Numéro d'ordre 2010-ISAL-0045
Année 2010

Stochasticité de l'expression génique et régulation transcriptionnelle

Modélisation de la dynamique spatiale et temporelle des structures multiprotéiques

Thèse présentée par
Antoine Coulon, ingénieur en informatique

Devant
L'Institut National des Sciences Appliquées de Lyon

Pour obtenir
Le grade de docteur

Formation doctorale
Informatique et Mathématiques

Spécialité
Informatique

Soutenue le 1er juillet 2010 devant le jury composé de :

Alain Arnéodo	Directeur de recherche, CNRS, École Normale Supérieure de Lyon (président)
Hugues Berry	Chargé de recherche HDR, INRIA (examinateur)
Guillaume Beslon	Professeur, INSA de Lyon (directeur de thèse)
Xavier Darzacq	Chargé de recherche HDR, CNRS, École Normale Supérieure (rapporteur)
Hidde de Jong	Directeur de recherche, INRIA (rapporteur)
Olivier Gandrillon	Directeur de recherche, CNRS, Université de Lyon (directeur de thèse)
Raphaël Métivier	Chargé de recherche, CNRS, Université Rennes 1 (examinateur)

INSA Direction de la Recherche - Ecoles Doctorales – Quadriennal 2007-2010

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
CHIMIE	**CHIMIE DE LYON** http://sakura.cpe.fr/ED206 M. Jean Marc LANCELIN Insa : R. GOURDON	M. Jean Marc LANCELIN Université Claude Bernard Lyon 1 Bât CPE 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cedex Tél : 04.72.43 13 95 Fax : lancelin@hikari.cpe.fr
E.E.A.	**ELECTRONIQUE, ELECTROTECHNIQUE, AUTOMATIQUE** http://www.insa-lyon.fr/eea M. Alain NICOLAS Insa : C. PLOSSU ede2a@insa-lyon.fr Secrétariat : M. LABOUNE AM. 64.43 – Fax : 64.54	M. Alain NICOLAS Ecole Centrale de Lyon Bâtiment H9 36 avenue Guy de Collongue 69134 ECULLY Tél : 04.72.18 60 97 Fax : 04 78 43 37 17 eea@ec-lyon.fr Secrétariat : M.C. HAVGOUDOUKIAN
E2M2	**EVOLUTION, ECOSYSTEME, MICROBIOLOGIE, MODELISATION** http://biomserv.univ-lyon1.fr/E2M2 M. Jean-Pierre FLANDROIS Insa : H. CHARLES	M. Jean-Pierre FLANDROIS CNRS UMR 5558 Université Claude Bernard Lyon 1 Bât G. Mendel 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cédex Tél : 04.26 23 59 50 Fax 04 26 23 59 49 06 07 53 89 13 e2m2@biomserv.univ-lyon1.fr
EDISS	**INTERDISCIPLINAIRE SCIENCES-SANTE** Sec : Safia Boudjema M. Didier REVEL Insa : M. LAGARDE	M. Didier REVEL Hôpital Cardiologique de Lyon Bâtiment Central 28 Avenue Doyen Lépine 69500 BRON Tél : 04.72.68 49 09 Fax :04 72 35 49 16 Didier.revel@creatis.uni-lyon1.fr
INFOMATHS	**INFORMATIQUE ET MATHEMATIQUES** http://infomaths.univ-lyon1.fr M. Alain MILLE Secrétariat : C. DAYEYAN	M. Alain MILLE Université Claude Bernard Lyon 1 LIRIS - INFOMATHS Bâtiment Nautibus 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cedex Tél : 04.72. 44 82 94 Fax 04 72 43 13 10 infomaths@bat710.univ-lyon1.fr - alain.mille@liris.cnrs.fr
Matériaux	**MATERIAUX DE LYON** M. Jean Marc PELLETIER Secrétariat : C. BERNAVON 83.85	M. Jean Marc PELLETIER INSA de Lyon MATEIS Bâtiment Blaise Pascal 7 avenue Jean Capelle 69621 VILLEURBANNE Cédex Tél : 04.72.43 83 18 Fax 04 72 43 85 28 Jean-marc.Pelletier@insa-lyon.fr
MEGA	**MECANIQUE, ENERGETIQUE, GENIE CIVIL, ACOUSTIQUE** M. Jean Louis GUYADER Secrétariat : M. LABOUNE PM : 71.70 –Fax : 87.12	M. Jean Louis GUYADER INSA de Lyon Laboratoire de Vibrations et Acoustique Bâtiment Antoine de Saint Exupéry 25 bis avenue Jean Capelle 69621 VILLEURBANNE Cedex Tél :04.72.18.71.70 Fax : 04 72 43 72 37 mega@lva.insa-lyon.fr
ScSo	**ScSo*** M. OBADIA Lionel Insa : J.Y. TOUSSAINT	M. OBADIA Lionel Université Lyon 2 86 rue Pasteur 69365 LYON Cedex 07 Tél : 04.78.69.72.76 Fax : 04.37.28.04.48 Lionel.Obadia@univ-lyon2.fr

*ScSo : Histoire, Geographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

Remerciements

Je souhaite tout d'abord exprimer mes profonds remerciements aux membres de mon jury : à Hidde de Jong et à Xavier Darzacq, pour avoir lu et rapporté mon manuscrit de thèse, et à Raphaël Métivier, Alain Arnéodo et Hugues Berry, pour avoir accepté de faire partie de mon jury. C'est un plaisir et un honneur que d'exposer mes travaux à leurs avis critiques et d'avoir à faire face à des compétences aussi diverses que les leurs. Chacun de leurs points de vue m'est précieux.

Mes directeurs de thèse ont eu et auront sur moi une influence capitale. En plus d'avoir stimulé mon intérêt pour la recherche, de m'avoir laissé une liberté de travail sans pareil et d'avoir été des guides avisés tout au long de cette thèse, c'est une façon de penser et d'aborder la science que Guillaume et Olivier m'ont enseignée. Je souhaite ici leur faire part de toute ma reconnaissance.

J'adresse mes sincères remerciements à Pierre Couble, Guy Mouchiroud, Atilla Baskurt et Bernard Perroche, non seulement pour m'avoir accueilli au sein de leurs laboratoires – le CGMC et le LIRIS – mais également pour croire aux approches pluridisciplinaires (parfois éloignées de leurs préoccupations de biologistes et d'informaticiens) et ainsi soutenir les travaux de mes deux équipes d'accueil. Je tiens également à remercier Jean-François Boulicaut, responsable de l'équipe Turing au LIRIS (très récemment devenue l'équipe Combining), pour s'être toujours soucié des bonnes conditions de travail de nos troupes et pour ne jamais hésiter à prendre le temps de discuter.

Il va de soi que ce travail n'aurait pas pu voir le jour sans financement. Plus qu'une simple source de moyens, c'est un soutien à des thématiques de recherche fondamentale que je souhaite remarquer ici. À ce titre, je remercie le cluster ISLE de la région Rhône-Alpes pour m'avoir octroyé une allocation doctorale, l'Institut Rhônalpin des Systèmes Complexes (IXXI) pour le financement du projet pluridisciplinaire dans lequel s'inscrivent mes travaux, ainsi que pour son accueil sur le site de l'institut, et le centre de calcul de l'IN2P3 pour nous avoir donné accès à ses moyens de calcul.

De nombreuses discussions avec des collègues d'horizons divers doivent être saluées ici. Merci aux membres du projet Stochagene, et en particulier à Jean-Jacques Kupiec, Andras Paldi et Carsten Carlberg, pour des débats riches et stimulants. Merci au groupe de travail SMABio – et par cet intermédiaire, au Programme d'Épigénomique – pour des discussions au début de ma thèse concernant la simulation spatiale. Je remercie également les membres du groupe de Biologie des Systèmes et Modélisation Cellulaire (BSMC) et en particulier Daniel Kahn dont les remarques m'ont mis sur la piste de ce qui constitue aujourd'hui un aspect important de mon travail de thèse.

Les membres de mes deux équipes d'accueil auront fait de mon environnement pendant ces années de thèse un cadre de travail agréable et un milieu scientifique stimulant. Je tiens à exprimer mon amitié à mes actuels ou anciens camarades thésards et postdoc. Parmi bien d'autres, je pense à Camila et à José, côté pipettes, et à Yolanda, Virginie, Hédi, David, Gaël et Loïc, côté claviers.

J'adresse un remerciement très particulier à Carole, sans qui beaucoup de choses n'auraient évidemment pas été ce qu'elles sont. Cette thèse – et moi-même – lui devons énormément. Merci aussi à Nox pour m'avoir régulièrement empêché de faire la biblio.

Une pensée s'envole vers la région parisienne. Je remercie Yaya, entre autres pour m'avoir laissé gérer nos TP de physique-chimie en 1ère S3.

Mes amis proches – Guille, Christophe, Emma et Alice – ont été à leur manière un soutien essentiel durant ces années. J'hésite encore entre les remercier ou leur en vouloir pour avoir sans cesse essayé de m'écarter du droit chemin !

Plus que des remerciements sont adressés à Léa. Elle a, entre autres choses, su me donner ce dont j'avais besoin pendant cette rédaction. Je la remercie de tout cœur.

Je tiens à terminer ces quelques lignes en remerciant chaleureusement ma famille. Leur soutien et leurs encouragements tout au long de mon parcours m'ont particulièrement touché. Merci à Elodie pour s'être bien occupée de son petit frère ! Je ne remercie pas seulement mon père pour avoir tenté de me distraire de ma rédaction avec ses problèmes d'astrophysique, ni ma mère pour avoir traqué jusqu'à la moindre faute dans une bonne partie de ce manuscrit. Je leur suis surtout reconnaissant pour m'avoir donné goût à la science et à tant d'autres choses.

Table des matières

Table des matières **7**

Table des figures **11**

Introduction **13**

I Origines moléculaires de la stochasticité de l'expression génique **17**

1 Séparation, quantification et identification des sources de stochasticité . . . 18
1.1 Mesures expérimentales . . . 18
1.2 Stochasticités intrinsèques *vs* extrinsèques . . . 19
1.3 Signatures spécifiques des différentes sources . . . 21
2 Sources de stochasticité gène-intrinsèques . . . 23
2.1 Modélisation de l'expression génique . . . 24
2.2 Synthèse/dégradation de protéines et des transcrits . . . 25
2.3 Bursts traductionnels . . . 27
2.4 Dynamique du promoteur . . . 28
2.5 Relations entre échelles de temps . . . 30
2.6 Distribution des délais : élongation, export, dégradations . . . 31
2.7 Aspects spatiaux . . . 32
3 Propagation et modulation de la stochasticité . . . 34
3.1 Transmission des fluctuations gène-extrinsèques . . . 34
3.2 Modulation par l'autorégulation . . . 35
3.3 Motifs de gènes : corrélations constructives et destructives . . . 35
3.4 Ergodicité : mémoire cellulaire et interactions cellule-cellule . . . 36
4 Organisation spatio-temporelle du noyau cellulaire . . . 39
4.1 Un espace hétérogène et dynamique . . . 39
4.2 Déplacements moléculaires . . . 40
4.3 Dynamique des corps nucléaires . . . 42
4.4 Architecture nucléaire et expression génique . . . 44
4.5 Le noyau, un système auto-organisé ? . . . 47
5 Dynamique de la régulation transcriptionnelle . . . 48
5.1 Promoteurs eucaryotes . . . 48
5.2 Mobilité des molécules sur le promoteur . . . 51
5.3 Interactions combinatoires . . . 55
5.4 Dynamique lente du promoteur . . . 56
5.5 Progression de l'état chromatinien . . . 59

II Organisation nucléaire dynamique et expression stochastique des gènes 61
1 Modélisations spatiales intracellulaires 62
1.1 Modèles macroscopiques 64
1.2 Simulation à l'échelle atomique 65
1.3 Coarse graining : structures moléculaires simplifiées 70
1.4 Approches mésoscopiques 75
2 Simulateur mésoscopique 4D de phénomènes nucléaires 80
2.1 Modèle(s) : définition d'un formalisme général 81
2.2 Implémentation et logiciel de simulation 88
2.3 Illustrations à différentes échelles 90
3 Hypothèses et mécanismes à étudier 92
3.1 Structuration de l'expression par des mécanismes de diffusion . . . 92
3.2 Régions inter-chromosomiques et usines à transcription 96
3.3 Corps nucléaires : transitions de phase et non-linéarités 98
4 Conclusion 99

