

INSA-Lyon – Université Claude Bernard Lyon I

**Action Concertée Incitative
– ImpBio –**

Modélisation Cellulaire Multi-Échelles
Application à la modélisation multi-agents du
comportement de la nucléoline au cours d'une infection
par le virus de l'Herpès

Partie II.B : DESCRIPTION DU PROJET

Coordination : G. Beslon (INSA-PRISMa)
INSA-IF/PRISMa, Bât. Blaise Pascal
gbeslon@prisma.insa-lyon.fr

Laboratoires partenaires :
PRISMa (INSA) – Maply (INSA) – LIRIS (UCBL) – CGMC (UCBL)

Version 2.6 – 9 mars 2004

B1 Objectif et contexte

B1.1 Introduction

Depuis quelques années, les projets de simulation informatique à l'échelle de la cellule se sont multipliés¹. Le but de ces modèles est d'intégrer "l'ensemble" des connaissances sur un modèle biologique afin de rendre compte de son fonctionnement global. C'est pourquoi ils sont basés sur une description *compartimentale* de l'objet cellulaire : le métabolisme est généralement décrit en termes d'équations différentielles régissant les échanges entre des compartiments cellulaires prédéfinis et considérés comme stables.

Une telle approche présente des intérêts indéniables : elle permet de choisir simplement la complexité du modèle (il "suffit" de choisir la granularité de la description compartimentale) et, surtout, elle est très résistante au passage à l'échelle. En outre, elle a l'avantage d'être bien maîtrisée par les biologistes et d'être relativement simple à mettre en œuvre au niveau informatique.

Pour autant, la modélisation compartimentale – et les simulations correspondantes – néglige de fait des pans entiers de la biologie moléculaire et cellulaire. C'est, en effet, dans le passage du niveau moléculaire au niveau métabolique que se situent les questions les plus prégnantes (auto-organisation par exemple). Or, les modèles de cellule virtuelle sont contruits dans une perspective descriptive alors que, si ces questions doivent être abordées, elles ne peuvent l'être que dans une perspective intégrative, partant de la partie (ici la molécule) pour remonter au tout (la cellule) afin de mettre à jour les mécanismes émergents qui sous-tendent son fonctionnement. C'est dans cette perspective que nous nous situons, c'est-à-dire dans une démarche relevant explicitement de la Biologie des Systèmes (section B1.2).

À travers le présent projet, nous nous proposons de développer un modèle de cellule virtuelle (ou plus modestement, d'une partie de cellule) mais en utilisant un formalisme plus conforme aux objectifs de la Biologie des Systèmes que le formalisme compartimental. Pour cela, nous proposons d'utiliser les méthodes et outils de l'Intelligence Artificielle Distribuée dans une démarche de modélisation individu-centrée. Une telle approche a déjà été appliquée avec succès en biologie des populations (voir section B1.3) mais son utilisation en biologie moléculaire et cellulaire reste encore très confidentielle. La raison en est probablement qu'elle demande que soit mise en œuvre une réelle démarche pluridisciplinaire réunissant *au quotidien* biologistes, informaticiens et mathématiciens dans une démarche de modélisation commune, parallèle et non plus séquentielle.

B1.2 De la Biologie des Systèmes

Jusqu'à ces dernières années, les expériences biologiques ne permettaient de mesurer simultanément qu'un tout petit nombre de paramètres tels que les niveaux d'expression de gènes ou de protéines. L'interprétation des patrons d'expression obtenus était alors facilitée par le faible nombre de points de mesure. Ceci autorisait généralement l'utilisation de modèles cybernétiques simples, réductionnistes, dérivés des travaux princeps sur l'opéron lactose. La mise en œuvre des techniques d'analyse à haut débit a pulvérisé ce cadre conceptuel. En effet, ces mêmes schémas explicatifs se révèlent incapables de rendre compte des données accumulées. Dans ce cadre sont appelés à se développer de nouveaux outils théoriques qui doivent traiter l'ensemble des données obtenues à l'échelle de la cellule dans un cadre incluant les interactions des produits des gènes (ARN et protéines) et de l'environnement.

¹On se référera, par exemple, aux projets E-cell (<http://www.e-cell.org>), BioSpice (<http://biospice.lbl.gov>) ou CyberCell (<http://cybercell.biochem.ualberta.ca>).

Différentes approches regroupées sous le vocable de “Biologie des Systèmes” visent à comprendre le fonctionnement des objets de la biologie (cellules, organes, organismes, populations), en tant que systèmes² (Ideker et al., 2001; Kitano, 2002). Dans un premier temps, les différents auteurs se sont d’abord attachés à dégager des caractéristiques essentielles de ces systèmes. On peut d’ores et déjà établir quelques propriétés remarquables des systèmes biologiques :

Ce sont des systèmes complexes, au sens où leur fonctionnement global ne peut pas être prédit de manière simple par une décomposition raisonnée en sous-parties. Ils sont, de plus, fortement interconnectés, ce qui rend difficile la prévision de l’effet de la modification de l’état d’un seul agent sur le comportement global du système.

Ce sont, néanmoins, des systèmes modulaires (Vespignani, 2003). Différentes analyses, basées sur des mesures variées (ARN ou protéines), produisent des réseaux d’interactions dans lesquels on voit apparaître des groupes d’agents présentant un niveau d’interconnection significativement plus élevé que la moyenne du réseau. L’analyse des fonctions portées par ces agents révèle souvent des fonctions biologiques communes (Stuart et al., 2003).

