ZWIRN, H. P. (2006). *Les systèmes complexes*. Odile Jacob, Paris. 224 pages.

Annexe A

Notice biographique

1 Informations Personnelles

Dr. Guillaume Beslon

INSA, Département Informatique,
Bâtiment Blaise Pascal (501) - Bureau No. 501.325,
69621 Villeurbanne Cédex, France

Tel. : +33-(0)4.72.43.84.87

Fax. : +33-(0)4.72.43.85.18

E-mail : guillaume.beslon@liris.cnrs.fr

<http://liris.cnrs.fr/gbeslon/>

Né le 13 mai 1969 à Caen (14)
Nationalité Française

2 Situation professionnelle

Maître de Conférences au département Informatique de l'Institut National des Sciences Appliquées de Lyon (INSA-Lyon - <http://www.insa-lyon.fr>). Membre de l'équipe Turing du laboratoire d'Informatique en Images et Systèmes d'Information (LIRIS, UMR CNRS 5205 - <http://liris.cnrs.fr>) et directeur adjoint de l'Institut Rhône-Alpin des Systèmes Complexes (IXXI - <http://www.ixxi.fr>). Titulaire de la Prime d'Encadrement Doctoral et de recherche (PEDR) depuis 2007.

3 Cursus universitaire – Expérience professionnelle

depuis 1998 : Maître de Conférences (titulaire depuis 1999) à l'Institut National des Sciences Appliquées de Lyon (département Informatique). Enseignement de l'architecture des ordinateurs, de la gestion de projets et de l'informatique bioinspirée ; recherche en informatique bioinspirée et bioinformatique. Rattaché au laboratoire de Productique et d'Informatique des Systèmes Manufacturier (PRISMa) de 1998 à 2006 puis au Laboratoire d'Informatique en Images et Systèmes d'information (LIRIS), équipe Turing (depuis octobre 2006).

1997-1998 : Attaché Temporaire d'Enseignement et de Recherche à l'Institut National des Sciences Appliquées de Lyon (département Informatique, laboratoire PRISMa).

Enseignement de l'architecture des ordinateurs; recherche en informatique bioinspirée.

1997 : Maître Assistant à l'Institut d'Informatique de l'Université de Fribourg (IIUF), laboratoire de Parallélisme et d'Intelligence Artificielle (PAI). Travaux sur la coordination d'agents réactifs dans un atelier de production. Enseignement de l'architecture des ordinateurs.

1995-1996 : Stage Post-Doctoral au Centre de Recherche en Automatique de Nancy (CRAN), dans le cadre du Service National actif (Objecteur de Conscience). Travaux en reconnaissance de caractères par réseaux de neurones modulaires et en ingénierie biomédicale.

1995 : Doctorat en Ingénierie Informatique au Laboratoire de "Productique et d'Informatique des Systèmes Manufacturiers" (PRISMa) de l'INSA de Lyon ; mention Très-Honorable.

1992 : Diplôme d'Études Approfondies en ingénierie informatique au Laboratoire d'Informatique des Systèmes de Production Industrielle (LISPI).

1992 : Diplôme d'ingénieur en informatique à l'Institut National des Sciences Appliquées de Lyon (INSA-Lyon)

4 Activités Scientifiques

Mes activités scientifiques se situent à la frontière entre l'informatique, les sciences du vivant et les sciences de la complexité. Plus précisément, je m'intéresse à la modélisation individu-centrée de systèmes biologiques complexes dans le double objectif de développer de nouvelles approches en informatique bio-inspirée (essentiellement en évolution artificielle) et de développer des modèles explicatifs en collaboration avec des biologistes ("Biologie In Silico"). Dans ce cadre, mes activités sont principalement centrées sur les réseaux de neurones artificiels, l'évolution artificielle et la modélisation cellulaire. Ces deux derniers points sont traités au sein d'une structure trans-disciplinaire que j'ai contribué à créer en 2001 à Lyon sur la campus de la Doua : le groupe de Biologie des Systèmes et Modélisation Cellulaire (BSMC - <http://bsmc.insa-lyon.fr>).

Réseaux de neurones récurrents L'objectif de mes travaux sur les réseaux de neurones récurrents est de développer des modèles de réseaux de neurones biologiquement plausibles pour le contrôle sensori-moteur d'aimats ("animaux-robots"). Dans ce cadre, nous utilisons des réseaux de neurones à impulsion ("Spiking Neurons") et une loi d'apprentissage inspirée de la loi STDP "Spike Timing Dependent Plasticity" afin de faire émerger un couplage entre les dynamiques environnementales et les dynamiques neuronales. Ce projet a été financé par une Action Concertée Incitative de 2002 à 2005 (projet "DYN", coordonné par le Professeur Manuel Samuelides) et je participe depuis 2007 à un projet de l'Agence Nationale de la Recherche sur cette thématique (projet "MAPS", coordonné par F. Alexandre). Dans le cadre de cette activité j'ai encadré plusieurs stages de master en 2002, 2003, 2005 and 2006. Une thèse a été soutenue sous ma supervision en juillet 2005¹.

¹Depuis mon transfert vers l'équipe Turing du LIRIS, cette thématique est peu à peu abandonnée au profit de la modélisation cellulaire.