III Dynamique stochastique de la régulation transcriptionnelle 101
1 Modélisations du promoteur 101
1.1 Simulation *vs* dérivations analytiques 102
1.2 Modèles de l'expression stochastique 102
1.3 Modèles de la régulation (modèles thermodynamiques) 104
1.4 Modèles de la dynamique du promoteur 105
2 Modèle biophysique 106
2.1 Formulation générale 106
2.2 Structure et contenu de la matrice de transition 111
2.3 Exemple d'instanciation : un promoteur eucaryote 113
2.4 Reformulation énergétique 116
2.5 Un cadre simplifié : association/dissociation de FT seuls 121
3 Dérivations générales 122
3.1 Dynamique du promoteur 122
3.2 Décomposition en modes propres 125
3.3 Dérivation d'indicateurs : ChIP, FRAP, corrélations croisées, spectres de puissance et distributions de durées 130
3.4 Consommation énergétique et périodicités 141
3.5 Systèmes idéalisés 142
3.6 Transmission aux niveaux d'ARN et de protéines 147
4 Exploitation du modèle et résultats biologiques 157
4.1 Variabilité induite par une concentration élevée d'un FT 158
4.2 Contrôle transcriptionnel de l'hétérogénéité de l'expression 161
4.3 Régulation complexe de la dynamique du promoteur 164
4.4 Activité périodique des promoteurs eucaryotes : consommation énergétique et longueur effective du cycle 169
4.5 Quantification sur données expérimentales 173
4.6 Progression de l'état chromatinien et mécanismes d'interaction des molécules sur/avec la chromatine 177
5 Perspectives 187

5.1 Analyse commune de données de ChIP, de FRAP et de FRET . . . 187
5.2 Fluctuations extrinsèques et résonance stochastique 188
5.3 Oscillateurs stochastiques couplés et spécificité du code des histones 189

Conclusion et perspectives **193**

Bibliographie **197**

Annexes **221**
A Décomposition propre de $\tilde{\mathbf{M}}$ et réduction de l'expression du FRAP 221
B Paramétrage (chapitre III) 222
B.1 Paramètres des différents systèmes 222
B.2 Optimisation paramétrique pour une activité périodique 224

Table des figures

I.1 Séparation des composantes intrinsèques et extrinsèques de la stochasticité 20
I.2 Modèle standard de la stochasticité gène-intrinsèque 26
I.3 Modes de déplacement d'un facteur de transcription 33
I.4 Observation expérimentale de l'organisation nucléaire 41
I.5 Organisation spatiale de l'expression génique 45
I.6 Rôle de la chromatine dans la régulation transcriptionnelle 49
I.7 Molécules et complexes impliqués dans la régulation transcriptionnelle . . 51
I.8 Mesure expérimentale de la mobilité des molécules sur le promoteur . . . 54
I.9 Dynamique périodique lente des promoteurs eucaryotes 57

II.1 Échelles spatio-temporelles et formalismes de simulation intracellulaire . . 63
II.2 Énergie potentielle d'interaction entre deux atomes d'oxygène 65
II.3 Interprétation et pouvoir d'expression d'un potentiel d'interaction de paire 67
II.4 Hypothèse de sur-amortissement pour la diffusion d'une particule 69
II.5 Définition des macromolécules et de leurs interactions 83
II.6 Disruption et équilibre de phase d'un corps nucléaire minimal 91
II.7 Influence du profil d'affinité d'un FT sur sa dynamique 93
II.8 (anti-)Corrélation de l'activité de deux gènes par des mécanismes spatiaux 95

III.1 Pouvoir descriptif du modèle 107
III.2 Instanciation et simulation d'un système très simple 109
III.3 Contraintes structurelles sur la matrice $\mathbf{M}$ 112
III.4 Représentation d'un promoteur eucaryote 114
III.5 Interprétation énergétique des transitions d'état du promoteur 119
III.6 Exponentielle matricielle $e^{\tau\mathbf{M}}$ pour l'exemple du promoteur eucaryote . . 124
III.7 Décomposition modale pour l'exemple de promoteur eucaryote 126
III.8 Allure des fonctions $e^{\lambda_i\tau}$ et facteur de cohérence κ 129
III.9 Prédiction des cinétiques de ChIP 131
III.10 Corrélations croisées de la présence des facteurs 133
III.11 Spectres de puissance et auto-corrélations 135
III.12 Distributions de durée et courbes de FRAP 137
III.13 Spectre de $\mathbf{M}$ pour le cas du cycle orienté et homogène de transitions . . 144
III.14 Relation entre cohérence, dépendance énergétique et longueur d'un cycle . 145
III.15 Graphe d'état et spectre de $\mathbf{M}$ pour deux structures indépendantes . . . 147
III.16 Propagation la stochasticité du promoteur jusqu'au niveau d'expression . 150
III.17 Délais de synthèse distribués et durées de vie non-exponentielles 152

III.18 Variabilité induite par une forte concentration de FT : système minimal . 158
III.19 Domaine de validité : déviations par rapport au système minimal 160
III.20 Régulation des propriétés stationnaires de la variabilité de l'expression . . 162
III.21 Modulation de la dynamique du promoteur par la concentration d'un FT 164
III.22 Structurations parallèles du graphe d'état du promoteur et de son spectre 166
III.23 Désynchronisation au sein d'une population et cohérence des oscillations . 168
III.24 Longueur effective d'un motif : hétérogénéités et chemins alternatifs . . . 170
III.25 Compromis entre longueur effective et consommation énergétique 172
III.26 Cohérence des oscillations sur des données expérimentales 173
III.27 Forme des profils de ChIP et cohérence du système 176
III.28 ChIP et FRAP : deux mesures complémentaires de la dynamique d'un FT 179
III.29 Interaction des FT sur la chromatine : instanciation du *return-to-template* 181
III.30 Décroissance énergétique progressive et dynamique hors d'équilibre 183

Introduction

Le support physique sur lequel se base tout système biologique est constitué de macromolécules, des entités discrètes, en nombre fini et au comportement intrinsèquement probabiliste. À cette échelle, la matière est en effet fortement soumise à l'agitation thermique et le déplacement et la rencontre des molécules, base de toute réaction (bio)chimique au sein d'une cellule, présente alors un caractère fortement aléatoire. En particulier, lorsqu'interviennent des entités en faibles quantités (comme c'est par exemple le cas des gènes, souvent en une ou deux copies dans une cellule) la nature discrète et probabiliste des événements moléculaires devient importante.

La vie s'est donc développée et a évolué dans un contexte stochastique. Cette propriété a longtemps été perçue comme ne pouvant être que négative pour les systèmes biologiques, imposant alors à ceux-ci d'en minimiser les effets. Cependant, comme le montrent de plus en plus d'études expérimentales, il apparaît clairement aujourd'hui que l'expression génique – un processus central au fonctionnement de tout système vivant – présente dans de nombreux cas un caractère significativement stochastique (pour des reviews, voir par exemple (Raser et O'Shea, 2005; Kærn *et al.*, 2005; Maheshri et O'Shea, 2007; Raj et van Oudenaarden, 2008)). En effet, deux cellules identiques, placées dans les mêmes conditions, n'expriment pas leurs gènes de la même façon, provoquant alors des hétérogénéités au sein d'une population clonale de cellules. De plus, bien que la vision négative de cette stochasticité perdure encore aujourd'hui (Fraser *et al.*, 2004; Martinez Arias et Hayward, 2006), il a été montré dans de nombreux cas qu'elle pouvait au contraire constituer un atout, se révélant utile – voire nécessaire – pour divers processus biologiques. Son implication a par exemple été montrée dans des cadres aussi variés que la différentiation cellulaire (Rao *et al.*, 2002; Chang *et al.*, 2008; Çağatay *et al.*, 2009), le développement (Samoilov *et al.*, 2006; Boettiger et Levine, 2009), la prises de décision chez certains virus (Rao *et al.*, 2002; Weinberger *et al.*, 2005) ou encore la survie d'une population bactérienne face à un stress environnemental (Süel *et al.*, 2007; Mettetal et van Oudenaarden, 2007; Veening *et al.*, 2008).

Ainsi, loin d'avoir été systématiquement éliminé au cours de l'évolution, le caractère stochastique des réactions biochimiques, et en particulier celui de l'expression génique, a au contraire pu être exploité par les systèmes biologiques. Il apparaît aujourd'hui clairement comme une composante à part entière de leur fonctionnement, à intégrer à nos modes de pensée et à notre compréhension des principes organisationnels du vivant. Cependant, si cette part d'aléatoire dans l'expression des gènes peut constituer un avantage dans cer-

taines situations, il va de soi que, pour en tirer parti, la cellule doit être en mesure de la moduler et de la contrôler. Afin de comprendre, non seulement comment cette stochasticité est générée, mais également comment elle peut être contrainte et structurée sur différents plans, il est alors nécessaire de s'intéresser à ses origines moléculaires. Dans ce cadre, les approches de modélisation constituent une aide particulière car elles permettent d'identifier différentes sources potentielles de stochasticité, d'en déterminer certaines singularités (reconnaissables par exemple dans des données expérimentales) et d'en étudier les mécanismes de contrôle. Pour cela, bien que l'expression génique soit un mécanisme connu et étudié depuis de très nombreuses années, il doit être ici reconsidéré sous le jour nouveau du caractère discret et probabiliste des interactions entre molécules individuelles et des réactions biochimiques.

De façon remarquable, divers travaux expérimentaux récents indiquent également qu'il est nécessaire de prendre en compte le caractère discret et stochastique des interactions moléculaires pour comprendre de nombreux phénomènes biologiques. En effet, les techniques récentes d'imagerie du noyau cellulaire par microscopie à fluorescence indiquent que celui-ci ne constitue pas, comme on l'a longtemps pensé, un espace homogène où les molécules peuvent se rencontrer librement, mais plutôt un système fortement hétérogène et organisé au sein duquel la disposition spatiale des molécules joue un rôle essentiel. Il semble de plus que cette structuration de l'espace ne soit pas régie par une architecture globale prédéterminée, mais fasse appel à des principes d'auto-organisation reposant essentiellement sur les interactions locales entre molécules et leur comportement probabiliste. Ces résultats expérimentaux en appellent alors à la prise en compte explicite, dans les travaux de modélisation, de l'emplacement spatial des molécules et du caractère discret de la matière.

De plus, des méthodes similaires d'imagerie ont récemment permis d'étudier l'association/dissociation des molécules sur l'ADN, et en particulier, sur le promoteur des gènes. Ces études permettent de mieux comprendre la dynamique des complexes multiprotéiques impliqués dans de la régulation et l'activation transcriptionnelle des gènes. Il apparaît alors que ces complexes macromoléculaires ne sont pas, comme on en a longtemps eu l'image, des structures stables se formant et s'associant à l'ADN de façon durable, mais qu'il s'agit au contraire de structures très dynamiques au sein desquelles les molécules n'ont souvent des temps de résidence que de l'ordre de quelques secondes. Les complexes de régulation de la transcription peuvent donc avoir une composition très variable et leur formation apparaît comme un processus intrinsèquement stochastique, ne suivant pas une unique voie d'assemblage prédéterminée. De façon surprenante, on observe également, pour ces complexes, un comportement dynamique à une échelle temporelle beaucoup plus lente, de l'ordre de plusieurs dizaines de minutes. Il peut aussi bien s'agir de l'activation/inactivation stochastique du promoteur que d'une activité périodique au contraire très régulière avec des cinétiques caractéristiques d'occupation du promoteur par les différentes molécules. Les complexes de régulation transcriptionnelle apparaissent donc comme des structures très dynamiques et pouvant démontrer de propriétés multi-échelles singulières sur la base de l'association/dissociation stochastique et des interactions locales et combinatoires entre molécules.

Bien que les aspects spatiaux à l'échelle du noyau cellulaire et combinatoires à l'échelle du

promoteur d'un gène soient clairement impliqués dans le processus d'expression génique et se basent sur le comportement probabiliste de la matière, ils ne sont que très rarement pris en compte dans l'étude du caractère stochastique de l'expression génique. Ces trois thématiques de recherche sont en effet relativement récentes et ne sont encore que peu associées les unes aux autres. Nous nous proposons, dans le cadre de cette thèse, d'étudier les relations entre la dynamique spatio-temporelle des complexes macromoléculaires à ces diverses échelles et le caractère stochastique de l'expression génique.

Le travail réalisé ici s'inscrit dans un contexte pluridisciplinaire et dans une démarche de *biologie des systèmes* (Kitano, 2002) faisant intervenir à la fois des travaux de modélisation et des expériences humides. Même si nous avons, dans ce cadre, participé à la définition, la mise au point et l'analyse d'expériences humides, nos travaux personnels restent essentiellement théoriques et nous ont de plus amené à explorer des voies différentes du cadre du projet initial[1]. Nous ne présentons donc ici que les aspects théoriques de nos travaux. Ce mémoire de thèse est organisé en trois chapitres.

Dans le premier chapitre, nous décrivons l'état des connaissances actuelles sur le champ de recherche relativement récent qu'est l'étude de l'expression stochastique des gènes. Nous nous concentrons spécifiquement sur son origine moléculaire (même si ses conséquences et son rôle biologique sont également des sujets de recherche très actifs) et mettons un accent particulier sur les résultats de modélisation dans ce domaine. Partant alors du constat que les phénomènes d'organisation au sein du noyau cellulaire et que la dynamique multi-échelle singulière des complexes de régulation de la transcription ne sont que très peu pris en compte dans ces études, nous décrivons alors la littérature, essentiellement expérimentale, relative à ces deux thématiques de recherche. Les deux chapitres suivants présentent alors nos travaux de modélisation sur chacun de ces deux aspects.