Ce sont des systèmes basés sur des phénomènes stochastiques. Il est établi que la transcription est, à l’échelle cellulaire, plus correctement décrite par des distributions probabilistes que par des valeurs moyennes (Levsky and Singer, 2003) ce qui entraîne des réponses cellulaires hétérogènes pour un stimulus donné (Blake et al., 2003). Comprendre comment de ces phénomènes stochastiques surgit l’ordre cellulaire reste un enjeu majeur de la biologie contemporaine (Kupiec, 1997).

Ce sont des systèmes dégénérés. Au sein d’un système biologique, des structures différentes peuvent remplir une même fonction tandis que, inversement, une même structure peut remplir des fonctions différentes (Tononi et al., 1999; Atamas, 1996). Il est probable que la description fonctionnelle des protéines doivent souvent plus au hasard de la situation biologique dans laquelle les molécules ont pour la première fois été identifiées qu’au spectre complet de leurs potentialités fonctionnelles dans un environnement cellulaire réel.

En dépit de l’implication forte de quelques organismes de recherche³, ces approches de “Biologie des Systèmes” restent encore balbutiantes, en particulier du fait de leur aspect trop théorique. Il est clair que cette vision se doit encore de produire des résultats expérimentaux convaincants pour asseoir son audience.

B1.3 De la modélisation multi-agents de processus biologiques

La plupart des outils existants pour modéliser les phénomènes biologiques sont développés sous un pré-supposé continu. Ces approches modélisent des comportements statistiquement moyens de populations *supposées grandes* de molécules semblables. Ce type de pré-supposé se révèle inadéquat lorsque le nombre de molécules est trop faible ou lorsque la variété de ces molécules ou de leurs interactions est trop grande. Parfois aussi, les modèles continus sont valides mais la complexité des phénomènes émergents rend difficile leur description analytique (Ball et al., 2001).

Par ailleurs, les modèles continus ne sont pas toujours adaptés à la représentation de phénomènes spatio-dynamiques, par exemple lorsqu’ils supposent une décomposition *a priori* en compartiments entre lesquels les flux sont étudiés. En effet, cette vision compartimentée ne permet pas d’expliquer l’extrême mobilité intra- et inter-compartiments des molécules biologiques, alors même que les données qui montrent cette mobilité s’accumulent (Mistelli, 2001). Dans un tel contexte, il apparaît

²C’est-à-dire en tant qu’entité *globale*.

³On notera d’ailleurs la très faible contribution de l’Europe – et de la France – dans ce type de recherche.

pertinent de proposer des modèles complémentaires permettant de rendre compte des propriétés spatio-dynamiques observées.

Ce besoin d'un outil de modélisation alternatif se ressent particulièrement dans le cas de la modélisation des structures nucléaires dans les cellules eucaryotes. Ces structures sont composées de molécules présentes en faible quantité dans le noyau cellulaire et leur observation à l'aide de techniques de microscopie confocale a permis d'en montrer le caractère spatio-dynamique : les molécules qui les composent apparaissent comme étant particulièrement mobiles à l'intérieur du noyau alors que les structures qu'elles forment restent, elles, extrêmement stables (Mistelli, 2001).

Plusieurs auteurs ont récemment proposé d'utiliser le formalisme multi-agents (ou "individu-centré") pour la modélisation des processus cellulaires (Amar et al., 2002) et, pour notre part, nous avons initié des travaux en ce sens dès 2002 en proposant un outil de modélisation (3DSpi) (Becquet et al., 2002; Soula et al., 2004). Les systèmes multi-agents présentent en effet le double intérêt de permettre la modélisation de populations très hétérogènes en taille et d'intégrer le caractère spatio-temporel du processus. En effet, dans cette approche de modélisation individu-centrée, les agents sont considérés comme des entités autonomes, dotées de comportements propres, mais influencées par leur environnement immédiat (Weiss, 1999). Cette influence peut conduire à des interactions de nature très différente mais qui, dans tous les cas, émergent des comportements locaux sans être guidées par des présupposés structurels globaux (Beslon, 2002). La modélisation individu-centrée a été appliquée avec succès à des problématiques biologiques, en particulier dans le domaine de la dynamique des populations où elle a permis de représenter les comportements dynamiques des groupes sociaux et des essaims (Theraulaz and Spitz, 1997). Pourtant, la biologie cellulaire n'a pour l'instant pas – ou peu – su tirer partie de cette approche, probablement parce que les problèmes posés par le "passage à l'échelle" semblent quasiment insurmontables : si des modèles multi-agents simulant quelques milliers, voire quelques dizaines de milliers de molécules sont envisageables, nous sommes encore très loin de pouvoir modéliser et calculer les interactions entre les millions de macromolécules que compte une cellule. C'est pourquoi il semble impératif d'associer, à la modélisation multi-agents, des mécanismes de modélisation multi-échelles (Duboz, 2004) ce qui peut s'envisager sous trois modalités :

À l'intérieur du paradigme multi-agents : Le paradigme multi-agents n'est pas limité au seul niveau de description moléculaire. Il est donc parfaitement possible d'*agentifier* des niveaux de description différents. Cette possibilité permet d'envisager un couplage entre différents types de modèles à l'intérieur du paradigme multi-agents. En outre, l'approche dite des "méta-agents"⁴ permet de proposer des interactions à différents niveaux d'organisation dans un système multi-agents (Bournez, 2001).