Évolution artificielle Dans ce domaine, j'ai proposé un nouveau modèle d'évolution artificielle respectant les principaux traits structuraux des génomes bactériens (modèle *aevol*). Ce modèle a été utilisé pour étudier les mécanismes de sélection indirecte des structures génétiques. Parallèlement, ce modèle a été utilisé pour développer de nouveaux algorithmes évolutionnaires à même de se structurer pour améliorer leur évolutivité (algorithme RBF-Gene). Dans le cadre de cette activité j'ai encadré plusieurs stages de master en 2003 et 2008. Deux thèses ont été soutenues sous ma supervision en décembre 2006 et juillet 2007.

Modélisation cellulaire L'objectif est ici de proposer des modèles individu-centré du noyau des cellules eucaryotes afin d'étudier la dynamique spatio-temporelle des structures nucléaires (nucléole, corps nucléaires, corps de Cajal, ...). En outre, un des objectifs de ce projet est d'approcher les causes moléculaire de la stochasticité de l'activité transcriptionnelle des cellules eucaryotes. Ce projet a été soutenu par une Action Concertée Incitative de 2004 à 2007 (projet "MoCeME" que je coordonne) et par la région Rhône-Alpes (dans le cadre du cluster ISLE). Enfin, dans le cadre de ce projet, j'ai participé à la création d'un groupe de travail national (SMABio).

5 Activités d'Encadrement Scientifique (depuis 1998)

Ne sont mentionnés que les projets effectués sous ma responsabilité au sein du laboratoire PRISMa (de 1998 à 2006) puis au sein du laboratoire LIRIS. Les projets en entreprise encadrés au seul titre de l'enseignement ne sont pas référencés ici.

5.1 Thèses

Depuis 2006 Co-encadrement (avec M. Olivier Gandrillon, CR-1 CNRS) de la thèse de M. Antoine Coulon. Sujet de la thèse : *Modélisation de structures nucléaires - vers une cellule virtuelle* (voir section II.4). Soutenance prévue pour septembre 2009.

Depuis 2005 Co-encadrement (avec M. le professeur Joël Favrel) de la thèse de Melle. Yolanda Sánchez-Dehesa Fernandez. Sujet de la thèse : *Modélisation de l'évolution des réseaux de régulation génétiques* (voir section II.3). Soutenance prévue pour décembre 2008.

2003-2007 Co-encadrement (avec M. le professeur Joël Favrel) de la thèse de Mme. Virginie Lefort. Thèse soutenue le 5 juillet 2007 devant Mme. Évelyne Lutton (rapporteur) et MM. G. Beslon (co-directeur de thèse), S. Christensen (examineur), Ph. Collard (rapporteur), J. Favrel (co-directeur de thèse) et J.-M. Peña (examineur, président du jury). Titre du mémoire de thèse : *Évolution de second ordre et algorithmes évolutionnaires : l'algorithme RBF-Gene*¹ (voir section II.3).

2003-2006 Co-encadrement (avec M. le professeur Jean-Michel Fayard) de la thèse de Melle. Carole Knibbe. Thèse soutenue le 4 décembre 2006 devant MM. G. Beslon (co-directeur de thèse), L. Duret (examineur), J.-M. Fayard (co-directeur de thèse), M. Morvan (rapporteur), E. Rocha (rapporteur), M. Schoenauer (examineur, président du jury) et F. Taddei (membre invité). Titre du mémoire de thèse :

¹Melle Virginie Lefort est actuellement ATER à l'Université Louis Lumière - Lyon 2.

*Structuration des génomes bactériens par sélection indirecte de la variabilité mutationnelle, une approche de modélisation et de simulation*¹ (voir section II.3).

2002-2005 Co-encadrement (avec M. le professeur Joël Favrel) de la thèse de M. Hédi Soula. Thèse soutenue le 11 juillet 2005 (mention très honorable) devant MM. G. Beslon (co-directeur de thèse), J. Favrel (co-directeur de thèse), D. Floréano (rapporteur), P. Meyrand (examinateur, président du jury), A. Mille (examinateur) et M. Samuelides (rapporteur). Titre du mémoire de thèse : *Dynamique et plasticité dans les réseaux de neurones à impulsion, étude du couplage temporel réseau/agent/environnement*² (voir section II.2).

1999-2001 Co-encadrement (avec M. le professeur Joël Favrel) de la thèse de Melle Carine Bournez. Thèse soutenue le 5 octobre 2001 (mention très honorable) devant MM. Ph. Beaune (examinateur), G. Beslon (co-directeur de thèse), J. Erschler (rapporteur), J. Favrel (co-directeur de thèse), K. Guedira (rapporteur), et J.P. Muller (examinateur). Titre du mémoire de thèse : *Une architecture multi-agents réflexive pour le contrôle de systèmes de production distribués et hétérogènes*³.

5.2 Stages de DEA/Master Recherche

2008 – Encadrement du stage de Master d’Informatique de Lyon (MASTRIA) de M. David Parsons. Sujet du projet : *Développement d’un modèle de recombinaison homologue pour le simulateur aeol*.

– Co-encadrement (avec Mme. Françoise Budar) du stage de Master de Bioinformatique (B2G) de M. Gaël Kaneko. Sujet du projet : *Modélisation de l’évolution de génomes mitochondriaux*.