Le chapitre II décrit un travail en cours et relativement préliminaire de conception d'un modèle et d'un logiciel de simulation spatiale, dédiés à l'étude des phénomènes d'organisation au sein du noyau cellulaire en relation avec le processus d'expression génique. Ce modèle emprunte en particulier différentes techniques à des formalismes de simulation intracellulaire à différentes échelles spatio-temporelles. Bien qu'à l'issue de ce chapitre, nous ayons identifié certaines limites potentielles de l'approche en termes de calculabilité et de pouvoir descriptif, ce travail nous a toutefois permis de développer une certaine réflexion sur le sujet, nous conduisant alors à formuler plusieurs hypothèses quant aux influences du contexte spatial du processus d'expression génique sur ses propriétés stochastiques. Les limitations de ce modèle de simulation nous ont cependant amenés à adopter une approche sensiblement différente pour étudier de façon plus précise la dynamique des complexes de régulation transcriptionnelle.

Ainsi, dans le chapitre III, nous proposons un modèle à l'intersection de différents champs de recherche théoriques et expérimentaux, pour étudier la dynamique stochastique spontanée des complexes multiprotéiques sur un promoteur. Ce modèle se base sur la description des événements d'association/dissociation de molécules et des modifications conformationnelles et covalentes de la chromatine, prenant en compte les influences mutuelles combinatoires entre tous ces facteurs (coopération/compétition, réactions enzymatiques et de

[1]Pour une description de ce projet pluridisciplinaire, le lecteur pourra se référer à (Coulon *et al.*, 2009).

remodelage). Les dérivations, essentiellement analytiques, nous permettent en particulier de décrire dans un cadre commun le comportement d'un promoteur selon des mesures employées dans les différents domaines expérimentaux évoqués précédemment : des mesures de la stochasticité de l'expression (distribution et spectres de puissance du niveau d'expression), des mesures de la dynamique rapide des complexes sur le promoteur (FRAP, FRET) et des mesures de son comportement lent (cinétiques de ChIP). Nous montrons alors, d'une part, que la dynamique du promoteur permet un contrôle fin et complexe de différents aspects de l'expression stochastique d'un gène et, d'autre part, que des mesures expérimentales peuvent révéler différentes caractéristiques de la structure du système moléculaire sous-jacent telles que sa taille, sa consommation énergétique, ou encore le mode d'interaction des molécules entre elles et avec la chromatine.

L'ensemble de ces travaux suggèrent alors que le gène, une entité souvent représentée de façon simple et déterministe, peut au contraire témoigner de comportements dynamiques et stochastiques très singuliers sur lesquels peut s'exercer un contrôle complexe.

Chapitre I

Origines moléculaires de la stochasticité de l'expression génique

Deux cellules génétiquement identiques – par exemple, issues d'une population clonale – et placées dans les mêmes conditions peuvent exprimer une protéine donnée à des niveaux très différents. Cette part d'aléatoire dans le processus d'expression génique, bien que mise en évidence depuis de nombreuses années (Kupiec, 1997, et références associées), fait l'objet d'intenses investigations depuis les travaux d'Elowitz *et al.* (2002) chez les procaryotes et de Levsky *et al.* (2002) et Blake *et al.* (2003) chez les eucaryotes. Ces nombreux travaux ont permis de mieux caractériser la stochasticité de l'expression génique chez divers organismes et ce dans des conditions très variées.

Les termes employés pour désigner ce phénomène sont nombreux. On parle aussi bien de stochasticité, de bruit, de variabilité, de fluctuations de l'expression génique, que d'hétérogénéité temporelle ou au sein d'une population, ou encore du caractère probabiliste ou aléatoire de l'expression. Ces termes ne sont cependant pas neutres et portent différentes connotations. En particulier, l'utilisation du terme *bruit* – encore souvent employé aujourd'hui – est assez révélatrice de la vision négative longtemps associée à cette stochasticité. Selon cette vision il ne s'agirait que d'un aspect nuisible de l'expression génique ne pouvant qu'entraver le bon déroulement d'un programme génétique déterministe et forçant la cellule à développer des stratégies pour le combattre. De nombreuses études ayant aujourd'hui démontré l'importance voire la nécessité de ce phénomène dans des contextes très différents, il apparaît maintenant clairement comme une composante essentielle du fonctionnement des systèmes biologiques, comme un mécanisme pertinent et fonctionnel, sujet à de multiples régulations.

De fait, si ce comportement probabiliste peut constituer un atout pour la cellule, il

convient alors pour celle-ci d'être en mesure de le contrôler – de le filtrer, de l'amplifier, de le canaliser, de le structurer – afin d'en tirer parti en fonction du contexte. On assiste alors à l'émergence d'un nouveau champ de recherche dont l'un des objectifs principaux est d'identifier et de quantifier les différentes sources de stochasticité – de découvrir quelle en est ou quelles en sont les causes primaires (Paulsson, 2005b) – ainsi que les mécanismes qui peuvent l'influencer. Dans ce cadre, il est souvent fait appel à la modélisation afin d'identifier les sources potentielles de stochasticité, de définir des expériences permettant de les mettre en évidence et d'en interpréter les résultats, plaçant la démarche de *biologie des systèmes* (Kitano, 2002) au cœur de ce champ de recherche.

Dans ce premier chapitre, nous posons tout d'abord les principes de base fréquemment employés dans la littérature pour la mise en évidence expérimentale de la stochasticité de l'expression génique et l'identification de ses diverses origines par l'intermédiaire de la modélisation. Nous décrivons ensuite (section 2) comment le processus d'expression génique peut comporter diverses sources de stochasticité et comment leurs mécanismes d'origine leur confèrent des caractéristiques spécifiques. Nous verrons alors, dans la section 3, comment des fluctuations peuvent être transmises et modifiées par un gène ou par un motif de gènes et structurées à un niveau plus global par le réseau. Cette revue de l'état de l'art nous conduira alors au constat essentiel que deux niveaux d'organisation sont généralement considérés de façon particulièrement simplifiée dans les études de modélisation. Il s'agit d'une part des aspects spatiaux du processus d'expression génique, en relation avec le caractère très hétérogène du noyau des cellules eucaryotes, et d'autre part de la régulation transcriptionnelle au niveau du promoteur d'un gène, processus complexe présentant des dynamiques temporelles multi-échelles très singulières. Hors du champ de la littérature consacré à la stochasticité de l'expression génique, ces deux problématiques constituent pourtant des champs de recherche très actifs et il existe une littérature abondante décrivant de nombreux phénomènes ayant de probables implications majeures sur le caractère stochastique de l'expression génique. Nous exposons alors dans les sections 4 et 5 de ce chapitre les connaissances relatives à ces deux champs de recherche en tentant de mettre à jours leurs diverses relations avec l'expression stochastique des gènes. Nos contributions en rapport avec ces deux domaines seront alors exposées respectivement aux chapitres II et III.

1 Séparation, quantification et identification des sources de stochasticité

1.1 Mesures expérimentales

Les techniques expérimentales employées pour identifier et séparer les différentes composantes de la stochasticité de l'expression génique sont variées mais se basent presque systématiquement sur des mesures sur cellules uniques. Beaucoup font en particulier usage d'un ou de plusieurs gènes rapporteurs codant pour des protéines fluorescentes, le niveau de fluorescence observé sur une cellule témoignant alors du niveau d'expression de la protéine en question. Deux techniques essentielles, maintenant couramment employées, sont

la cytométrie de flux – permettant de mesurer à un instant donné la fluorescence de chaque cellule d'une large population – et la vidéo-microscopie – permettant de mesurer l'évolution au cours du temps de la fluorescence de quelques cellules. À partir de telles données, il est possible de dériver plusieurs indicateurs fournissant chacun différents points de vue sur le système étudié. En combinant les mesures obtenues sur les différents rapporteurs et/ou dans plusieurs conditions, il est alors possible de distinguer plusieurs composantes de la stochasticité, d'en quantifier la contribution et d'y reconnaître des signatures spécifiques permettant de discriminer parmi plusieurs hypothèses quant à leur origine.

1.2 Stochasticités intrinsèques *vs* extrinsèques

Depuis les travaux expérimentaux d'Elowitz *et al.* (2002), mettant en application les travaux théoriques de Swain *et al.* (2002), il est souvent fait une distinction entre stochasticité *intrinsèque* et stochasticité *extrinsèque*. Comme le souligne Paulsson (2005a), ces deux termes n'ont cependant de sens que s'ils se rapportent à un système préalablement défini. La part intrinsèque de stochasticité est alors celle dont l'origine se trouve au sein du système considéré, tandis que la part extrinsèque correspond aux fluctuations extérieures au système que celui-ci ne fait que transmettre. Pourtant, bien souvent, l'utilisation de ces termes souffre de l'absence d'une identification claire du système en question, ce qui peut parfois conduire à certaines confusions. En effet, le système n'est généralement défini qu'implicitement par la description de l'expérience réalisée.

Afin de séparer ces deux composantes, l'approche classique, reprise aujourd'hui dans de nombreuses études (Raser et O'Shea, 2004; Colman-Lerner *et al.*, 2005; Rosenfeld *et al.*, 2005; Raj *et al.*, 2006; Sigal *et al.*, 2006; Neildez-Nguyen *et al.*, 2007; Raser et O'Shea, 2005; Maheshri et O'Shea, 2007; Kærn *et al.*, 2005; Raj et van Oudenaarden, 2008), consiste à comparer les niveaux d'expression de deux protéines fluorescentes contrôlées par deux promoteurs identiques. Les sources de stochasticité qui affectent de la même façon les deux gènes rapporteurs et provoquent des fluctuations corrélées entre les niveaux de fluorescence au sein de chaque cellule sont extérieures au système « dupliqué » – ici, le gène[1] (figure I.1a1). Il peut s'agir par exemple de fluctuations de la concentration d'un facteur de transcription (FT), de la quantité d'ARN polymérase, ... À l'inverse, les sources de stochasticité provenant du système dupliqué (état actif/inactif du promoteur, événement de synthèse/dégradation d'ARNm ou de protéines, ...) sont indépendantes d'un système à l'autre et produisent donc des fluctuations décorrélées entre les deux niveaux de fluorescence (figure I.1a2). Dans le cas général, il est possible de montrer (Swain *et al.*, 2002) que la variance normalisée du niveau d'expression est la *somme* d'une contribution intrinsèque et d'un contribution extrinsèque[2] $\eta^2_{\text{tot}} = \eta^2_{\text{ext}} + \eta^2_{\text{int}}$ qui peuvent être directement déterminées à partir de la corrélation entre les deux niveaux de fluorescence mesurés, par exemple, par cytométrie de flux (figure I.1a3).

[1] Plus précisément, il s'agit du promoteur du gène, de l'ensemble de ses transcrits et des protéines produites à partir de ces derniers.

[2] η^2_{ext} correspond au η^2_{tot} que l'on obtiendrait si le système considéré était déterministe et ne faisait que transmettre une stochasticité reçue en amont. η^2_{int} correspond en revance à la moyenne, pour toutes les trajectoires possibles des facteurs extrinsèques, du η^2_{tot} que l'on obtiendrait pour une trajectoire donnée.