Intégration de modèles géométriques : Certains compartiments cellulaires pourraient être décrits sur la base de modèles plus "simples" (au moins en termes de calculabilité). Il semble ainsi particulièrement intéressant de coupler des modèles multi-agents spatialisés avec des modèles géométriques (description d'une entité sous la forme d'un maillage représentant sa structure tridimensionnelle). Une telle description permet en effet de limiter les besoins de calcul (puisque plusieurs milliers de molécules sont modélisées sous la forme de quelques centaines de triangles) tout en permettant les interactions physiques avec les différents agents. Cependant, les interactions entre modèles multi-agents et modèles géométriques ne sont pas encore très bien maîtrisées et la question de la méthodologie pour y parvenir est encore ouverte. Nous nous proposons, dans un premier temps, de modéliser les différentes membranes (cellulaire et/ou nucléaire) sous une forme géométrique afin de permettre à la cellule et/ou au noyau de se déformer sous l'influence des entités internes (modélisées, elles, sous une

⁴Un méta-agent est un agent lui-même composé d'agents.

forme multi-agents).

Intégration de modèle mathématiques : Pour modéliser de grands écosystèmes, Duboz et al. (2003) proposent de coupler des modèles multi-agents avec des modèles mathématiques afin de relier les différentes échelles de temps et d'espace dans le modèle. Dans le cas de la modélisation cellulaire, une telle approche permettrait de remonter à l'échelle de la cellule sans pour autant simuler l'ensemble des entités en interaction. Cependant, en ayant à l'esprit les limites de ces modèles déterministes décrivant des comportements moyens, nous proposons d'utiliser plutôt des modèles stochastiques pour extrapoler, à l'échelle de la cellule, les comportements collectifs observés à l'échelle moléculaire. L'intérêt d'une telle approche est que la communication inter-échelles est alors bi-directionnelle : les modifications observées à l'échelle de la cellule peuvent être transmises en retour aux agents moléculaires qui adapteront alors leur comportement.

Bien entendu, ces modalités ne sont pas exclusives : si, dans la démarche de modélisation, elles sont basées sur des approches différentes, elles devront être unifiées lors du passage à la simulation. Dans ce cadre, on sera attentif à ce que la méthodologie proposée permette d'“agentifier” différents types de modèles, en s'attachant tout particulièrement aux interfaces (échanges de matières, de forces, de signaux, ...) entre les différents niveaux de description.

B1.4 Modèle biologique : modulation de la localisation de la nucléoline au cours d'une infection par le virus de l'herpès

Le nucléole est le domaine nucléaire le plus volumineux des cellules eucaryotes. Du fait de ses caractéristiques de taille et de densité, il a été mis en évidence par microscopie photonique il y a plus de 100 ans. C'est une structure nucléaire complexe aux fonctions multiples. Il a ainsi été montré que cette structure, en plus du rôle qu'elle joue dans la synthèse des ribosomes, participe au contrôle de plusieurs processus fondamentaux de la vie cellulaire : la maturation et le transport de certains petits ARN, le vieillissement cellulaire, l'apoptose et la division cellulaire (Olson et al., 2000; Peterson, 1998; Politz et al., 2000; Visintin and Amon, 2000). Pour contribuer aux études qui visent à élucider les multiples fonctions des nucléoles et leur dynamique, nous – l'équipe “Régulation post-transcriptionnelle de l'expression Génique Cellulaire et Virale” du CGMC – avons développé un modèle expérimental constitué de cellules humaines infectées par le virus Herpès Simplex de type 1 (HSV-1). En effet, l'infection par HSV-1, virus à réplication intranucléaire, induit un profond remodelage de l'architecture nucléaire de la cellule hôte (Monier et al., 2000). Suite à l'infection par HSV-1, les nucléoles subissent d'importantes modifications morphologiques (Puvion-Dutilleul et al., 1997), mais le rôle de ces modifications pour l'établissement de l'infection reste à déterminer. Récemment, nous avons entrepris l'analyse protéomique du nucléole. Plus de 200 protéines différentes contenues dans les fractions nucléolaires des cellules non infectées ont été identifiées, ce qui a permis de confirmer certains rôles supposés pour les nucléoles et également de proposer de nouveaux rôles (Scherl et al., 2002). Parmi les protéines du nucléole, la nucléoline est l'une des plus abondantes (Srivastava and Pollard, 1999). Il apparaît clairement que cette protéine joue un rôle majeur dans la régulation de la synthèse des ribosomes et plus généralement dans le contrôle de l'expression génique. Par ailleurs, nous avons démontré que la nucléoline interagit avec la protéine US11 de HSV-1 (résultats non publiés). Dans une cellule en interphase, la localisation de cette protéine est principalement nucléolaire mais différentes conditions physiologiques et pathologiques, dont l'infection par HSV-1, induisent une forte délocalisation de cette dernière. C'est ce phénomène biologique que nous souhaitons explorer dans le cadre du présent projet.

B2 Description du projet

B2.1 Organisation générale

B2.1.1 Une démarche intégrée

Dans ce projet, nous nous proposons d'adopter une démarche intégrée associant des biologistes, des informaticiens et des mathématiciens. Le but de cette approche est de permettre à ces différents champs disciplinaires de développer une vision commune sur le modèle biologique d'une part et sur le modèle informatique d'autre part, par l'entremise, entre autres, de modèles mathématiques. Il nous semble en effet clair que les approches de Biologie des Systèmes ne sont viables que sous la condition de développement de cette vision intégrée. Elle seule permet de garantir la validité des hypothèses formulées, *sous la forme* d'un modèle mathématique, *à partir* du modèle informatique *sur* le modèle biologique, les différents modèles étant en permanence construits, observés et validés conjointement par les trois communautés. Il s'agit en particulier d'observer les mécanismes auto-organisateur qui conduisent à l'apparition de macrostructures et, lorsque ces mécanismes peuvent être décrits comme des concentrations de mesures dans des modèles stochastiques (par exemple des transitions de phase), de mettre en évidence les processus sous-jacents qui conduisent à cette transition. Le modèle multi-agents et la description stochastique permettraient ainsi de relier deux niveaux d'observation d'un même phénomène : un niveau microscopique (moléculaire) qui, bien que totalement a-structuré, produit un niveau macroscopique (le nucléole) fortement structuré. Il s'agit donc de développer un outil de modélisation permettant de programmer le comportement *local* d'entités (ici des protéines) et d'observer l'évolution spatio-temporelle *globale* qui en résulte. Un tel outil permettrait alors au biologiste d'émettre, en retour, des hypothèses sur la causalité biologique à l'origine des structures globales.