2006 – Encadrement du stage de Master d’Informatique de Lyon (MASTRIA) de M. Antoine Coulon. Sujet du projet : *Étude de la réponse d’un réseau de neurones à impulsion à un signal complexe ; influence de la STDP sur la dynamique de la réponse*⁴.

– Encadrement du stage de Master d’Informatique de Lyon (MASTRIA) de M. Damien Peelman. Sujet du projet : *Codage d’un réseau de neurones récurrent par algorithmes évolutionnaires*.

2005 – Encadrement du stage de Master d’Informatique de Lyon (MASTRIA) de M. Frédéric Henry. Sujet du projet : *Influence de la règle STPD sur le comportement dynamique d’un réseau récurrent. Application à l’apprentissage par renforcement*⁵.

¹Après un stage post-doctoral à l’hôpital Necker (Paris), Melle Carole Knibbe a été recrutée comme Maître de Conférences à l’Université Claude Bernard - Lyon 1 où elle exerce actuellement.

²Après un stage post-doctoral à la NIH (National Institute of Health, Bethesda), M. Hédi Soula a été recruté comme Maître de Conférences au département BioSciences de l’INSA de Lyon où il exerce actuellement

³Melle Carine Bournez est actuellement ingénieur de recherche à l’INRIA Sophia-Antipolis dans le cadres de l’European Research Consortium for Informatics and Mathematics (W3C-ERCIM).

⁴Master soutenu en juillet 2006 (mention Bien, major de promotion). M. Antoine Coulon est actuellement en thèse sous ma direction (en co-encadrement avec M. Olivier Gandrillon).

⁵M. Frédéric Henry est actuellement inscrit en thèse au laboratoire “Perception et Mouvement” de Marseille.

- Encadrement du stage de Master de l’Université de Jendouba (Tunisie) de M. Jaber Kouki. Sujet du projet : *Optimisation des interactions protéines-protéines dans le logiciel 3DSPI*.
- Encadrement du stage de Master de l’Université de Jendouba (Tunisie) de M. Anis Tissaoui. Sujet du projet : *Modélisation 3D de dynamique protéiques avec le langage SMBL*.
- 2003** – Encadrement du stage de DEA DISIC de Mme. Virginie Lefort. Sujet du projet : *Utilisation du modèle RBF-Gene en intelligence artificielle*¹.
- Co-encadrement (avec M. Jacques Skubich) du stage de DEA DISIC de M. Thomas Tempé. Sujet du projet : *Pilotage d’un Khepera par réseaux de neurones à spike*.
- Co-encadrement (avec M. le professeur Jean-Michel Fayard) du stage de DEA AMSB de Melle. Carole Knibbe. Sujet du projet : *Étude de l’évolution structurelle des génomes avec le modèle RBF-Gene*².
- 2002** – Encadrement du stage de DEA ECD de M. Patrice Vienne. Sujet du projet : *Apprentissage d’un comportement de prédation avec une architecture de critique adaptative*³.
- Co-encadrement (avec M. le professeur Joël Favrel) du stage de DEA ISCE de M. Hédi Soula. Sujet du projet : *Apprentissage de comportements par réseaux de neurones temporels récurrents*⁴.

6 Autres activités scientifiques

6.1 Organisation de conférences scientifiques

Depuis 2006 Membre du comité d’organisation des “Séminaires de Modélisation du Vivant” (SeMoVi), Lyon, France.

Depuis 2004 Membre du comité scientifique et du comité d’organisation de la conférence annuelle *Integrative Post Genomics*, Lyon (France), octobre 2004, novembre 2005, décembre 2006, novembre 2007, novembre 2008.

BGRS 2008 Membre du comité scientifique du workshop “Knowledge Discovery in Bioinformatics” dans le cadre de la sixième International Conference on Bioinformatics of Genome Regulation and Structure, Novosibirsk (Russie), juin 2008

ECML-PKDD 2007 Membre du comité d’organisation du workshop “Data Mining in Functional Genomics and Proteomics” dans le cadre de la conférence ECML-PKDD’07, Varsovie (Pologne), septembre 2007.

¹DEA soutenu en juillet 2003 (mention Très Bien). Suite à son DEA Mme. Virginie Lefort a effectué une thèse sous ma direction (en co-direction avec M. le professeur Joël Favrel).

²DEA soutenu en septembre 2003 (mention Très Bien, major de promotion). Suite à son DEA Melle. Carole Knibbe a effectué une thèse sous ma direction (en co-direction avec M. le professeur Jean-Michel Fayard).

³DEA soutenu en juillet 2002 (mention Bien). Suite à son DEA, M. Patrice Vienne a effectué une thèse au laboratoire PRISMa sous la direction de M. le professeur Jean-Louis Sourrouille.

⁴DEA soutenu en juillet 2002 (mention Bien, major de promotion). Suite à son DEA M. Hédi Soula a effectué une thèse sous ma direction (en co-direction avec M. le professeur Joël Favrel).