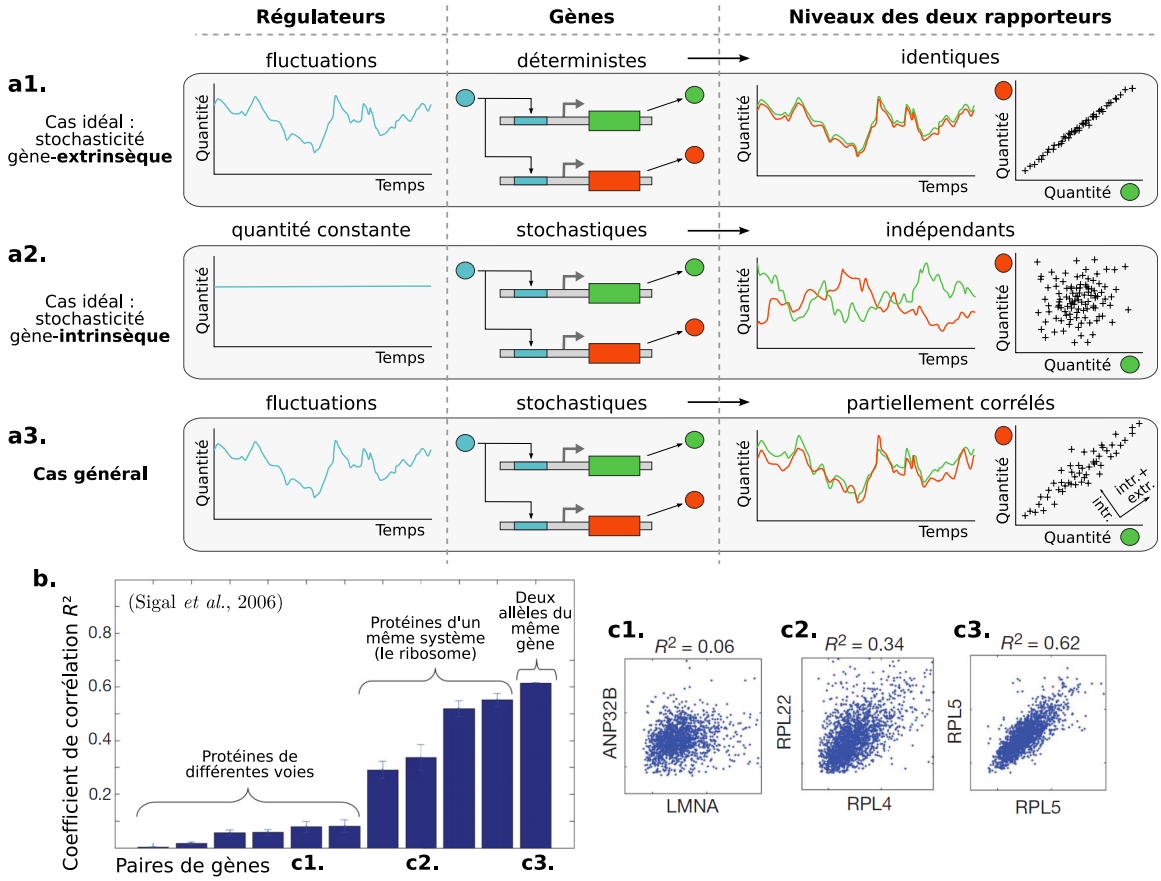


Fig. I.1: Séparation des composantes intrinsèques et extrinsèques de la stochasticité. (a1) Dans le cas idéal où l'expression d'un gène n'est pas probabiliste, les fluctuations dans la quantité de protéines produites proviennent nécessairement de fluctuations dans la concentration des molécules contrôlant l'expression du gène (régulateurs, polymérase, ...) : elles sont donc gène-extrinsèques. Deux copies du gène auront alors une expression totalement identique. (a2) Dans le cas inverse où l'expression d'un gène est probabiliste mais où tous ses régulateurs sont en concentrations constantes, les fluctuations du niveau d'expression proviennent exclusivement des événements stochastiques propres au processus d'expression du gène : elles sont gène-intrinsèques. Deux copies différentes du même gène auront alors une expression totalement indépendante. (a3) Dans le cas général, la corrélation entre les niveaux d'expression de deux gènes identiques permet de quantifier la contribution respective de ces deux composantes (Swain *et al.*, 2002) et peut par exemple être obtenue simplement à partir de données de cytométrie de flux. Notons par ailleurs que la définition exacte de ces deux composantes[2] est en réalité moins restrictive que les deux cas idéaux a1 et a2. (b) L'utilisation de paires de gènes différents rend intrinsèques toutes fluctuations n'affectant pas à la fois les deux gènes et permet donc d'avoir une description de la propagation des fluctuations au sein du réseau. (c1-3) En effet, plus deux gènes sont fonctionnellement proches, plus on observe que leur expression est corrélée. (b, c1-3 : d'après Sigal *et al.* (2006)).

L'utilisation de constructions différentes permet de raffiner la séparation des différentes sources. Ainsi, en comparant le résultat de l'expérience décrite précédemment avec le cas où les deux promoteur sont différents et appartiennent à des voies de régulation distinctes, il est possible de séparer les fluctuations de facteurs globaux (quantité de polymérase, FT

généraux, ...) de celles de facteurs spécifiques à chaque promoteur (Raser et O'Shea, 2004; Colman-Lerner *et al.*, 2005; Ramanathan et Swain, 2005; Kærn *et al.*, 2005; Sigal *et al.*, 2006; Maheshri et O'Shea, 2007). La propagation de la stochasticité au sein du réseau de gènes peut alors être hiérarchisée en mesurant la corrélation entre paires de gènes plus ou moins proches (figures I.1b et I.1c1-3) (Sigal *et al.*, 2006; Pedraza et van Oudenaarden, 2005). De même, placer les deux promoteurs identiques l'un après l'autre dans le génome rend extrinsèque les fluctuations locales (ouverture/fermeture de la chromatine, position spatiale du locus chromosomique ...) et permet de les quantifier (Raj *et al.*, 2006).

Il est à noter que les composantes intrinsèques et extrinsèques peuvent être séparées en faisant appel à d'autres techniques que celle du double rapporteur fluorescent. En particulier, l'intégration en un nombre variable de copies d'un seul rapporteur fluorescent n'influence que la mesure de la part intrinsèque de la stochasticité (Becskei *et al.*, 2005; Volfson *et al.*, 2006).

Dans certains cas, cette distinction peut cependant rester ambiguë et les sources de variabilité ou les mécanismes l'influençant ne peuvent pas toujours être placés dans l'une ou dans l'autre de ces catégories. En particulier, comme nous le verrons au chapitre II, l'identification claire du système peut devenir particulièrement délicate dès lors que l'on prend en compte les aspects spatiaux. De plus, lorsqu'on considère des rétrocontrôles, la propriété essentielle d'additivité évoquée précédemment n'est plus valable (Tănase-Nicola *et al.*, 2006).

1.3 Signatures spécifiques des différentes sources

La modélisation permet de prédire des caractéristiques spécifiques aux divers mécanismes générant ou influençant la variabilité de l'expression génique. Ces signatures peuvent ensuite être utilisées pour identifier et quantifier une ou plusieurs sources à partir de données expérimentales. Un indicateur très couramment employé est la relation entre la variabilité de la quantité d'une certaine molécule et son abondance (sa quantité moyenne). En effet, dans les modèles simples d'expression génique, chaque source possède sa signature propre dans la relation entre la variance normalisée $\sigma_P^2/\langle P\rangle^2$ et la moyenne $\langle P\rangle$ du nombre de protéines. De très nombreux auteurs (Elowitz *et al.*, 2002; Ozbudak *et al.*, 2002; Blake *et al.*, 2003; Raser et O'Shea, 2004; Golding *et al.*, 2005; Pedraza et van Oudenaarden, 2005; Becskei *et al.*, 2005; Bar-Even *et al.*, 2006; Newman *et al.*, 2006; Volfson *et al.*, 2006; Mettetal *et al.*, 2006; Murphy *et al.*, 2007; Sánchez et Kondev, 2008) utilisent cette relation (mesurée pour différentes concentrations de l'inducteur du gène considéré, à différents instants pendant l'induction, pour différentes mutations ponctuelles affectant l'efficacité des différentes étapes des l'expression, ...) afin de discriminer les différentes sources de stochasticité.

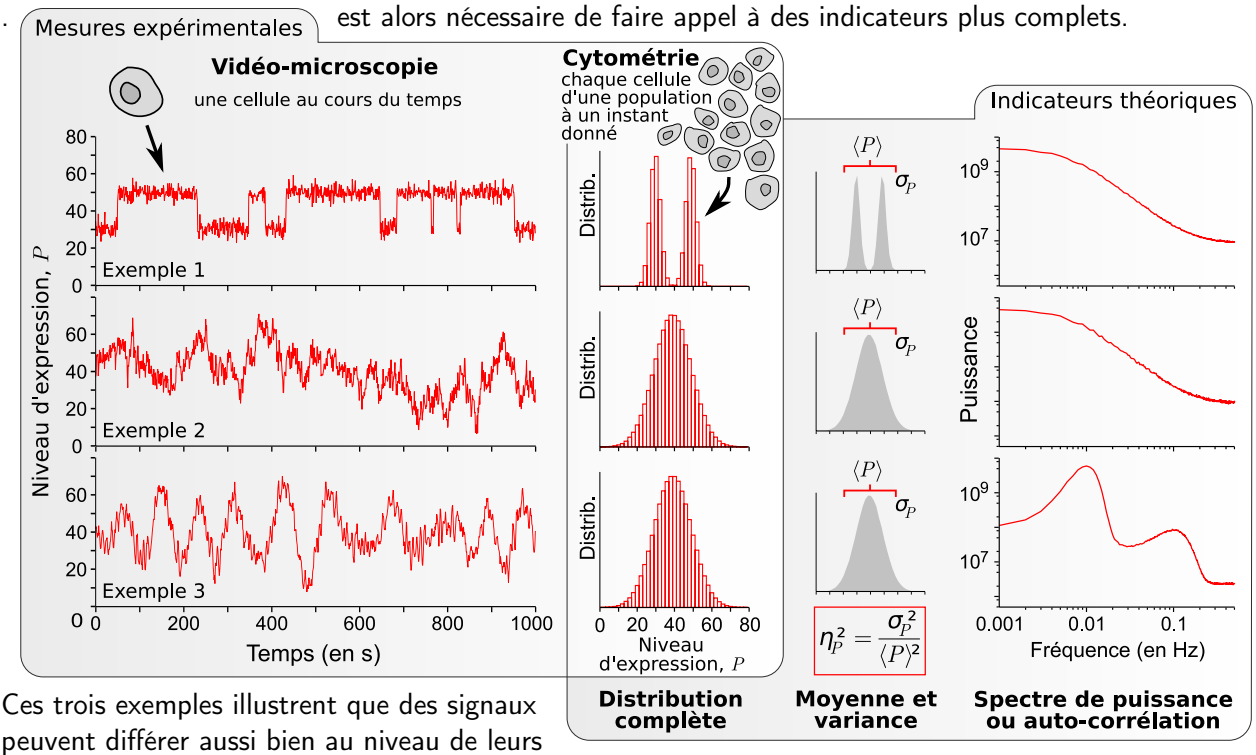
Le contenu fréquentiel d'un niveau d'expression mesuré au cours du temps par des techniques de vidéo-microscopie peut également comporter des singularités exploitables. En particulier, un gène autorégulé ou un motif particulier de quelques gènes produisent une signature reconnaissable dans des mesures spectrales telles que le spectre de puissance, l'autocorrélation d'un niveau d'expression ou la corrélation croisée entre deux niveaux

Box 1 : Indicateurs de la stochasticité de l'expression génique

La stochasticité de l'expression d'un gène peut être caractérisée par différents indicateurs pouvant aussi bien être obtenus à partir de mesures expérimentales que prédits à partir de modèles théoriques.

- La *variance normalisée*, notée η_P^2, est obtenue en divisant la variance σ_P^2 de la quantité P d'une certaine molécule par le carré de sa moyenne $\langle P \rangle^2$. Elle correspond donc au carré du *coefficient de variation* η_P. Ces mesures, η_P et η_P^2, parfois appelées simplement *noise*, sont sans dimension et décrivent de façon équivalente à quel point les fluctuations sont importantes relativement à la moyenne. Le *facteur de Fano* $\sigma_P^2/\langle P \rangle$, également appelé *noise strength*, permet de quantifier l'écart d'une distribution à une loi de Poisson mais peut parfois être source de confusions (cf section 2.3).

Ces indicateurs synthétiques (une seule valeur) ne donnent cependant qu'une vision très partielle du système observé. De fait, ils ne mettent pas en évidence les spécificités des trois exemple présentés ci-dessous. Il est alors nécessaire de faire appel à des indicateurs plus complets.



Ces trois exemples illustrent que des signaux peuvent différer aussi bien au niveau de leurs valeurs (bistabilité dans l'exemple 1) que d'un point de vue temporel (périodicités dans l'exemple 3).

- La *distribution* complète des valeurs d'un niveau d'expression – que mesurent directement des données de cytométrie – caractérise bien plus précisément la structure de la variabilité (nombre/taille/position/forme des différents modes). Elle met ainsi en évidence la différence entre les exemples 1 et 2 ci-dessus, montrant le caractère bistable du premier. L'ensemble des moments de la distribution, bien que théoriquement équivalent, ne permet pas une interprétation aussi explicite.

		Moyenne et variance $\langle P \rangle$ σ_P	Distribution complète	Spectre de puissance
Peut être déduit de données de	Cytométrie	✓	✓	✗
	Vidéo-micro.	✓*	✓*	✓
Informations dynamiques		✗	✗	✓
Ordre de la statistique		2	∞	2

* sous une condition d'ergodicité (cf section 3.4)

- Le *spectre de puissance* décrit le contenu fréquentiel d'un signal de vidéo-microscopie (il s'agit du carré du module de la transformée de Fourier), mettant ici en évidence la différence entre les exemples 2 et 3. Un pic plus ou moins fin dans le spectre révèle une périodicité plus ou moins régulière du signal. L'*autocorrélation* d'un niveau d'expression (corrélation entre deux valeurs du signal en fonction du délai qui les séparent) est un indicateur strictement équivalent.

Tous ces indicateurs peuvent être généralisés au cas où la quantité de plusieurs espèces moléculaires sont mesurées simultanément (deux protéines différentes, une protéine et son ARN, ...). On considérera alors des *covariances*, des *distributions jointes* et des *corrélations croisées*.

d'expression (cf Box 1) (Simpson *et al.*, 2003, 2004; Cox *et al.*, 2006; Warren *et al.*, 2006; Lestas *et al.*, 2008). Sur des données expérimentales, la mesure de ce type d'information peut alors révéler l'architecture du réseau de gènes sous-jacent (Austin *et al.*, 2006; Dunlop *et al.*, 2008; Cox *et al.*, 2008).