Une telle démarche ne peut cependant aboutir que si elle est ancrée sur une problématique biologique ou, tout au moins, sur un modèle biologique. Ainsi, les projets de cellule virtuelle cités ci-dessus (paragraphe B1.1) se sont tous donnés de tels modèles⁵. Dans notre cas, nous ne souhaitons évidemment pas rivaliser avec de tels projets; c'est pourquoi nous nous proposons de modéliser des structures relativement simples (de par le nombre de protéines mis en jeu mais aussi de par la variété limitée des interactions entre ces protéines). Nous centrerons donc nos travaux sur un sous-système cellulaire (le nucléole) et sur une condition métabolique particulière qui modifie ce sous-système (l'infection par le virus HSV-1 qui induit une délocalisation de la nucléoline). Ce sous-système présente donc l'intérêt de poser une question *directe* au modèle, ce qui permettra une maturation réciproque des démarches méthodologique et biologique.

B2.1.2 Partenaires

Ce projet s'inscrit dans la perspective, plus large, de création d'un groupe multi-laboratoires de recherche en Biologie des Systèmes et Modélisation Cellulaire (BSMC⁶) sur le campus de la Doua. BSMC regroupe actuellement des équipes de recherche appartenant à cinq laboratoires du campus (BF2I, CGMC, LIRIS, Maply et PRISMa) dans le but de développer des recherches pluridisciplinaires plus particulièrement centrées sur l'analyse du transcriptome (qui apparaît comme un niveau d'étude pertinent pour de nombreux mécanismes cellulaires). Cependant, le développement d'une vision intégrée impose d'étudier d'autres niveaux d'organisation ainsi que les questions méthodologiques. C'est dans ce cadre que le présent projet est proposé par quatre des laboratoires du groupe :

⁵ *Mycoplasma genitalium* pour le projet E-Cell, *Bacillus subtilis* pour BioSpice et *E. Coli* pour CyberCell.

⁶ <http://bsmc.insa-lyon.fr>

PRISMa (Coordinateur) : Dans le cadre de sa thématique de recherche “Systèmes Autonomes, Architecture et Comportements”, le laboratoire PRISMa (EA INSA-UCBL 2058) développe depuis plusieurs années des méthodes et des outils de modélisation objet et/ou agents. L’équipe INSA-ALAB (Artificial Life And Behaviour), en particulier, développe des approches de modélisation qualitative bio-inspirées (Bournez et al., 2001; Beslon, 2002) ou bio-orientées (Knibbe et al., 2003). Il a en particulier développé un premier outil de modélisation multi-agents pour la simulation de grands amas de protéines (3DSpi) (Soula et al., 2004). Ce projet entre donc pleinement dans le domaine de compétence de cette équipe qui en assurera la coordination.

CGMC : Le Centre de Génétique Moléculaire et Cellulaire (UMR CNRS 5534) apporte au projet son expertise dans le domaine du métabolisme nucléaire. Pour cela, deux équipes du CGMC seront impliquées dans le projet :

- Signalisations et Identités Cellulaires (SIC). Cette équipe a participé au développement de la première version de 3DSpi. Elle apportera donc ses compétences en termes de biologie cellulaire, mais aussi en termes de modélisation et de Biologie des Systèmes.
- L’équipe “Régulation Post-Transcriptionnelle de l’Expression Génique Cellulaire et Virale” qui apportera la connaissance biologique sur les objets biologiques modélisés (nucléole et HSV-1). C’est aussi à ce partenaire que reviendra la tâche de validation biologique des hypothèses formulées dans le cadre du modèle.

Maply : Le laboratoire de Mathématiques APpliquées de LYon (UMR CNRS 5585) comprend une thématique de recherche “modélisation mathématique des sciences du vivant” et mène des travaux sur les systèmes de particules en interaction. Les outils stochastiques développés dans ce cadre semblent particulièrement appropriés puisqu’ils permettent de préciser les mécanismes régissant la dynamique du modèle individu-centré.

LIRIS : Le Laboratoire d’InfoRmatique en Image et Systèmes d’information (FRE 2672 - CNRS), à travers son axe ”Modélisation et réalité augmentée”, développe des recherches sur la modélisation géométrique et à la simulation d’objets complexes. Le présent projet constitue un champ d’application supplémentaire pour des travaux déjà menés sur les systèmes de particules et la modélisation surfacique, qu’il s’agira d’adapter au cas de la représentation de différents compartiments cellulaires et d’intégrer dans l’outil 3DSpi.

B2.2 Aspects méthodologiques : vers une cellule virtuelle

B2.2.1 3DSpi : un outil de modélisation multi-agents à l’intention des biologistes

Depuis trois ans, l’équipe ALAB (Artificial Life And Behaviour) du laboratoire PRISMa développe un logiciel de modélisation et de simulation individu-centrée : SPEAR (Simulator of Physical Environment for Animat Research). Ce logiciel s’appuie sur une méthodologie de modélisation stricte puisque le modélisateur n’a accès qu’aux entités locales, indépendamment des conditions globales de la simulation. SPEAR est particulièrement bien adapté à la modélisation multi-agents de grands amas de protéines car :

- il permet de modéliser le comportement local des protéines et d’observer les résultats structurels globaux,
- il permet de calculer la dynamique des objets émergents (ici la dynamique des amas multi-protéiques).