- ECCS 2007** Organisateur de la session “Biological Modelling” pour le workshop “Multi-Agent Modelling for Complex Systems” (MA4CS) dans le cadre de la troisième European Conference on Complex Systems, Desden (Germany), octobre 2007
- InSciT 2006** Membre du comité scientifique du workshop “Knowledge Discovery and Management in Life Sciences” dans le cadre de l’International Conference on Multidisciplinary Information Sciences and Technologies, Merida (Espagne), octobre 2006
- JOBIM 2005** Membre du comité d’organisation des *Journées Ouvertes en Biologie Informatique et Mathématiques*, Lyon (France), juillet 2005
- JPGD 2003** Membre du comité d’organisation des *Journées de Post-Génomiques de la DOUA*, Lyon (France), mai 2003
- ECAI 2002** Membre du comité d’organisation de la *European Conference on Artificial Intelligence*, Lyon (France), juin 2002.
- EWLR 2001** Membre du comité scientifique du *European Workshop on Learning Robots*, Prague, septembre 2001.
- 1998** Organisation du workshop “Reproduction sexuée et évolution” à l’Institut d’Informatique de l’Université de Fribourg (Switzerland), juin 1998.

6.2 Conférences invitées

- 2008** – Workshop “Evolutionary Dynamics” (Utrecht, Pays-Bas). Titre de la conférence : *Evolution of genome structure - a computational approach* (mars 2008)
- Séminaire du laboratoire Joliot Curie (Lyon, France). Titre de la conférence : *Evolution of genome structure - a computational approach* (janvier 2008)
- Séminaire du laboratoire BF2I (Lyon, France). Titre de la conférence : *A long-term evolutionary pressure on the amount of noncoding DNA* (janvier 2008)
- 2007** – Cavanilles Seminar on Complexity and Evolution (Valence, Espagne). Titre de la conférence : *Evolution of genome structure - a computational approach* (décembre 2007)
- Séminaire de l’équipe Éco-Évolution Mathématique (EEM) de l’École Normale Supérieure (Paris, France). Titre de la conférence : *A long-term evolutionary pressure on the amount of noncoding DNA* (décembre 2007)
- Journées communes des clusters environnement et informatique (INRIA Montbonnot, France). Titre de la conférence : *Digital Genetics : modéliser la dynamique évolutive des génomes* (novembre 2007)
- Santa-Fe Institute (Santa-Fe, Nouveau Mexique). Titre de la conférence : *Systems Biology, where is the network ?* (juillet 2007)
- École de printemps “Hasard et Fluctuations en Biologie” (Berder, France). Titre de la conférence : *Évolution de second ordre : le hasard sélectionné* (mai 2007)
- Journées du Réseau National des Systèmes Complexe (RNSC, Paris, France). Titre de la conférence : *De l’évolution à l’évolvabilité : exploration par la simulation* (mars 2007)
- 2006** – Journées de Gerland “Biologie et Complexité” (Lyon, France). Titre de la conférence : *Évolution et robustesse : un équilibre complexe aux conséquences inattendues* (septembre 2006)

- École de printemps “Modélisation des Systèmes Biologiques” (Bombannes, France).
Titre de la conférence : *Multi-agent modeling of nuclear structures* (mars 2006)
- 2005** Séminaire d’Épigénomique (Évry, France). Titre de la conférence : *Introducing Proteins into Genetic Algorithms* (juillet 2005)
- 2004** Séminaire du TIMC (Grenoble, France). Titre de la conférence : *An artificial life model for the structural evolution of bacterial genomes* (décembre 2004)
- 2001** Séminaire du LIP6 (Paris, France). Titre de la conférence : *Les assemblées de neurones, ou la modularité faible du comportement* (juin 2001)
- 1999** Séminaire de l’INRETS (Lyon, France). Titre de la conférence : *Une architecture neuro-réactive en robotique mobile* (octobre 1999)

6.3 Évaluation scientifique - Jury de thèses

- 2008** Membre du jury de thèse de M. Milan van Hoek (Theoretical Biology and Bioinformatic Group, Univ. Utrecht, Pays-Bas).
- 2007** Membre du jury de thèse de M. Hunor Albert-Lorincz (LIRIS, INSA-Lyon).
- 2005** Membre du jury de thèse de M. Fabien Flacher (LIP6, Univ. Paris 6).
- MBE** Reviewer pour la revue internationale “Molecular Biology and Evolution” (Oxford University Press)
- JTB** Reviewer pour la revue internationale “Journal of Theoretical Biology” (Elsevier)
- TAAS** Reviewer pour la revue internationale “Transactions on Autonomous and Adaptive Systems” (ACM Press)

6.4 Vulgarisation scientifique

- 2008** Conseiller scientifique pour le spectacle “The looking glass”, produit par la compagnie Zabzka autour de la vie de de l’œuvre de Lewis Carroll. Mise en scène : Benoît Bradel.
- 2008** Conférence-débat sur la robotique intelligente au Centre Culturel Charlie Chaplin (Vaulx en Velin) dans le cadre des trophées de robotique.
- 2007** Conseiller scientifique pour le spectacle “Performance Emergence” produit par la section Théâtre-Études de l’INSA de Lyon avec le soutien de l’Institut Rhône-Alpin des Systèmes Complexes (IXXI) et mis en scène par Benoît Bradel.
- 2007** Collaboration à un article de vulgarisation dans la revue “Science & Vie” (août 2007) sur le thème du hasard (titre de l’article : “Et si la vie devait tout au hasard?”).
- depuis 2006** Correspondant référent pour la collaboration CNRS/Club Elek (Club de robotique de l’INSA de Lyon)
- 2003** Stand, démonstrations et conférences sur la robotique intelligente à la cité des sciences de Paris la Villette dans le cadre de la finale des trophées de robotique.
- 2001** Article “Des réseaux de neurones auto-organisés” paru dans la revue “Pour la Science” (août 2001).