De même, la distribution complète d'un niveau d'expression au sein d'une population – par opposition à sa description partielle par la moyenne et la variance (cf Box 1) – témoigne de certaines régularités (par exemple, la multimodalité) dues à la dynamique propre d'un gène (Kepler et Elston, 2001; Hornos *et al.*, 2005; Innocentini et Hornos, 2007; Shahrezaei et Swain, 2008a) et/ou à la transmission de fluctuations extrinsèques (Thattai et van Oudenaarden, 2001; Walczak *et al.*, 2009). L'allure de cette distribution pour différents niveaux d'induction et en réponse à des perturbations peut être utilisée, par exemple, pour l'interprétation de données de cytométrie (Blake *et al.*, 2003; Mettetal *et al.*, 2006; Volfson *et al.*, 2006).

Ainsi, la modélisation, parce qu'elle permet une interprétation quantitative des mesures expérimentales, mais aussi parce qu'elle peut être un guide pour la définition de nouvelles expériences, constitue un outil indispensable à la compréhension de la stochasticité de l'expression génique, jouant de fait un rôle central dans la majorité des études. Dans la section suivante, nous verrons plus précisément quelles sont les principales sources de cette stochasticité, quelles en sont les signatures caractéristiques et comment elles ont pu être modélisées.

2 Sources de stochasticité gène-intrinsèques

Le caractère stochastique de tout processus au sein d'une cellule vivante tient au fait que les événements moléculaires (synthèse, dégradation, association, dissociation, ...) sont discrets, probabilistes et ont lieu en nombre fini. De façon générale, la nature aléatoire de ces événements n'est plus négligeable dès lors que certains d'entre eux sont rares et/ou affectent de petites quantités. C'est pourquoi la stochasticité est souvent associée aux faibles concentrations, même si cette vision est assez simpliste car la rareté des événements n'est pas nécessairement due à l'existence de faibles concentrations. En particulier, nous montrerons au chapitre III que l'on peut obtenir la relation inverse dans de nombreux cas.

Pour étudier les différentes sources de stochasticité, il peut être fait appel à différents modèles détaillant plus ou moins les nombreuses étapes de l'expression génique et basés sur divers formalismes de modélisation (intégrant ou non l'espace, décrivant les molécules de façon discrète ou non, ...) (Gillespie, 1977; Turner *et al.*, 2004; Takahashi *et al.*, 2005; Chabrier-Rivier *et al.*, 2005; Kuttler, 2006; Wilkinson, 2009). Ceux-ci peuvent ensuite aussi bien être étudiés à l'aide de la simulation qu'en faisant appel à des méthodes de résolution analytiques et/ou computationnelles. Chaque choix impose à la modélisation certaines simplifications de la réalité et il convient de garder à l'esprit qu'ils doivent être dictés en premier lieu par leur pertinence par rapport à la représentation de l'objet d'étude et non par rapport aux techniques d'analyse qu'ils permettent d'appliquer. L'utilisation de techniques de résolution analytiques ou computationnelles sur des modèles génériques

permet de comparer et d'exprimer en termes communs un grand nombre de sources de stochasticité. En particulier, un certain nombre d'auteurs (Paulsson, 2005a; Warren *et al.*, 2006; Cox *et al.*, 2006; Lestas *et al.*, 2008) proposent des dérivations de divers indicateurs (variance normalisée, spectres de puissance, ...) pour un système arbitraire de réactions biochimiques. Instanciées sur des systèmes particuliers simples, ces dérivations donnent lieu à des expressions dont chaque terme peut être interprété d'un point de vue biophysique.

2.1 Modélisation de l'expression génique

Comme le décrit le principe de séparation intrinsèque/extrinsèque, la variabilité observée dans le niveau d'expression d'un gène donné provient en partie des événements stochastiques liés à son expression et en partie de la propagation de fluctuations de la concentration de ses régulateurs. Mais ces dernières sont elle-même dues au caractère stochastique de l'expression d'autres gènes. Il est donc avant tout nécessaire de comprendre comment la stochasticité peut être générée avant de comprendre comment elle peut être transmise et modifiée. Nous commençons donc ici par décrire les différentes sources gène-intrinsèques et considérerons, dans la section 3, les propriétés de propagation et de modulation des fluctuations gène-extrinsèques par un gène ou au sein d'un réseau.

La figure I.2a présente un modèle de l'expression génique utilisé dans de nombreuses études[3] (Rigney et Schieve, 1977; Ko, 1991; Peccoud et Ycart, 1995; Kepler et Elston, 2001; Paulsson, 2004, 2005a; Raser et O'Shea, 2004; Bar-Even *et al.*, 2006; Kærn *et al.*, 2005; Kaufmann et van Oudenaarden, 2007; Shahrezaei et Swain, 2008b,a; Simpson *et al.*, 2004; Cox *et al.*, 2006; Hornos *et al.*, 2005; Innocentini et Hornos, 2007; Lipniacki *et al.*, 2006; Tkačik *et al.*, 2008; Paszek, 2007). Bien qu'impliquant plusieurs simplifications (dont nous verrons l'impact ultérieurement), ce modèle permet de rendre compte de nombreuses sources de stochasticité gène-intrinsèques. Il correspond au système de réactions biochimiques suivant :

$$\mathcal{G}_{\text{off}} \underset{k^-}{\overset{k^+}{\rightleftarrows}} \mathcal{G}_{\text{on}} \tag{I.1}$$

$$\mathcal{G}_{\text{on}} \xrightarrow{\rho} \mathcal{G}_{\text{on}} + \mathcal{R} \tag{I.2}$$

$$\mathcal{R} \xrightarrow{\gamma} \varnothing \tag{I.3}$$

$$\mathcal{R} \xrightarrow{\tilde{\rho}} \mathcal{R} + \mathcal{P} \tag{I.4}$$

$$\mathcal{P} \xrightarrow{\tilde{\gamma}} \varnothing \tag{I.5}$$

Les protéines $\mathcal{P}$ sont synthétisées à partir des molécules d'ARNm $\mathcal{R}$ (I.4), elles-mêmes synthétisées à partir du gène dans un état actif $\mathcal{G}_{\text{on}}$ (I.2). Le promoteur du gène passe spontanément d'un état actif $\mathcal{G}_{\text{on}}$ à un état inactif $\mathcal{G}_{\text{off}}$, représentant par exemple la présence/absence d'un FT ou l'état ouvert/fermé de la chromatine (I.1). Les protéines $\mathcal{P}$ et les

[3] À quelques différences près : par exemple certains auteurs ne considèrent qu'une seule étape pour la transcription et la traduction, d'autres considèrent de l'autorégulation, d'autres encore considèrent plusieurs copies du gène. Mais toutes ces études considèrent un promoteur à deux états et des processus simples de synthèse/dégradation comme nous les décrivons ici.

messagers $\mathcal{R}$ se dégradent spontanément (I.3 et I.5). La figure I.2b présente un exemple de la dynamique résultant de la simulation de ce système pour un jeu de paramètres donné. À partir de ce modèle simple, il est déjà possible de mettre en évidence différentes sources de stochasticité correspondant à des comportements très différents. Nous en présentons les principaux mécanismes dans les sections suivantes.

2.2 Synthèse/dégradation de protéines et des transcrits

Considérons uniquement les réactions de synthèse (I.4) et de dégradation (I.5) des protéines pour un nombre constant R de messagers $\mathcal{R}$. Le nombre P des protéines $\mathcal{P}$ fluctue alors autour d'une valeur moyenne $\langle P\rangle = \tilde{\rho}R/\tilde{\gamma}$ proportionnelle à la quantité d'ARNm. Ces fluctuations sont d'autant plus importantes, relativement à la moyenne, que le niveau moyen $\langle P\rangle$ est bas (figures I.2c1 et I.2c2). On parle alors d'effets de *taille finie* car chaque événement stochastique de synthèse ou de dégradation provoque un changement discret d'autant plus « visible » que la quantité de protéines est faible. En effet, plus le nombre de molécules est important plus on se rapproche du comportement déterministe décrit par la loi d'action de masse. Dans le cas d'un système où le taux de synthèse $\tilde{\rho}R$ est constant, la distribution du nombre de protéines (ce que l'on observerait sur une population de cellules par cytométrie de flux par exemple) suit alors une loi de Poisson (Kaufmann et van Oudenaarden, 2007). Une caractéristique importante de cette distribution est que sa variance est égale à sa moyenne $\sigma_P^2 = \langle P\rangle$. Ainsi, pour un tel système, la variance normalisée par le carré de la moyenne vaut

$$\frac{\sigma_P^2}{\langle P\rangle^2} = \frac{1}{\langle P\rangle} \tag{I.6}$$

et diminue donc avec $\langle P\rangle$. Une variance normalisée différente de cette valeur témoigne donc que la synthèse des protéines ne s'effectue pas à un taux constant au cours du temps.

Si l'on considère maintenant également les réactions de synthèse (I.2) et de dégradation (I.3) d'ARNm pour un gène actif en permanence, le nombre R de transcrits $\mathcal{R}$ suit exactement le même type de processus. Les fluctuations de R autour de sa moyenne $\langle R\rangle = \rho/\gamma$ sont d'autant plus importantes que celle-ci est faible $\sigma_R^2/\langle R\rangle^2 = 1/\langle R\rangle$ et se transmettent en partie au niveau protéique (figure I.2c3) (Paulsson, 2005a) :

$$\frac{\sigma_P^2}{\langle P\rangle^2} = \frac{1}{\langle P\rangle} + \frac{1}{\langle R\rangle}\frac{\tilde{\gamma}}{\gamma+\tilde{\gamma}} \tag{I.7}$$

Le terme supplémentaire provient de l'inhomogénéité du taux de protéosynthèse résultant des fluctuations de la quantité d'ARN et varie également comme $1/\langle P\rangle$ car $\langle R\rangle = \langle P\rangle\tilde{\gamma}/\tilde{\rho}$.

Ainsi, un comportement en $1/\langle P\rangle$ de la variance normalisée témoigne d'un bruit Poisson de synthèse/dégradation. Celui-ci provient essentiellement du faible nombre de protéines si le coefficient de proportionnalité de cette relation est proche de 1 et du faible nombre de transcrits si ce coefficient est très supérieur à 1. Cette signature couramment utilisée a permis de montrer qu'une part significative de la stochasticité de l'expression génique peut provenir de ces effets de taille finie dus au faible nombre de transcrits ou de protéines, aussi bien dans des systèmes procaryotes (Elowitz *et al.*, 2002; Ozbudak *et al.*, 2002)