Nous avons proposé une version de SPEAR dédiée à ce type de modélisation, 3DSpi, dans laquelle les protéines sont modélisées par leur domaine spatial d’interaction et un ensemble de règles d’interaction et de déplacement (la biochimie des protéines est donc résumée à ses conséquences en

termes d'interaction). À partir de ce modèle local (un ou plusieurs types de protéines en nombre variable), 3DSpi modélise un environnement fini et simule le déplacement et les interactions entre les entités. On peut alors observer l'émergence de structures macroscopiques telles que des agrégats de protéines et observer leur évolution spatio-temporelle (Soula et al., 2004). Les résultats obtenus à l'issue des simulations sont alors des surfaces d'évolution (évolution de la distribution du nombre d'agrégats ou de leur taille, ...). 3DSpi peut aussi fournir des images et des vidéos qui, bien peu informatives quantitativement, permettent aux biologistes d'appréhender visuellement le comportement du système et constituent ainsi un guide pertinent pour la formulation d'hypothèses. 3DSpi permet en outre de réaliser, sur le modèle informatique, des expériences similaires à celles réalisées sur le modèle biologique (marquage "fluorescent" et observation du devenir des protéines marquées par exemple). Il est ainsi possible de comparer les résultats obtenus dans les deux expériences et ainsi de renforcer (ou, au contraire, d'infirmier) les hypothèses formulées.

B2.2.2 De 3DSpi à un modèle de cellule virtuelle

Pour l'instant, le modèle développé est resté relativement limité puisque nous nous sommes contentés de modéliser des domaines d'interaction sphériques isotropes. De plus, le nombre de protéines modélisées était limité à quelques centaines pour des raisons de calculabilité des simulations. Dans le cadre du présent projet, nous proposons d'accroître les capacités de modélisation de 3DSpi dans différentes directions :

Augmenter le nombre d'objets en interaction ; Il s'agit pour cela d'augmenter la capacité de calcul, d'une part en faisant appel à des outils logiciels (optimisation) et d'autre part à des outils matériels (parallélisation). Dans le cadre du présent projet, ces deux directions seront poursuivies simultanément puisque nous utiliserons le cluster de calcul du groupe BSMC (20 nœuds à 2.7 GHz) pour simuler le modèle.

Augmenter le nombre de classes d'objets ; 3DSpi est pour l'instant limité à la modélisation de deux types d'objets différents. Cette limitation devra être dépassée pour permettre la modélisation d'un nombre quelconque de types d'objets. Pour cela, il pourra être nécessaire de développer un langage d'expression des relations entre les objets modélisés pour permettre aux biologistes d'exprimer simplement le modèle.

Augmenter le pouvoir de description ; Les domaines d'interaction des entités sont pour l'instant supposés isotropes. Si cette hypothèse semble réaliste dans le cas de la modélisation du nucléole, il est évident qu'elle limite fondamentalement le pouvoir de description de 3DSpi. Il est donc nécessaire d'affiner le modèle moléculaire sous-jacent, sans accroître outre mesure la complexité algorithmique du modèle. Une approche serait de modéliser les molécules par une séquence de domaines d'interaction organisés dans l'espace (en analogie avec les structures secondaire et tertiaire des protéines).

Agentification multi-niveaux ; La modélisation multi-échelles sera progressivement introduite dans 3DSpi en intégrant des entités plus complexes que les molécules considérées individuellement. Dans un premier temps, nous introduirons des membranes déformables⁷. Nous nous intéresserons ensuite aux propriétés moléculaires de ces membranes pour affiner les modalités d'interactions entre les différents niveaux de modélisation (introduction de protéines membranaires par exemple). Enfin, cette démarche sera étendue pour permettre la modélisation d'autres compartiments cellulaires (par exemple le processus de transcription/traduction) dont la modélisation fine n'est pas nécessaire mais dont le fonctionnement influe sur le métabolisme cellulaire.

⁷Dans le cas de la modélisation du nucléole, il s'agira de la membrane nucléaire.

B2.3 Aspects biologiques : étude de la délocalisation de la nucléoline au cours de l'infection par le virus HSV-1

B2.3.1 Modélisation du nucléole

Dans un premier temps, nous nous attacherons à modéliser le nucléole. Pour cela nous ferons quelques hypothèses simplificatrices nécessaires mais toujours basées sur des connaissances biologiques. Nous utiliserons 3DSpi pour modéliser une structure à trois composants :

1. La nucléoline, dont on sait qu'elle représente 10% des protéines nucléolaires totales.
2. Une deuxième protéine, capable d'interagir avec la nucléoline, représentant ses partenaires au sein du nucléole. Les connaissances biologiques à ce sujet restent encore peu précises et devront donc être affinées et/ou approximées.
3. Une troisième entité sera chargée de représenter le reste des protéines du nucléole.

Les paramètres des interactions inter-protéiques seront alors ajustés pour obtenir une structure nucléaire (un ou plusieurs "nucléoles virtuels") qualitativement comparable aux structures biologiques.

B2.3.2 Modélisation de l'infection par le virus de l'Herpès

Nous utiliserons 3DSpi pour poser une double question : est-ce que la délocalisation de la nucléoline suite à une infection est susceptible d'être causée,

1. par une modification post-traductionnelle de cette protéine ?
2. par une interaction préférentielle avec un composant viral ?