7 Activités d'Enseignement (pour l'année 2007/2008)

7.1 Au département Informatique de l'INSA de Lyon

Troisième année – Travaux dirigés et travaux pratiques dans les domaines “architectures des ordinateurs” (projet “micromachine”) et “bases techniques pour les réseaux” (projet “khepif”),

- Encadrement des projets “fil rouge” (projets libres de développement informatique et internet).

Cinquième année – Séminaire d'informatique bioinspirée et sensibilisation à la complexité,

- Encadrement de projets de fin d'étude.

7.2 Au département BioSciences de l'INSA de Lyon

Quatrième année Cours d'informatique bioinspirée pour la filière Bioinformatique et Modélisation (BiM).

Cinquième année – Cours et travaux pratiques de bioinformatique pour la filière Biochimie (BC).

- Encadrement de projets de fin d'études pour la filière BiM.

7.3 Enseignement de troisième cycle

Mastère MASTRIA (Lyon) Cours de recherche en Informatique Bio-Inspirée (réseaux de neurones, algorithmes génétiques, systèmes multi-agents ; cours réalisé en collaboration avec Melles. Salima Hassas et Carole Knibbe).

Mastère de Sciences Cognitive (Lyon) Séminaire sur la vie artificielle.

Mastère d'Informatique pour la Médecine (Lyon) Séminaire sur la modélisation informatique du vivant.

8 Activités Administratives, Responsabilités Collectives

IXXI Directeur-Adjoint de l'Institut Rhône-Alpin des Systèmes Complexes (IXXI) depuis novembre 2007.

Région Rhône-Alpes Co-responsable¹ (avec Alain Viari) du projet “Modèles formels et complexité en biologie” du cluster de recherche ISLE (Informatique, Signal, Logiciel Embarqué).

INSA-Lyon – Membre de la commission de spécialistes d'établissement (section 61).

- Membre de la commission de recrutement des enseignants du second degré (spécialité : éducation physique et sportive).
- Membre extérieur du conseil de département au département des humanités.
- Responsable de l'option “Lumière et Son” de la section artistique “Théâtre-Études”.

¹depuis février 2008

Annexe B

Bibliographie personnelle classée

1 Revues internationales

C. Knibbe, J.-M. Fayard, **G. Beslon**, (2008) *The topology of the protein network influences the dynamics of gene order : from systems biology to a systemic understanding of evolution*, Artificial Life Journal, 2008, Vol. 14, No. 1, p. 149-156

C. Knibbe, O. Mazet, F. Chaudier, J.-M. Fayard, **G. Beslon**, (2007) *Evolutionary coupling between the deleteriousness of gene mutations and the amount of non-coding sequences*, Journal of Theoretical Biology, 2006, Vol. 44, No. 4, p. 621-630

C. Knibbe, A. Coulon, O. Mazet, J.-M. Fayard, **G. Beslon**, (2007) *A long-term evolutionary pressure on the amount of noncoding DNA*, Molecular Biology and Evolution, 2007, Vol. 24, No. 10, p. 2344-2353.

W. Banzhaf, **G. Beslon**, S. Christensen, J. A. Foster, F. Képès, V. Lefort, J. F. Miller, M. Radman, J. J. Ramsden, (2006) *From artificial evolution to computational evolution : a research agenda*, Nature Genetic Review, Vol. 7, p. 729-735.

H. Soula, **G. Beslon**, O. Mazet, (2006) *Spontaneous dynamics of asymmetric random recurrent spiking neural networks*, Neural Computation, Vol. 18, No. 1, p. 60-79.

H. Soula , C. Robardet , F. Perrin , S. Gripon , **G. Beslon**, O. Gandrillon, (2005) *Modeling the emergence of multi-protein dynamic structures by principles of self-organization through the use of 3DSpi, a Multi-Agent-based Software*, BMC Bioinformatics 6 :228, 11 p.

N. Reymond, H. Charles, L. Duret, F. Calevro, **G. Beslon**, J.-M. Fayard, (2004). *ROSO : Optimizing oligonucleotide probes for microarrays*, bioinformatics, Vol. 20, No. 2, p. 271-273.

2 Revues nationales, chapitres d'ouvrages

A. Coulon, H. Soula, O. Mazet, O. Gandrillon, **G. Beslon** (2008) *Large multi-protein structures modeling and simulation : the need for mesoscopic models*, Methods in Molecular Biology Series, vol. 484 pp. 537-558.

A. Coulon, H. Soula, O. Mazet, O. Gandrillon, **G. Beslon** (2007) *Modélisation cellulaire pour l'émergence de structures protéiques auto-organisées*, Technique et Science Informatiques (RSTI/TSI), 2007, Vol. 26, No. 1-2, p. 123-148.

G. Beslon, H. Soula, (2003). *Laissez-les vivre; la codépendance temporelle comme clé de l'apprentissage de comportements chez un animat*, Technique et Science Informatiques (RSTI/TSI), 2003, Vol. 22, No. 2, pp. 145-170.