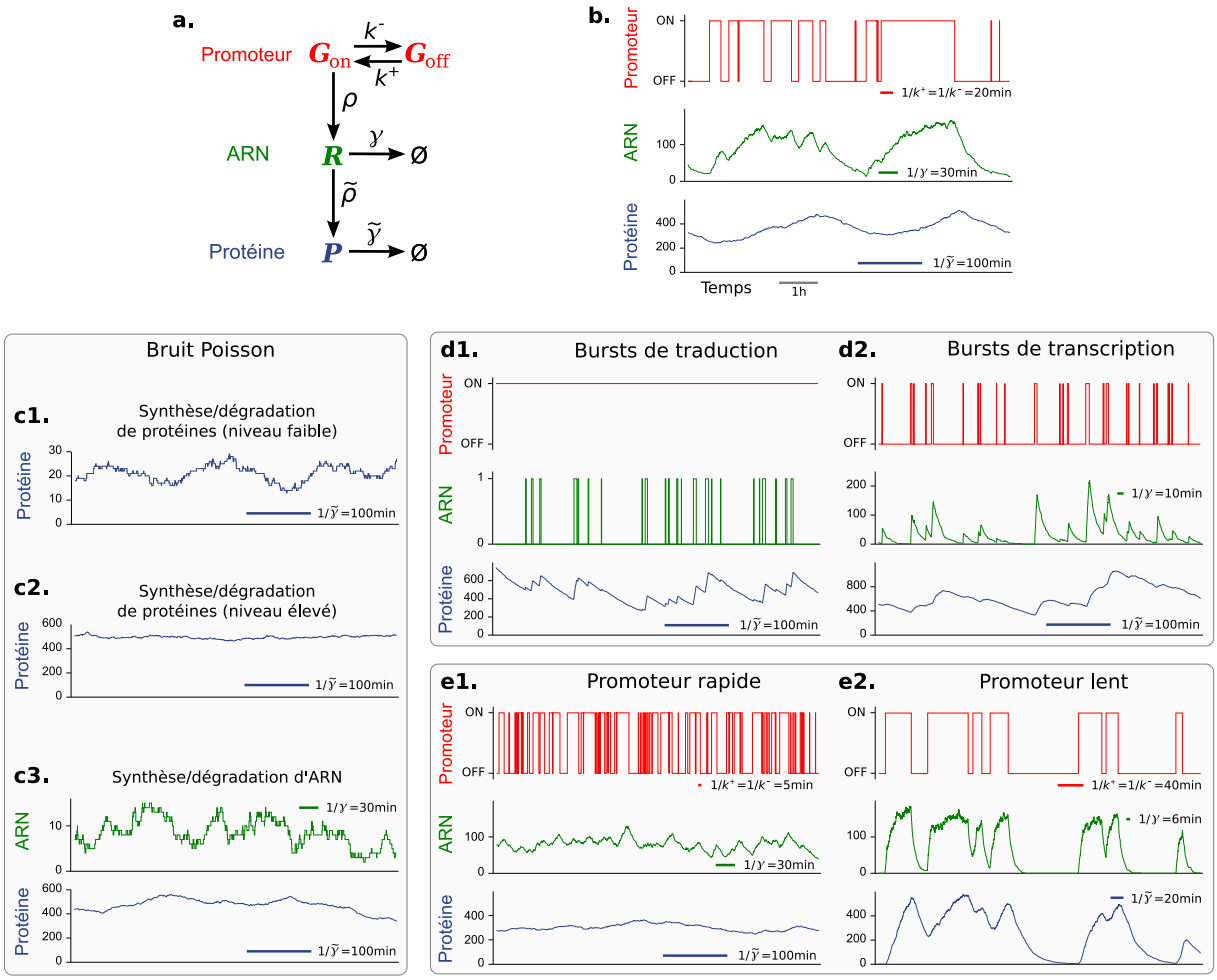


Fig. I.2: Modèle standard de la stochasticité gène-intrinsèque. Ce modèle très simple (a) et couramment employé dans la littérature permet de comprendre un grand nombre de sources de stochasticité gène-intrinsèques. La simulation de ce système par un algorithme de Gillespie (1977) donne un exemple de sa dynamique générale (b) et permet de mettre en évidence plusieurs phénomènes en fonction des valeurs des paramètres cinétiques. (c) Le bruit Poisson correspond aux fluctuations spontanées d'une quantité de molécules soumises à un processus de synthèse/dégradation stochastique. (c1,c2) Ces fluctuations sont d'autant plus importantes relativement au niveau moyen que ce dernier est faible et (c3) peuvent être transmises d'un niveau à l'autre. La durée de vie moyenne des molécules (barres horizontales) détermine la durée caractéristique des fluctuations. (d) Les phénomènes de bursts dans la synthèse des protéines (d1) et des ARN (d2) sont des régimes très particuliers et ont lieu respectivement lorsque la durée de vie des ARN et les périodes actives du promoteur sont très courtes. (e) Le rapport temporel entre les différents niveaux du processus d'expression a une influence majeure sur la propagation de la stochasticité d'un niveau à l'autre : une dynamique rapide du promoteur par rapport aux durées de vie des ARN et des protéines n'influence que très marginalement le niveau d'expression (e1) alors qu'une dynamique lente du promoteur peut être la source prédominante de la variabilité du niveau d'expression (e2).

qu'eucaryotes (Blake *et al.*, 2003; Raser et O'Shea, 2004; Bar-Even *et al.*, 2006; Newman *et al.*, 2006).

2.3 Bursts traductionnels

Dans la littérature, la signature précédente a parfois été caractérisée par le facteur de Fano $\sigma_P^2/\langle P\rangle$ (cf Box 1) et interprétée en termes de *bursts traductionnels*, correspondant à de rares et brèves périodes d'intense synthèse protéique (figure I.2d1) (Thattai et van Oudenaarden, 2001; Ozbudak *et al.*, 2002). Ce phénomène, bien que démontré par des techniques de détection de molécules uniques dans certaines situations chez le procaryote (Yu *et al.*, 2006) et chez l'eucaryote (Cai *et al.*, 2006) n'est pas représentatif du cas général. En effet, comme le souligne Paulsson (2005a), ce régime d'expression très particulier nécessite (i) que le nombre de transcrits soit nul la plupart du temps ($\rho/\gamma \ll 1$) et (ii) que ces derniers aient une durée de vie très courte comparée à celle des protéines ($1/\gamma \ll 1/\tilde{\gamma}$). Dans ce cas, la synthèse protéique à lieu en « paquets » de distribution géométrique[4] et de taille moyenne $b = \tilde{\rho}/\gamma$, correspondant au nombre de protéines synthétisées par transcrit. Le facteur de Fano peut alors s'écrire

$$\frac{\sigma_P^2}{\langle P\rangle} \simeq 1 + b \tag{I.8}$$

Cependant, outre le fait que le facteur de Fano ne soit pas une mesure pertinente de la variabilité dans le cas général, des mesures cohérentes avec l'équation (I.8) (c'est-à-dire une indépendance à $\langle P\rangle$ et une valeur élevée de b) ne signifient pas nécessairement que la traduction a lieu en bursts (Paulsson, 2005a). En effet, ces caractéristiques ne sont autres que celles évoquées sur la relation (I.7) (resp. une dépendance en $1/\langle P\rangle$ de $\sigma_P^2/\langle P\rangle^2$ et un coefficient de proportionnalité important) et signifient uniquement que la variabilité est due aux événements de synthèse/dégradation des transcrits. Les deux conditions pour un régime d'expression en bursts traductionnels n'ont donc nullement besoin d'être remplies. De plus, dans le cas général, la valeur de b mesurée ne correspond pas au nombre de protéines produites par transcrit. Par exemple, dans le cas de la figure I.2c3, on observerait bien $\sigma_P^2/\langle P\rangle = 1 + b$ avec $b \simeq 11,5$ quel que soit le niveau d'induction, alors que la protéosynthèse n'a pas lieu en bursts et que le nombre de protéines produites par transcrit est de 15. Notons aussi que, contrairement à ce que suggère le terme de bursts traductionnels, ceux-ci ne reflètent pas le caractère aléatoire de la traduction, mais proviennent des événements stochastiques de transcription et de dégradation des transcrits. En effet, plus la synthèse/dégradation protéique s'approche du cas déterministe, plus b devient grand.

L'interprétation en termes de bursts traductionnels a conduit un certain nombre de modèles à agréger la synthèse d'un transcrit, sa traduction (en une quantité suivant une loi géométrique) et sa dégradation en une seule étape (Mettetal *et al.*, 2006; Friedman *et al.*, 2006; Sánchez et Kondev, 2008). Ce choix de modélisation revient à supposer *a priori* ce régime d'expression très particulier et limite donc significativement l'expressivité du système.

[4]Plus précisément, il s'agit d'une loi de Poisson dont le paramètre suit une loi exponentielle, ce qui est très proche d'une loi géométrique.

2.4 Dynamique du promoteur

L'activation/inactivation du promoteur due à la présence/absence d'un FT ou à l'ouverture/fermeture de la chromatine a depuis longtemps été prise en compte comme source de stochasticité dans les approches de modélisation (Rigney et Schieve, 1977; Ko, 1991; Peccoud et Ycart, 1995; Kepler et Elston, 2001). Elle est aujourd'hui couramment employée dans l'analyse la variabilité de données d'expression (Blake *et al.*, 2003; Raser et O'Shea, 2004; Becskei *et al.*, 2005; Bar-Even *et al.*, 2006; Newman *et al.*, 2006). En effet, Becskei *et al.* (2005) et Raj *et al.* (2006), en observant que l'intégration de deux rapporteurs fluorescents à un même locus chromosomique provoque une forte corrélation de leur niveau d'expression, ont démontré que l'environnement chromatinien d'un gène joue un rôle essentiel sur son niveau de variabilité. De même, des observations directes de la synthèse d'ARNm ont révélé que l'activité transcriptionnelle pouvait présenter de fortes hétérogénéités (Newlands *et al.*, 1998; Golding *et al.*, 2005; Chubb *et al.*, 2006; Zenklusen *et al.*, 2008), alternant entre des périodes de synthèse et des périodes d'inactivité. La distribution exponentielle de la durée de ces périodes (ou la distribution géométrique de la quantité de transcrits produits pendant chaque période active) a conduit la grande majorité des approches de modélisation considérant une dynamique explicite du promoteur à ne représenter celui-ci que comme un système à deux états (Simpson *et al.*, 2004; Paulsson, 2004, 2005a; Raser et O'Shea, 2004; Bar-Even *et al.*, 2006; Kærn *et al.*, 2005; Kaufmann et van Oudenaarden, 2007; Shahrezaei et Swain, 2008b; Hornos *et al.*, 2005; Golding et Cox, 2006; Cox *et al.*, 2006; Lipniacki *et al.*, 2006; van Zon *et al.*, 2006; Innocentini et Hornos, 2007; Okabe *et al.*, 2007; Paszek, 2007; Tkačik *et al.*, 2008; Shahrezaei et Swain, 2008a), tel que représenté par la réaction (I.1).

Dans ce cas simple où l'on considère que le promoteur bascule entre un état actif et un état inactif (figures I.2a et I.2b), la variance normalisée peut être exprimée analytiquement, chacun des termes de l'expression pouvant alors être interprété d'un point de vue biophysique (reformulation de Paulsson, 2005a) :

$$\frac{\sigma_P^2}{\langle P\rangle^2} = \overbrace{\underbrace{\frac{1}{\langle P\rangle}}_{\text{Poisson}}}^{\text{Synthèse/dégradation des protéines}} + \overbrace{\underbrace{\frac{1}{\langle R\rangle}}_{\text{Poisson}} \underbrace{\frac{\tilde{\gamma}}{\gamma+\tilde{\gamma}}}_{\text{Lissage temporel ARN}\to\text{protéines}}}^{\text{Synthèse/dégradation des ARN}} + \overbrace{\underbrace{\frac{\langle G_{\text{off}}\rangle}{\langle G_{\text{on}}\rangle}}_{\text{Bernoulli}} \underbrace{\frac{\tilde{\gamma}}{\gamma+\tilde{\gamma}} \frac{1+\gamma/(\tilde{\gamma}+k^+ + k^-)}{1+(k^+ + k^-)/\gamma}}_{\text{Lissage temporel promoteur}\to\text{ARN}\to\text{protéines}}}^{\text{Activation/inactivation du promoteur}} \tag{I.9}$$

En plus des deux termes dus aux événements de synthèse/dégradation déjà présents dans l'équation (I.7), le troisième terme représente la variabilité provenant de la dynamique stochastique d'activation/inactivation du promoteur. Il s'agit de la variance normalisée de l'état du promoteur $\langle G_{\text{off}}\rangle/\langle G_{\text{on}}\rangle$ (rapport des probabilités que le promoteur soit inactif et actif, sa distribution stationnaire suivant une loi de Bernoulli) pondérée par un facteur prenant en compte les rapports temporels entre les différents niveaux (voir section suivante). Cependant, même dans le cas simple d'un promoteur à deux états, il n'est pas possible de prédire une signature claire dans la relation entre $\sigma_P^2/\langle P\rangle^2$ et $\langle P\rangle$. Une telle signature est en effet observable uniquement dans le régime très particulier où les transitions d'état du promoteur sont très fréquentes comparées à la durée de vie des ARN

et si les périodes actives sont sensiblement plus courtes que les périodes d'inactivité (figure I.2d2). Dans ce régime de *burst transcriptionnel*, le processus de synthèse des ARN est analogue à celui des protéines dans le cas de bursts traductionnels (figure I.2d1) et on observe la même signature en $1/\langle P \rangle$ (Paulsson, 2005a). Ces deux sources de stochasticité ne peuvent alors pas être séparées sans recourir à des mesures complémentaires, telles que des expériences de mutagenèse du promoteur (Raser et O'Shea, 2004) et/ou du site de fixation du ribosome (Ozbudak *et al.*, 2002).

Bien que ce modèle à deux états soit employé dans la très grande majorité des approches qui considèrent explicitement une dynamique du promoteur, il ne permet pas de rendre compte d'un grand nombre de phénomènes. En effet, aussi bien chez les procaryotes que chez les eucaryotes, la régulation fait intervenir un grand nombre de molécules (Browning et Busby, 2004; Perissi et Rosenfeld, 2005; Thomas et Chiang, 2006) avec des rôles et des propriétés cinétiques très variés (Hager *et al.*, 2009). L'approximation de la dynamique du promoteur par la simple réaction bidirectionnelle (I.1) constitue donc une simplification très forte. Elle implique que le passage d'un état à l'autre se fait avec un délai exponentiel, ce qui revient à considérer que la transition a lieu en une seule étape ou qu'une des étapes du processus est largement plus limitante que toutes les autres. Or, même si certains promoteurs présentent de telles caractéristiques, cette propriété est loin d'être une généralité. De fait, les cas où des distributions exponentielles de la durée des deux états (ou géométriques du nombre de transcrits par période active) ont effectivement été directement observées et mesurées sont plutôt rares (Chubb *et al.*, 2006; Golding *et al.*, 2005) et des dynamiques d'activation/inactivation singulièrement différentes peuvent être observées d'un gène à l'autre (Zenklusen *et al.*, 2008). En outre, plusieurs études basées sur des cinétiques de ChIP[5] ont démontré un comportement cyclique chez de nombreux promoteurs eucaryotes (Shang *et al.*, 2000; Métivier *et al.*, 2003; Karpova *et al.*, 2008; Degenhardt *et al.*, 2009), ce qui peut alors conduire à des comportements périodiques dans la synthèse des transcrits. De façon générale, des études récentes à l'échelle du génome entier montrent l'existence de modes de régulation très différents (van Werven *et al.*, 2009; Cairns, 2009).