Pour répondre à la première question, nous changerons les valeurs du paramètre d'interaction entre la nucléoline et les autres protéines et analyserons le comportement du système. Pour répondre à la deuxième question, nous modéliserons une quatrième entité représentant la protéine virale US11 (voir paragraphe B1.4). Nous ferons alors varier les interactions entre la nucléoline et la protéine US11 et analyserons le comportement du système. Dans les deux cas, nous nous baserons sur les observations faites en microscopie confocale, notamment la cinétique de diffusion de la nucléoline hors du nucléole, pour estimer la pertinence et ajuster les paramètres du modèle.

B2.3.3 Analyse du modèle et retour à la biologie

La réalisation d'une cellule virtuelle n'est pas un but en soit ; une telle démarche n'acquiert son intérêt que s'il est possible de "fermer la boucle" et de revenir ainsi à l'objet biologique. Or, ce retour à la biologie ne peut pas être immédiat puisque, si la simulation peut exprimer *qualitativement* les propriétés recherchées, elle ne représentera pas *quantitativement* l'objet biologique. C'est pourquoi le couple biologie-informatique ne peut pas fonctionner seul. Il est indispensable d'introduire une démarche de modélisation mathématique, à *partir* du modèle informatique, pour exprimer les propriétés émergentes avec une généricité suffisante pour englober le modèle informatique et l'objet biologique.

Ainsi, la modélisation stochastique, guidée par la simulation, permet de relier les distributions de probabilités au niveau local (interactions bruitées) avec les probabilités au niveau global (stabilité des structures émergentes)(Durrett, 1998; Blake et al., 2003). Cette modélisation peut, contrairement à la simulation, être comparée aux distributions de probabilités mesurées expérimentalement et revêtir ainsi un caractère explicatif. C'est à travers la création de ce cercle vertueux (Biologie → Informatique → Mathématiques → Biologie) que nous pensons pouvoir aboutir à la mise en œuvre d'une démarche de Biologie des Systèmes fructueuse.

B3 Résultats attendus

Du fait de la démarche proposée, intégrant une dimension méthodologique et une dimension biologique, nous attendons des résultats de deux ordres. D’une part, le projet devra avoir des retombées en termes de modélisation cellulaire individu-centrée (méthodes et outils), d’autre part les modèles mis en œuvre devront permettre d’éclairer le mécanisme de délocalisation de la nucléoline au cours de l’infection par le virus HSV-1.

B3.1 Aspects méthodologiques

Il est bien évident que nous n’entendons pas, dans ce projet, rivaliser avec les projets internationaux que sont E-Cell, BioSpice et CyberCell. Sur un plan strictement méthodologique, nous proposons une approche transdisciplinaire de la Biologie des Systèmes et ce projet nous permettra de la mettre en œuvre sur une problématique réelle. Nous espérons ainsi pouvoir démontrer la pertinence d’une approche individu-centrée dans le cadre de la Biologie des Systèmes et de la modélisation cellulaire.

Par ailleurs, même si les résultats en termes de théorie informatique ne sont pas directement attendus dans le strict cadre de ce projet, la question de l’intégration multi-modèles et multi-échelles est encore largement ouverte. Une telle intégration présente des avantages indéniables pour la modélisation cellulaire mais reste encore à mettre en œuvre ; une des retombées attendues de ce projet est donc de proposer une ou plusieurs approches pour une telle intégration. Pour cela, nous nous proposons d’adopter une démarche itérative au cours de laquelle les différents types de modèles seront intégrés successivement :

Modèle multi-agents : Dans un premier temps, le modèle multi-agents tel qu’il est actuellement implémenté dans l’outil 3DSpi sera étendu de façon à en accroître les capacités descriptives (voir section B2.2.2). À l’issue de cette première phase, un premier modèle de nucléole pourra être construit. Par ailleurs, des modèles d’autres “compartiments cellulaires” pourront être envisagés pour valider l’approche méthodologique indépendamment de la modélisation du nucléole (modélisation de protéines membranaires, de l’actine, de la tubuline, ...).

Modèle géométrique : Une ou plusieurs entités modélisées sous une forme géométrique seront alors intégrées au modèle multi-agents. Nous intégrerons en particulier un modèle de membrane nucléaire. En effet, le nucléole étant un domaine nucléaire important (en taille comme en densité), il est susceptible d’interagir *physiquement* avec la membrane nucléaire.

Modèle mathématique : Dans un troisième temps, différents compartiments cellulaires seront intégrés au modèle sous une forme “mathématique”. Nous intégrerons en particulier la machinerie de transcription/traduction (ainsi que la dégradation des protéines). En effet, dans la version actuelle du modèle, les concentrations de protéines sont considérées constantes. Cette évolution nous permettra de faire varier les taux de transcription – et donc les concentrations – des différentes protéines en interaction et d’observer les effets induits sur les macro-structures.

À l’issue de ces trois étapes, nous disposerons d’une part d’un modèle multi-échelles du nucléole mais aussi d’une méthodologie *testée* (sinon éprouvée) de modélisation cellulaire.

L’analyse des modèles produits se place dans une thématique émergente au sein des sciences mathématiques, à savoir l’étude de la complexité. Il serait prétentieux de prétendre résoudre, à travers ce projet, la question de la caractérisation mathématique de l’émergence ; cependant, nous espérons pouvoir apporter une contribution significative dans ce domaine, de par l’utilisation des modèles stochastiques. En effet, ces modèles sont pour l’instant restés essentiellement cantonnés

à l'étude des phénomènes physiques (modélisation des systèmes de particules en interaction (Taglagrand, 2003)) et, plus confidentiellement, à la biologie des populations ou à la sociologie. Les premiers résultats obtenus nous laissent penser qu'ils pourraient être bien adaptés à l'étude des mécanismes d'auto-organisation moléculaires et cellulaires.