G. Beslon, E. Daucé, (2002). *Modularité et apprentissage dans les réseaux de neurones récurrents*, In : Guillot A., Daucé E. (Coord.) *Approche Dynamique de la Cognition Artificielle*, Paris (France) : Eds. Hermès-Lavoisier, pp. 61-80.

G. Beslon, (2002). *Coopération implicite et émergence dans des groupes d'agents*, In : Guillot A., Daucé E. (Coord.) *Approche Dynamique de la Cognition Artificielle*, Paris (France) : Eds. Hermès-Lavoisier, pp. 267-283.

3 Conférences internationales avec comité de lecture

Y. Sanchez-Dehesa, L. Cerf, J.-M. Peña, J.-F. Boulicaut, **G. Beslon** (2007) *Artificial regulatory networks evolution*, Proceedings of the first International Workshop on Machine Learning for Systems Biology (MLSB'07), Evry (France), p. 47-52.

Y. Sanchez-Dehesa, J.-M. Peña, **G. Beslon** (2007) *Generating data form the evolution of artificial regulatory networks*, ECML/PKDD Workshop on Data Mining in Functional Genomics and Proteomics, Varsovie (Pologne), octobre 2007, p. 91-103.

Y. Sanchez-Dehesa, **G. Beslon**, J.-M. Peña (2007) *Modeling evolution of regulatory networks in artificial organisms*, Proceedings of the 3rd International Symposium on Computational Life Science (CompLife'07), Utrecht (Pays-Bas), octobre 2007, p. 87-98.

H. Soula, **G. Beslon** (2006). *Spike-timing dependent plasticity learning for visual-based obstacle avoidance*, International Conference on Simulation of Adaptive Behavior, SAB 2006, LNAI 4095, Springer, 11 p.

E. Daucé, H. Soula, **G. Beslon** (2005) *Learning methods for dynamic neural networks*, International Symposium on Nonlinear Theory and its Applications (NOLTA 2005), 4 p.

C. Knibbe, **G. Beslon**, V. Lefort, F. Chaudier, J.-M. Fayard (2005). *Self-adaptation of genome size in artificial organisms*, In Capcarrere et al., *Advances in Artificial Life*, Proceedings of the 8th European Conference, ECAL 2005, LNAI 3630, Springer, pp. 423-432

C. Knibbe, **G. Beslon**, F. Chaudier, J.-M. Fayard, (2005). *Designing artificial organisms to study the influence of gene pleiotropy on genome evolution*, CD-Rom Proc. of the Systems Biology Workshop at ECAL 2005, 10 p.

V. Lefort, **G. Beslon**, J. Favrel, (2005). *Simultaneous optimization of weights and structure of an RBF Neural Network*, Talbi et al., *Artificial Evolution*, Proceedings of International Conference on Evolutionary Algorithms (EA'05), LNCS 3871, Springer, 10 p.

H. Soula, **G. Beslon**, J. Favrel, (2005). *Evolution versus learning in temporal neural networks*, Procs. of the Int. Conf. on Adaptive and Natural Computing (ICANNGA'05), Cumbria (Portugal), 2005, p. 42-45.

H. Soula, A. Alwan, **G. Beslon**, (2005). *Learning at the Edge of Chaos : Temporal Coupling of Spiking Neuron Controller for Autonomous Robotics*, AAI Spring symposia on developmental robotics, Stanford (USA), 2005.

V. Lefort, C. Knibbe, **G. Beslon**, J. Favrel, (2004). *A Bio-Inspired Genetic Algorithm with a Self-Organized Genome : The RBF-Gene*, Procs. of the Genetic and Evolutionary Computation Conference (GECCO'04), Seattle (Washington), 2004, p. 406-407

V. Lefort, C. Knibbe, **G. Beslon**, J. Favrel, (2004). *Introducing "Proteins" into Genetic Algorithms*, Proc. of CSIMTA'04 Int. Conf., Cherbourg (France), 2004, p. 181-186.

G. Beslon, C. Knibbe, H. Soula, J-M. Fayard, (2003). *The RBF-Gene Model*, Proc. of the Int. Conf. on Artificial Neural Nets and Genetic Algorithms (ICANNGA'2003), Roannes (France), 2003, pp. 187-192.

H. Soula, **G. Beslon**, J. Favrel, (2003). *Evolving Spiking Neurons Nets to Control an Animat*, Proc. of the Int. Conf. on Artificial Neural Nets and Genetic Algorithms (ICANNGA'2003), Roannes (France), 2003, pp. 193-197.

G. Beslon, H. Soula, J. Favrel, (2001). *A Neural Model for Animats Brain*, Proc. of the Int. Conf. on Artificial Neural Nets and Genetic Algorithms (ICANNGA'2001), Prague (Czech Republic), 2001, pp. 352-355.

C. Bournez, **G. Beslon**, J. Favrel, (2001). *OCEAN : A Multi-Agent Market for Distributed System Control*, Proc. of the European Simulation Symposium (ESS'2001), Marseille (France), 2001.

H. Soula, **G. Beslon**, J. Favrel, (2001). *Controlling an Animat with a Self-Organized Modular Neural Model*, Proc. of the European Workshop on Learning Robots (EWLR'2001), Prague (Czech Republic), 2001, pp. 39-46.