Certaines données expérimentales ne peuvent pas être expliquées avec un modèle de promoteur à deux états et nécessitent une description plus précise (Blake *et al.*, 2003, 2006; Murphy *et al.*, 2007). Plusieurs études prennent ainsi en compte la dynamique du promoteur en considérant par exemple que l'état du promoteur est déterminé par la présence de deux molécules s'influençant mutuellement (Blake *et al.*, 2003, 2006), ou en prenant en compte plusieurs étapes entre la fixation de l'ARN polymérase et l'initiation de la transcription (Mitarai *et al.*, 2008). D'autre modèles permettent de représenter de façon plus générale l'interaction d'un nombre arbitraire de molécules sur un promoteur (Vilar et Saiz, 2005; Saiz et Vilar, 2006a; Sánchez et Kondev, 2008). Ces études ont permis de montrer d'un point de vue théorique qu'un tel système permettait un contrôle fin de la stochasticité de l'expression génique.

[5]L'immunoprécipitation de la chromatine (ChIP) est une technique expérimentale permettant d'observer, au sein d'une population de cellules, le taux d'occupation d'une séquence d'ADN par certaines protéines (ici l'occupation d'un promoteur par des facteurs de transcription). Une cinétique de ChIP, réalisée après synchronisation de la population, permet de décrire l'évolution de cette occupation au cours de l'activation transcriptionnelle d'un gène. Cette technique sera décrite plus en détail dans la section 5.

Il paraît donc essentiel de prendre en compte la dynamique complexe résultant des interactions moléculaires sur le promoteur pour l'étude de la génération et du contrôle de la variabilité de l'expression génique. De fait, des études à grande échelle, telles que celle réalisée par Newman *et al.* (2006), ont montré que le niveau de stochasticité d'un gène était fortement corrélé à de nombreuses caractéristiques de son promoteur (présence d'une TATA box, susceptibilité à différents types de remodelages chromatiniens et de modifications d'histones, ...). Pourtant, bien que la littérature expérimentale décrivant la dynamique de la régulation transcriptionnelle soit très vaste et témoigne de propriétés dynamiques riches et diverses, très peu de modèles de l'expression stochastique des gènes la prennent en compte. Nous reviendrons plus en détails sur cette littérature dans la section 5 de ce chapitre, une partie essentielle de notre travail consistant en l'élaboration et l'utilisation d'un modèle générique permettant précisément de prendre en compte ces aspects (chapitre III).

2.5 Relations entre échelles de temps

Comme nous l'avons vu, la stochasticité de l'expression génique met en jeu de nombreux mécanismes à différents niveaux, depuis les interactions moléculaires sur le promoteur jusqu'à la traduction et la dégradation des protéines. Or, les fluctuations produites à un niveau donné du processus d'expression peuvent être plus ou moins transmises par les niveaux suivants. Un des aspects centraux dans cette transmission est la relation entre les échelles de temps des différents processus. La durée de vie d'une molécule dont le niveau est soumis à un processus de synthèse/dégradation détermine le temps de relaxation de ce dernier, c'est-à-dire le temps qu'il met pour retourner à l'équilibre à partir d'une valeur hors d'équilibre. Elle détermine alors la durée caractéristique de ses fluctuations : intuitivement, la durée d'un écart à sa moyenne[6]. Par exemple, P sur la figure I.2c1 et R sur la figure I.2c3 suivent exactement le même processus de synthèse/dégradation, mais ont des durées de vie différentes.

Le temps de relaxation d'un niveau (donc la durée de vie de la molécule correspondante) détermine également la vitesse à laquelle celui-ci est capable de réagir à des changements en amont. Ainsi, si la durée de vie moyenne des transcrits ($1/\gamma$) est sensiblement plus longue que le temps moyen entre deux activations du promoteur (à savoir $1/k^- + 1/k^+$), alors ces dernières n'auront que peu d'influence sur la variabilité finale du niveau d'expression (figure I.2e1). À l'inverse, des fluctuations du promoteur sensiblement plus lentes que la durée de vie des transcrits et des protéines seront transmises jusqu'à ces niveaux (figure I.2e2). Ainsi, dans un cas intermédiaire, les deux processus de synthèse/dégradation effectuent un lissage temporel des fluctuations du promoteur, ne laissant passer que les variations globales (figure I.2b).

Ce phénomène est décrit quantitativement par les coefficients de lissage temporel dans les équations (I.7) et (I.9). Par exemple, dans l'équation (I.7) (et dans la figure I.2c3), les fluctuations du niveau d'ARN – le terme $1/\langle R\rangle$ – sont d'autant plus amorties qu'elles

[6]Plus formellement, il s'agit du coefficient de l'exponentielle dans l'autocorrélation du processus ou de la fréquence à laquelle commence la décroissance dans son spectre de puissance, appelée fréquence de coupure.

sont rapides par rapport à la durée de vie des protéines ($\gamma \gg \tilde{\gamma}$). Dans l'équation (I.9), les changements d'état du promoteur – dont la variance normalisée est $\langle G_{\text{off}} \rangle / \langle G_{\text{on}} \rangle$ – sont filtrés par un terme prenant en compte la relation entre les échelles de temps des trois processus impliqués (promoteur, ARN et protéines).

L'utilisation d'indicateurs décrivant les aspects dynamiques du niveau d'expression, tels que le spectre de puissance ou l'auto-corrélation (cf Box 1), permet de caractériser ce phénomène de façon précise. Ainsi, plusieurs auteurs ont montré que la transmission des fluctuations d'un niveau à un autre (par exemple de R à P) faisait intervenir un filtrage passe-bas dont la fréquence de coupure est déterminée par la durée de vie des molécules du second niveau (Simpson *et al.*, 2003; Warren *et al.*, 2006; Cox *et al.*, 2006; Lestas *et al.*, 2008). Cela signifie par exemple que le contenu fréquentiel du niveau d'ARN à des fréquences sensiblement inférieures à $\tilde{\gamma}$ (le taux de dégradation protéique) est toujours présent dans le spectre de puissance du niveau de protéines, tandis que les fréquences plus élevées disparaissent d'autant plus que l'on s'éloigne de la fréquence de coupure $\tilde{\gamma}$.

2.6 Distribution des délais : élongation, export, dégradations ...

Le modèle de la figure I.2a considère que les délais des différentes réactions[7] sont distribués exponentiellement, ce qui constitue une approximation très forte. Comme nous l'avons évoqué dans la section 2.4, les nombreuses réactions impliquées dans le processus de régulation transcriptionnelle rendent généralement non-exponentiels les délais entre l'activation et l'inactivation du promoteur (I.1). De la même façon, on peut argumenter que chacune des autres réactions de ce système est en réalité non-exponentielle. Par exemple, la dégradation des transcrits (I.3) n'est ni spontanée ni instantanée puisqu'elle se produit par digestion enzymatique progressive à partir de la queue poly-A du transcrit. De même, pour la réaction (I.5), une protéine doit subir plusieurs réactions d'ubiquitylation avant d'être ciblée par le protéasome. Les distributions non-exponentielles des délais de ces deux réactions ainsi que celle de l'activation/inactivation du promoteur ont été prises en compte de façon générique par Pedraza et Paulsson (2008), montrant alors l'importance de ces aspects sur la variabilité du niveau d'expression du gène.

De même, les réactions (I.2) et (I.4) sont très loin d'avoir des durées exponentielles. En particulier, la réaction (I.2) suppose nul le délai entre l'initiation de la transcription et l'apparition d'un ARN messager prêt à être traduit. On sait pourtant que de nombreux mécanismes aux dynamiques temporelles singulières interviennent dans ce processus, tels l'élongation de la transcription (s'effectuant de façon très irrégulière avec de nombreuses pauses), le splicing, l'export hors du noyau (dont les propriétés dépendent de nombreux facteurs comme la position spatiale du gène dans le noyau), ... Bien que ces mécanismes soient décrits de façon relativement précise d'un point de vue expérimental (Darzacq *et al.*, 2005; Shav-Tal *et al.*, 2006) et fassent l'objet d'études de modélisation en tant que tels (Roussel et Zhu, 2006; Voliotis *et al.*, 2008), leurs implications spécifiques dans le cadre de l'expression stochastique des gènes reste à ce jour encore peu explorées (Dobr-

[7] Le délai d'une réaction correspond à la durée entre l'instant où tous les réactants sont disponibles et l'instant où la réaction a lieu. Pour une réaction chimique élémentaire (se faisant en seule étape), cette durée a une distribution exponentielle dont la valeur moyenne est l'inverse du taux de la réaction.

zyński et Bruggeman, 2009). Nous verrons au chapitre III, section 3.6, qu'il est possible de prendre en compte simplement des délais non-exponentiels arbitraires pour chacune des réactions (I.2) à (I.5), permettant alors de rendre compte, au moins partiellement, des divers processus couramment approximés comme instantanés dans les modèles d'expression stochastique des gènes.

2.7 Aspects spatiaux

La grande majorité des modèles d'expression stochastique des gènes considèrent un espace homogène et implicite dans lequel les molécules sont réparties de façon uniforme. Bien que cette supposition permette l'application de nombreuses techniques de modélisation et de simulation (Gillespie, 1977; Turner *et al.*, 2004; Wilkinson, 2009), elle reste cependant une simplification forte de la réalité et néglige plusieurs sources essentielles de la variabilité de l'expression. En effet, des études expérimentales chez différents organismes eucaryotes ont montré que l'intégration de deux gènes rapporteurs en un même locus chromosomique provoque une corrélation significative de leurs niveaux d'expression (Becskei *et al.*, 2005; Raj *et al.*, 2006), indiquant alors l'implication de mécanismes locaux tels que la condensation/décondensation de la chromatine et/ou le déplacement du locus chromosomique dans différentes régions de l'espace nucléaire. De plus, en employant des techniques de détection de molécules uniques, Elf *et al.* (2007) ont montré chez *E. coli* que le mouvement observé des facteurs de transcription (FT) était compatible avec un modèle où le déplacement s'effectue par une combinaison de diffusion libre en 3D dans l'espace cellulaire et de diffusion en 1D ou de *sliding* le long de l'ADN (figure I.3a1). Ce principe de diffusion « 3D+1D » a été très largement étudié du point de vue de la modélisation pour expliquer comment les FT trouvent efficacement leurs sites de fixation par l'alternance de périodes de recherche globale et de recherche locale (Hołyst *et al.*, 2000; Halford et Marko, 2004; Gowers *et al.*, 2005; Wunderlich et Mirny, 2008; Halford, 2009).

Cependant, la présence d'autres molécules sur l'ADN est susceptible de modifier sensiblement les propriétés d'une telle diffusion (Li *et al.*, 2009). En particulier, chez les eucaryotes, la chromatine, structure moléculaire dans laquelle est enroulé l'ADN, rend difficile une diffusion libre en 1D. Dans ce contexte, les mesures expérimentales (sur lesquelles nous reviendrons dans la section 4) suggèrent que les molécules se déplacent plutôt par une série d'associations transitoires à l'ADN séparées par de brèves périodes de diffusion 3D sur de courtes distances – un processus qualifié de *hopping* ou de *hit-and-run* (figure I.3a2) (Misteli, 2001, 2007) – tandis que les phénomènes de *sliding* seraient plutôt couplés à des moteurs moléculaires tels que la machinerie de transcription ou de réplication (Talbert et Henikoff, 2006).