Enfin, le dernier résultat attendu est d'ordre pratique puisqu'il s'agit de mettre à la disposition de la communauté scientifique un outil de modélisation multi-agents adapté à la biologie cellulaire. Dans ce cadre, il s'agit d'affiner l'outil existant (3DSpi). Nous avons déjà évoqué l'accroissement de son pouvoir descriptif; il s'agira ensuite d'en accroître l'"utilisabilité". Pour cela il sera probablement nécessaire de définir un langage de développement (par exemple basé sur SBML⁸) pour permettre aux biologistes de décrire directement le comportement des entités modélisées.

B3.2 Aspects biologiques

La question des modifications viro-induites dans le cas de l'infection par HSV-1 est encore ouverte, même si les interactions avec la nucléoline – et donc avec le nucléole – semblent établies. Dans ce cadre, la démarche de modélisation proposée apparaît comme clairement complémentaire des analyses protéomiques menées par ailleurs au CGMC. En effet, elle permet de franchir la "barrière des échelles" en montrant les conséquences macroscopiques (donc ici à l'échelle du nucléole) induites par des modifications microscopiques (à l'échelle moléculaire). Pour autant, cette démarche ne revêt aucun caractère de preuve. C'est pourquoi, sur le plan biologique, nous attendons de ce projet, dans un premier temps, qu'il nous permette de proposer des hypothèses pertinentes *et testables* sur le fonctionnement du nucléole et les modulations induites par une infection HSV-1 (en identifiant les paramètres locaux responsables de la délocalisation). Grâce à ces hypothèses, nous espérons être en mesure de "retourner à la cellule" pour vérifier expérimentalement les mécanismes proposés. On pourra, par exemple, envisager de transfecter les cellules par un vecteur d'expression de la protéine US11 et observer le comportement de la nucléoline.

Dans un second temps, à travers la modélisation mathématique issue de l'observation des modèles informatiques, nous chercherons à isoler les mécanismes émergents observés. Ce n'est qu'à l'issue de cette deuxième démarche que le modèle proposé, mathématisé, pourra revêtir un caractère explicatif, apportant alors des connaissances biologiques (et non seulement des hypothèses) et confortant ainsi la démarche méthodologique elle-même.

B3.3 Échéancier

Le projet est proposé pour une durée de 36 mois au cours de laquelle biologistes, informaticiens et mathématiciens travailleront de concert. Cette démarche intégrée est cependant soumise à des résultats intermédiaires importants en termes de modélisation et de simulation : il est en effet primordial que les premières versions du "nucléole virtuel" soient disponibles suffisamment rapidement pour permettre aux mathématiciens d'analyser les comportements observés et aux biologistes de mettre en œuvre des expérimentations permettant de tester les hypothèses formulées. C'est pourquoi nous distinguons trois phases :

Première phase : mise en place D'une durée de six mois, cette phase de mise en place permettra aux différentes équipes d'acquérir les connaissances transdisciplinaires nécessaires au bon déroulement du projet. C'est aussi durant cette première phase que le laboratoire PRISMA conduira les premières améliorations du logiciel 3DSpi afin de permettre la modélisation du nucléole.

⁸Systems Biology Markup Language, <http://www.sbw-sbml.org>. L'utilisation de SBML permettrait en outre d'intégrer 3DSpi à d'autres outils de modélisation via SBW (Systems Biology Workbench).

Deuxième phase : modélisation D'une durée de 18 mois, cette phase sera entièrement dévolue au travail de modélisation du nucléole. Elle sera divisée en trois sous-phases permettant d'accroître progressivement le pouvoir d'expression du modèle en intégrant les compétences des différentes équipes partenaires (voir section B2.2.2). C'est dans cette optique que nous demandons, pour toute la durée de cette phase, le financement d'un contrat de post-doctorant de 18 mois au laboratoire PRISMa. À l'issue de cette deuxième phase nous serons en mesure de mettre une première version de l'outil 3DSpi à disposition de la communauté scientifique et de produire un premier rapport d'étape.

Troisième phase : analyse D'une durée de 12 mois cette phase permettra aux mathématiciens et aux biologistes d'exploiter les données issues du modèle (nous demandons pour cela le financement d'un contrat post-doctoral d'une durée de 18 mois au laboratoire Maply). À cette fin, il sera probablement nécessaire de mener des campagnes d'expérimentation (par exemple en microscopie confocale) pour estimer les distributions de probabilités locales et globales permettant, d'une part d'affiner les simulations, et d'autre part de valider les résultats. Parallèlement, les deux équipes PRISMa et LIRIS continueront le travail sur l'outil de simulation pour permettre la simulation d'autres compartiments cellulaires.

Bien entendu, cette décomposition en trois étapes ne constitue pas un cadre rigide pour l'intervention des différentes équipes partenaires. Nous nous proposons en effet d'adopter une démarche de modélisation itérative, ce qui nous permettra d'accroître progressivement – et parallèlement – la complexité du modèle et de la simulation, donc la complexité de l'analyse mathématique et la portée des résultats biologiques obtenus.