G. Beslon, F. Biennier, J. Favrel, (2000). *Towards NeuroReactivity and Neural Assemblies Emergence : The Asynchronous Spike Propagation Model*, Proc. of Neural Computation (NC'2000), Berlin (Germany), 2000, pp. 456-462.

G. Beslon, C. Bournez, J. Favrel, (2000). *Adaptive Reactive Agents for Industrial System Control : The ARAPSI Project*, Proc. of the Int. Conf. on Management and Control of Production and Logistics (MCPL'2000), Grenoble (France), 2000, 7 p.

C. Bournez, **G. Beslon**, J. Favrel, (2000). *A Multi-Agent System for Distributed Factory Decentralized Emergent Control*, Proc. of the ISPE Int. Conf. on Concurrent Engineering (CE'2000), Lyon (France), 2000, pp. 445-452.

G. Beslon, J. Favrel, (1999). *Genetically Determined Conveying Agents In a Flexible Material Handling System*, Proc. of the Int. Conf. on Systemics, Cybernetics and Informatics (SCI'99), Orlando (USA), 1999, Vol. 3, p. 288-295.

C. Bournez, **G. Beslon**, J. Favrel, (1999). *A Multi-Agent Market-Driven Contract Net for Emergent Manufacturing Control*, Proc. of the Int. Conf. on Systemics, Cybernetics and Informatics (SCI'99), Orlando (USA), 1999, Vol. 3, p. 296-303.

G. Beslon, F. Biennier, B. Hirsbrunner, (1998). *Multi-Robot Path-Planning Based on Implicit Cooperation in a Robotic Swarm*, Procs. of Autonomous Agents'98, Minneapolis (USA), 1998, pp. 39-46.

G. Beslon, F. Biennier, B. Hirsbrunner, J. Favrel, (1998). *Reactive Path-Planning Through Implicit Cooperation in a Distributed Material Handling System*, Proc. of Mechatronics'98, Kitakyushu (Japan), 1998, p. 406-411.

G. Beslon, P. Ambroise, B. Heit, J. Bremont, F. Guillemin, (1996). *Intravesical Dosimetry Applied to Photodynamic Therapy*, Proc. of the International Society for Optical Engineering (SPIE'96), Viennes (Austria), 1996, p. 116-124.

E. Mauget, **G. Beslon**, M. Henryon, V. Bombardier, J. Bremont, (1996). *Artificial Vision Applied to Production Tracking : A Neural Networks Approach*, Procs. of Mechatronics'96, Besançon (France), 1996, p. 267-271.

G. Beslon, F. Biennier, J. Favrel, (1995). *A Neural Process Based on Sensori-Motor Loops*, Procs. of Neural, Parallel and Scientific Computation (NPSC95), Atlanta (USA), 1995, Vol. 1, pp. 37-42.

G. Beslon, F. Biennier, J. Favrel, (1995). *Flexible Conveying System Based on Autonomous Vehicles*, Procs. of the Int. Conf. on CAD/CAM, Robotics and Factories of the Future (CARS&FOF 95), Pereira (Colombia), 1995, Vol. 1, pp. 115-120.

F. Biennier, **G. Beslon**, J. Favrel, (1995). *Robust Control for Manufacturing Systems*, Procs. of IFAC/IFIP/IFORS Workshop on Intelligent Manufacturing Systems, Bucarest (Hungaria), 1995, pp. 349-354.

G. Beslon, F. Biennier, J. Favrel, (1994). *Structured Knowledge Acquisition Thanks to Neural Networks*, Procs. of Data and Knowledge Systems for Manufacturing and Engineering (DKSME94), Hong-Kong (Hong-Kong), 1994, Vol. 2, pp. 515-520.

G. Beslon, F. Biennier, J. Favrel, (1993). *Integrated Knowledge System Based on Neural Controllers*, Procs. of the Int. Conf. on CAD/CAM, Robotics and Factories of the Future (CARS&FOF 93), Newark (USA), 1993, pp. 451-456.

F. Biennier, J. Favrel, **G. Beslon**, (1993). *Integration of Information and Knowledge from the Engineering Activity to the Workshop Control*, Procs. of IEEE Int. Conf. on Robotics and Automation (RA'93), Atlanta (USA), 1993, Vol. 1, pp. 840-845.

F. Biennier, **G. Beslon**, J. Favrel, (1992). *Utilisation de Réseaux Neuronaux pour le Pilotage d'Ateliers Flexibles*, Procs. of Neuro-Nîmes'92, Nîmes (France), 1992, p. 465-476.

4 Autres conférences, résumés, posters

V. Lefort, **G. Beslon** (2008). *Optimisation de réseaux de neurones RBF pour la régression via un algorithme évolutionnaire : RBF-Gene*, Proceedings of EGC'08, Sophia-Antipolis (France), janvier 2008, p. 331-336.

A. Coulon, **G. Beslon**, O. Gandrillon (2007). *Binding cooperation and competition among transcription factors can cause complex stochastic properties in gene expression*, Integrative Post-Genomics (IPG'07), Lyon (France), novembre 2007, p. 1

Y. Sanchez-Dehesa, J.-M. Peña, **G. Beslon** (2007). *RAevol, un modèle degénétique digitale des réseaux de régulation*, Proceedings of the workshop Réseaux d'Interactions : Analyse, Modélisation, Simulation (RIAMS'07), Lyon (France), novembre 2007.