Références

- Amar, P., Ballet, P., Barlovatz-Meimon, G., Benecke, A., Bernot, G., Bouligand, Y., Bourguine, P., Delaplace, F., Delosme, J.-M., Demarty, M., Fishov, I., Fourmentin-Guilbert, J., Fralick, J., Giavitto, J.-L., Gleyse, B., Godin, C., Incitti, R., Kepes, F., Lange, C., Sceller, L. L., Loutellier, C., Michel, O., Molina, F., Monnier, C., Natowicz, R., Norris, V., Orange, N., Pollard, H., Raine, D., Ripoll, C., Rouviere-Yaniv, J., Jr, M. S., Soler, P., Tambourin, P., Thellier, M., Tracqui, P., Ussery, D., Vincent, J.-C., Vannier, J.-P., Wiggins, P., and Zemirline, A. (2002). Hyperstructures, genome analysis and i-cell. *Acta Biotheor.*, 50(4) :357–373.
- Atamas, S. (1996). Self-organization in computer simulated selective systems. *Biosystems*, 39 :143–151.
- Ball, R., Bowler, N., Sander, L., and Somfai, E. (2001). Off-lattice noise reduction and the ultimate scaling of DLA in two dimensions. *ArXiv Condensed Matter e-prints*.
- Becquet, C., Charavay, C., Herr, A., Martin, T., Morlé, F., Déandréa, A., Beslon, G., and Gandrillon, O. (2002). Modeling protein-protein interactions by using multi-agents systems. In *Journées Post-Génomiques de la Doua (JPGD)*.
- Beslon, G. (2002). Coopération implicite et émergence dans des groupes d'agents. In *Guillot A., Daucé E. (Coord.) Approche Dynamique de la Cognition Artificielle*, pages 267–283. Hermès (Paris).
- Blake, W., M, K., Cantor, C., and Collins, J. (2003). Noise in eukaryotic gene expression. *Nature*, 422 :633–637.
- Bournez, C. (2001). *Une architecture multi-agents réflexive pour le contrôle de systèmes de production distribués et hétérogènes*. PhD Thesis, INSA de Lyon - PRISMa.
- Bournez, C., Beslon, G., and Favrel, J. (2001). Ocean : A multi-agent market for distributed system control. In *Proc. of the Multiagent Based Modelling and Simulation Workshop*, pages 934–940.
- Duboz, R. (2004). *Intégration de modèles hétérogènes pour la modélisation et la simulation de systèmes complexes*. PhD Thesis, Université du Littoral.

- Duboz, R., Ramat, E., and Preux, P. (2003). Scale transfert modeling : using emergent computation for coupling an ordinary differential equation system with a reactive agent model. *Systems Analysis Modelling Simulation*, 43 :793–814.
- Durrett, R., editor (1998). *Lecture Notes on Particle Systems and Percolation*. Brooks Cole Statistics/Probability Series.
- Ideker, T., Galitski, T., and Hood, L. (2001). A new approach to decoding life : systems biology. *Annu. Rev. Genomics Hum. Genet.*, 2 :343–372.
- Kitano, H. (2002). Systems biology : a brief overview. *Science*, 295 :1662–1664.
- Knibbe, C., Beslon, G., and Fayard, J.-M. (2003). Modeling evolutionary interactions between genomic and functional structures. In *Proc. of the 4th Int. Conf. on Systems Biology*, pages 163–164.
- Kupiec, J. (1997). A darwinian theory for the origin of cellular differentiation. *Mol Gen Genet*, 255 :201–208.
- Levsky, J. and Singer, R. (2003). Gene expression and the myth of the average cell. *Trends Cell Biol.*, 13 :4–6.
- Mistelli, T. (2001). Protein dynamics : implication for nuclear architecture and gene expression. *Science*, 291 :843–847.
- Monier, K., Armas, J., Etteldorf, S., Ghazal, P., and Sullivan, K. (2000). Annexation of the interchromosomal space during viral infection. *Nat Cell Biol*, 2 :661–665.
- Olson, M., Dundr, M., and Szebeni, A. (2000). The nucleolus : an old factory with unexpected capabilities. *Trends Cell Biol*, 10 :189–196.
- Peterson, T. (1998). The plurifunctional nucleolus. *Nucleic Acids Res*, 26 :3871–3876.
- Politz, J., Yarovoi, S., Kilroy, S., Gowda, K., Zwieb, C., and Pederson, T. (2000). Signal recognition particle components in the nucleolus. *Proc Natl Acad Sci USA*, 97 :55–60.
- Puvion-Dutilleul, F., Besse, S., Diaz, J.-J., Kindbeiter, K., Vigneron, M., Warren, S., Kedinger, C., Madjar, J.-J., and Puvion, E. (1997). Identification of transcription factories in nuclei of HeLa cells transiently expressing the us11 gene of herpes simplex type 1. *Gene Expression*, 6 :315–332.
- Scherl, A., Couté, Y., Déon, C., Callé, A., Kindbeiter, K., Sanchez, J.-C., Greco, A., Hochstrasser, D., and Diaz, J.-J. (2002). Functional proteomic analysis of the human nucleolus. *Mol Biol Cell*, 13 :4100–4109.
- Soula, H., Robardet, C., Perrin, F., Gripon, S., Beslon, G., and Gandrillon, O. (2004). 3DSpi, a multi-agent-based software for generating and analyzing dynamic structures generated through protein-protein interactions. *Submitted to Genome Biology*.
- Stuart, J., Segal, E., Koller, D., and Kim, S. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science*, 302 :249–255.
- Talagrand, M. (2003). Self organization in the low temperature region of a spin glass model. *Reviews in Mathematical Physics*, 15(1) :1–78.
- Theraulaz, G. and Spitz, F. (1997). *Auto-organisation et comportement*. Hermès (Paris).
- Tononi, G., Sporns, O., and Edelman, G. (1999). Measures of degeneracy and redundancy in biological networks. *Proc. Natl. Acad. Sci. USA*, 96 :3257–3262.
- Vespignani, A. (2003). Evolution thinks modular. *Nat Genet*, 35 :118–119.
- Visintin, R. and Amon, A. (2000). The nucleolus : the magician’s hat for cell cycle tricks. *Curr Opin Cell Biol*, 12 :372–377.
- Weiss, G., editor (1999). *Multiagent Systems, a Modern Approach to Distributed Artificial Intelligence*. MIT Press (Cambridge).