C. Knibbe, **G. Beslon**, J.-M. Fayard (2006). *Long-term needs for evolvability and robustness shape genome structure : A simulation study*, Journées Ouvertes en Biologie, Informatique et Mathématiques (Jobim'2006), Bordeaux (France), 2006.

C. Knibbe, A. Coulon, O. Mazet, J.-M. Fayard, **G. Beslon**, J.-M. Fayard (2006). *Amount of Non-Coding Sequences Depends on Mutation Rate*, 10th Evolutionary Biology Meeting at Marseilles, Marseille (France), 2006.

C. Knibbe, **G. Beslon**, J.-M. Fayard, (2005). *Evolutionary influence of the protein network topology on gene organisation in artificial organisms*, Proceedings of European Conference on Complex Systems (ECCS'2005), Paris (France), 2 p.

C. Robardet, H. Soula, K. Frugier, S. Belin, J.-J. Diaz, O. Gandrillon, **G. Beslon**, (2005). *Multi-agent modeling of dynamic multi-protein cellular structures*, Journées Ouvertes en Biologie, Informatique et Mathématiques (Jobim'2005), Lyon (France), 2005.

C. Knibbe, V. Lefort, **G. Beslon**, J.-M. Fayard, (2004). *Evolutionary Dynamics of Overlapping Genes in Virtual Organisms*, Integrated Post-Genomics Conference (IPG'04), Lyon (France), p. 40.

H. Soula, C. Robardet, K. Frugier, F. Perrin, S. Gripon, **G. Beslon**, O. Gandrillon, (2004). *Evidence that Multi-Protein Dynamic Structures are Formed and Maintained by Principles of Self-Organization through the use of 3DSpi, a Multi-Agent-Based Software*, Integrated Post-Genomics Conference (IPG'04), Lyon (France), p. 55.

H. Soula, A. Alwan, **G. Beslon**, (2004). *Obstacle Avoidance Learning in a Network of Spiking Neurons*, Last Minute Results of Simulation of Adaptive Behavior (SAB'04), Santa-Monica (USA), 2004.

C. Knibbe, V. Mathivet, **G. Beslon**, H. Soula, J.-M. Fayard, (2003). *The RBF-Gene Model : A Tool to Study Interactions Between Genomic Structures and Evolution*, Journées de Post-Génomiques de la Doua (JPGD'03), Lyon (France), 2003.

C. Knibbe, **G. Beslon**, J.-M. Fayard, (2003). *Modeling Evolutionary Interactions Between Genomic and Functional Structures*, Proc. of the Int. Conf. on System Biology (ICSB'2003), St. Louis (Washington DC), 2003, p. 163-164

- N. Reymond, H. Charles, **G. Beslon**, J.-M. Fayard, (2002). *ROSO : a software to search optimized oligonucleotide probes for microarrays*, Intelligent Systems for Molecular Biology (ISMB'02), Edmonton (Canada)
- C. Becquet, C. Charavay, A. Herr, T. Martin, F. Morle, A. Deandrea, **G. Beslon**, O. Gandrillon, (2002). *Modeling Protein-Protein Interactions by Using Multi-Agents Systems*, Journées de Post-Génomiques de la Doua (JPGD'02), Lyon (France), 2002.
- G. Beslon**, H. Charles, J.-M. Fayard, O. Gandrillon, O. Mazet, H. Soula, (2002). *Functional Space Approach for Evolutionary Algorithms*, Journées de Post-Génomiques de la Doua (JPGD'02), Lyon (France), 2002.
- G. Beslon**, H. Charles, J.-M. Fayard, O. Gandrillon, O. Mazet, (2002). *Modeling Evolution Through a Genetic Regulation Network*, Journées de Post-Génomiques de la Doua (JPGD'02), Lyon (France), 2002.
- N. Reymond, H. Charles, **G. Beslon**, J.-M. Fayard, (2002). *ROSO : A Software to Search Optimized Oligonucleotides Probes for Microarrays*, Journées de Post-Génomiques de la Doua (JPGD'02), Lyon (France), 2002.
- N. Reymond, H. Charles, **G. Beslon**, J.-M. Fayard, (2002). *ROSO : a software to search optimized oligonucleotide probes for microarrays*, Journées Ouvertes en Biologie Informatiques et Mathématiques (Jobim'02), Saint-Malo (France).
- N. Reymond, H. Charles, **G. Beslon**, J.-M. Fayard, (2001). *Développement d'un Logiciel d'Optimisation de Sondes Oligonucléotidiques Destinées aux Puces à ADN*, Journées de Post-Génomiques de la Doua (JPGD'01), Lyon (France), 2001.
- G. Beslon**, J. Favrel, (2000). *Optimisation de Chemins par Algorithmie Génétique et Réseaux Neuromimétiques dans un Système de Transport Distribué*, Proc. of ROADEF'2000, Nantes (France), 2000, pp. 24-25.
- G. Beslon**, F. Biennier, (1994). *Architectures Connexionnistes pour le Behavior-Based Control*, Actes de Neurosciences et Sciences de l'Ingénieur (NSI94), Chamonix (France), 1994, pp. 149-152.