L'influence spécifique de ces aspects spatiaux sur la stochasticité de l'expression génique fait cependant rarement l'objet d'études de modélisation. Les quelques exemples témoignent pourtant de l'importance de ces phénomènes. Par exemple, même s'il ne s'agit dans cette étude que d'une diffusion en 3D dans un espace homogène, van Zon *et al.* (2006) montrent que la diffusion d'un FT dans la cellule peut provoquer d'importantes hétérogénéités transcriptionnelles lorsque celui-ci est présent en un très faible nombre de copies. Typiquement, s'il n'existe qu'une seule copie de ce facteur dans la cellule, le FT peut

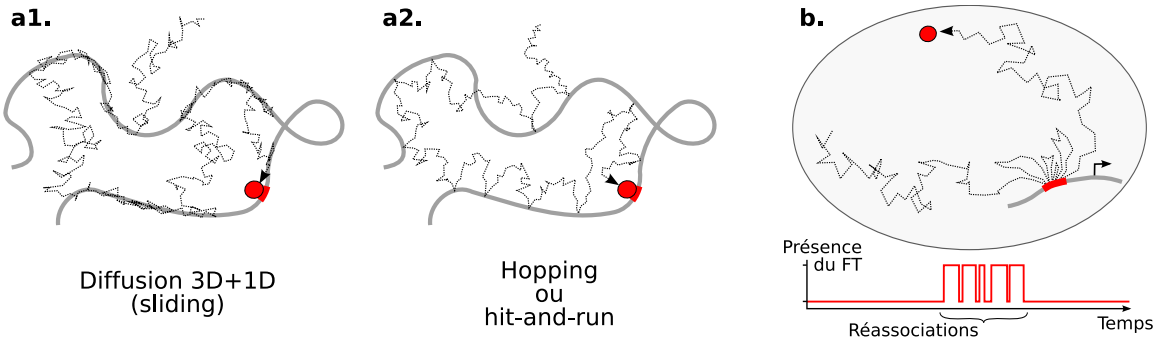


Fig. I.3: Modes de déplacement d'un facteur de transcription. Pour expliquer la façon dont un facteur de transcription (FT) trouve son site de fixation sur l'ADN, plusieurs modes de déplacement sont proposés dans la littérature. Par exemple, le FT pourrait se déplacer par (a1) une alternance entre des périodes de diffusion libre en 3D et de diffusion en 1D le long de l'ADN (*sliding*), ou par (a2) un mécanisme de *hopping* ou de *hit-and-run* consistant en une suite d'associations (sans déplacement) séparées par de courtes périodes de diffusion en 3D. Diverses mesures expérimentales chez les procaryotes et chez les eucaryotes peuvent être en accord avec ces modes de déplacement et leur efficacité dans la recherche du site de fixation est une question largement étudiée. Leur implication sur la stochasticité de l'expression génique reste cependant peu explorée. (b) En ne considérant que la diffusion 3D d'un FT et son association à un unique site de fixation, van Zon *et al.* (2006) ont montré que la réassociation rapide du FT pouvait significativement augmenter l'hétérogénéité dans l'activité de transcription d'un gène.

diffuser loin du promoteur pendant une longue période, puis se rapprocher et provoquer plusieurs association/dissociation successives avant de s'éloigner à nouveau (figure I.3b). De plus, Tkačik et Bialek (2009) montrent que, de façon inattendue, un processus de diffusion « 3D+1D » ne réduit pas la variabilité du processus d'expression génique, du fait d'importantes corrélations temporelles lors de la diffusion en 1D.

Ainsi, la prise en compte explicite de l'espace permet de mettre en évidence des mécanismes pertinents quant à la stochasticité de l'expression génique. Or, si les travaux théoriques se basant sur le déplacement des FT sont plutôt rares, ceux qui tiennent compte de l'organisation spatiale du noyau eucaryote à plus grande échelle le sont encore davantage. La littérature expérimentale récente sur cette architecture nucléaire dynamique est pourtant très abondante et témoigne d'un paysage hétérogène et complexe, sujet à de nombreux phénomènes d'auto-organisation basés sur le comportement probabiliste des molécules et décrivant l'expression génique comme un processus intrinsèquement couplé à une dynamique spatio-temporelle. Une partie de nos travaux consiste en l'élaboration d'un modèle et d'un logiciel de simulation ayant pour but d'étudier ces phénomènes d'auto-organisation au sein du noyau cellulaire ainsi que leur influence sur la stochasticité de l'expression génique (chapitre II). Nous reviendrons en détail sur cette littérature expérimentale dans la section 4 du présent chapitre de façon à mettre en avant les nombreux liens possibles avec l'expression stochastique des gènes.

3 Propagation et modulation de la stochasticité

Nous avons jusqu'ici abordé la question de la génération de la stochasticité et décrit les différentes sources relatives au processus d'expression génique. Nous présentons dans cette section comment l'interaction de gènes entre eux et/ou la rétroaction directe de l'activité d'un gène sur sa propre transcription permet à la fois de structurer la stochasticité qu'ils génèrent et de moduler leur capacité à transmettre ou à filtrer des fluctuations extérieures.

3.1 Transmission des fluctuations gène-extrinsèques

Les fluctuations observées dans l'expression d'un gène donné proviennent en partie du processus d'expression génique mais peuvent également refléter des fluctuations extérieures au gène : même si l'on considère qu'un gène est déterministe, les fluctuations de la quantité de régulateurs ou de polymérase, par exemple, vont se répercuter sur son niveau d'expression (figure I.1a1). Cette stochasticité gène-extrinsèque peut être prépondérante dans certains cas (Becskei *et al.*, 2005; Volfson *et al.*, 2006).

Une approche classique de modélisation pour étudier ce mécanisme est de considérer un modèle dans lequel un gène A régule un gène B. On peut alors étudier les propriétés de l'expression stochastique du gène B en fonction de celles du gène A, cette méthode pouvant également être généralisée à une cascade de gènes (Thattai et van Oudenaarden, 2001; Simpson *et al.*, 2003; Pedraza et van Oudenaarden, 2005; Maheshri et O'Shea, 2007; Warmflash et Dinner, 2008; Walczak *et al.*, 2009). Dans un tel système (dépourvu de rétrocontrôles), la part gène-extrinsèque des fluctuations de B provient des fluctuations de A et est déterminée non seulement par les rapports temporels entre les différentes dynamiques (similairement à ce qui est décrit dans la section 2.5), mais également par l'allure de la fonction de régulation (le niveau moyen de B en fonction de celui de A). Cette propriété est généralisable à plusieurs régulateurs et pour une CRIF[8] arbitraire (Warmflash et Dinner, 2008). Dans le cas de fluctuations extrinsèques de faible amplitude, on peut appliquer l'approximation linéaire du bruit (*linear noise approximation*) (Thattai et van Oudenaarden, 2001; Paulsson, 2005a). Les fluctuations de B sont alors proportionnelles à celles de A et à la pente de la CRIF (au niveau de la valeur moyenne de A). Dans le cas de fluctuations extrinsèques importantes, la variabilité de B dépendant alors de la forme de la CRIF sur toute la plage de valeurs atteintes par A, ce qui peut conduire à un régime d'expression bimodal (Kærn *et al.*, 2005).

Ainsi, un gène peut aussi bien générer de la stochasticité que transmettre une stochasticité reçue en amont. Cependant, lorsque les fluctuations de concentration des régulateurs d'un gène sont corrélées entre elles et/ou corrélées au niveau d'expression du gène considéré, les dépendances entre les différentes sources de variabilité peuvent conduire à l'apparition de propriétés particulières – l'un des cas de figure les plus simples étant celui de l'autorégulation.

[8]CRIF (*cis*-regulatory input function) : fonction décrivant le niveau moyen d'un gène en fonction de la concentration de tous ses régulateurs.

3.2 Modulation par l'autorégulation

Il a été montré par de nombreuses études théoriques (Thattai et van Oudenaarden, 2001; Simpson *et al.*, 2003; Swain, 2004; Hornos *et al.*, 2005; Cox *et al.*, 2006; Tao *et al.*, 2007) et/ou expérimentales (Becskei et Serrano, 2000; Dublanche *et al.*, 2006; Nevozhay *et al.*, 2009), que l'auto-répression réduit la variabilité de l'expression d'un gène, que son origine soit gène-intrinsèque ou gène-extrinsèque (Paulsson, 2004, 2005a). En effet, une fluctuation positive provoquant par rétroaction une diminution de l'expression et *vice versa*, la variabilité autour de la moyenne est globalement réduite. D'un point de vue spectral, l'auto-répression modifie le contenu fréquentiel du niveau d'expression en déplaçant les fluctuations basses-fréquences vers des fréquences plus élevées (Simpson *et al.*, 2003; Cox *et al.*, 2006). Notons que l'on peut également retrouver ce phénomène – parfois qualifié de *noise shaping* – dans d'autres systèmes biologiques utilisant un rétrocontrôle négatif en présence de bruit, par exemple les réseaux de neurones (Mar *et al.*, 1999).

À l'inverse, l'auto-activation provoque une augmentation de la variabilité de l'expression en amplifiant les fluctuations (Kepler et Elston, 2001; Thattai et van Oudenaarden, 2001; Lestas *et al.*, 2008). Pour une auto-activation importante, cela peut conduire le système à être bistable, alternant entre un régime d'expression faible et un régime d'expression élevé. Le temps d'échappement entre ces deux états dépend alors de l'amplitude des fluctuations et de l'intensité de l'auto-activation. Il est toutefois à noter que dans certains cas, l'auto-activation peut également améliorer les propriétés de sensitivité et de transmission d'information du gène (Hornung et Barkai, 2008).

D'autres rétrocontrôles au sein même des processus de transcription et de traduction d'un gène peuvent également provoquer des propriétés de modulation de la stochasticité (Swain, 2004). D'un point de vue général, la prise en compte de rétrocontrôles se traduit par l'apparition de coefficients dits d'*élasticité* dans l'équation (I.9) (Paulsson, 2004, 2005a). Chaque élasticité décrit à quel point un niveau donné (par exemple la quantité de protéines) est capable d'influencer le processus de synthèse/dégradation d'un autre niveau (par exemple, l'activation/inactivation du promoteur). Leur combinaison détermine alors comment les propriétés de transmission ou de génération de la stochasticité du système sont modifiées par un rétrocontrôle donné.

3.3 Motifs de gènes : corrélations constructives et destructives

On peut retrouver des propriétés similaires à l'autorégulation en considérant l'interaction de plusieurs gènes. Par exemple, un gène A activant un gène B, lui-même réprimant A, provoque de propriétés similaires à l'auto-répression (Simpson *et al.*, 2003; Maheshri et O'Shea, 2007). D'autres motifs de quelques gènes faisant intervenir des rétrocontrôles positifs et négatifs avec des délais différents peuvent provoquer des propriétés riches de filtrage et de transmission des sources motif-extrinsèques ou générer des fluctuations aux propriétés particulières (Thattai et van Oudenaarden, 2001; Cox *et al.*, 2006; Pedraza et van Oudenaarden, 2005; Maheshri et O'Shea, 2007; Shahrezaei *et al.*, 2008; Walczak *et al.*, 2009). Par exemple, si deux gènes A et B sont tous deux sujets à des fluctuations globales (par exemple de la quantité d'ARN polymérase) et que A réprime B, alors l'effet

des fluctuations globales sur B sera en partie compensé par la réaction de A – il s'agit là d'une corrélation destructive entre deux sources de fluctuations. D'un point de vue spectral, Cox *et al.* (2006) montrent que, selon ce principe, l'interaction de plusieurs gènes peut provoquer des dynamiques stochastiques relativement structurées, avec en particulier des comportements oscillants prononcés.

Ainsi, la dynamique couplée de nombreux gènes – dont le comportement individuel est représenté façon simple – permet d'obtenir des comportements aux propriétés variées. Il semble alors probable que ces mécanismes de contrôle et de structuration de la stochasticité soient exploités avantageusement par la cellule en fonction du contexte. Nous verrons au chapitre III comment de nombreuses propriétés similaires peuvent également être obtenues au niveau d'un gène unique dès lors que l'on considère explicitement la dynamique des différentes molécules sur son promoteur.

3.4 Ergodicité : mémoire cellulaire et interactions cellule-cellule

Les interactions mutuelles entre plusieurs gènes sont susceptibles de provoquer des propriétés de multistabilité. Par exemple, deux gènes s'inhibant mutuellement constituent un système bistable dans lequel le temps de passage d'un état à l'autre dépend de l'intensité de l'inhibition et l'importance des fluctuations dans l'expression des deux gènes (Kepler et Elston, 2001) et peut alors théoriquement être rendu arbitrairement lent. Une propriété bien connue des réseaux en général est que le nombre d'états stables du système augmente avec la taille du réseau (Kauffman, 1993). Ainsi, dans le cadre d'un grand réseau de gènes dont l'expression est stochastique, le système peut facilement présenter une dynamique plus ou moins lente de transitions stochastiques entre des nombreux états. Comme nous allons le voir, c'est en effet ce que suggèrent les données expérimentales soulevant alors des questions quant au caractère ergodique du comportement des cellules.

Un processus stochastique est dit *ergodique* si les statistiques obtenues sur une réalisation donnée du processus sont identiques à celles obtenues sur un ensemble de réalisations indépendantes observées à un instant donné. Cela revient simplement, dans notre cas, à poser la question : le comportement d'une cellule est-il représentatif de celui de la population ? En effet, une même distribution observée en cytométrie peut dériver de dynamiques très différentes au niveau de la cellule unique (Sigal *et al.*, 2006; Brock *et al.*, 2009).

Considérons l'exemple simple d'un système bistable. S'il est impossible pour une cellule dans l'un des deux états de passer dans l'autre état, l'observation d'une cellule unique (qui reste donc indéfiniment dans l'un ou l'autre des deux états) n'est clairement pas représentative de la population. Le système n'est donc pas ergodique. À l'inverse, s'il est possible de passer d'un état à l'autre, on peut alors reproduire la distribution que l'on obtiendrait sur une population de cellules à partir de l'observation d'une seule cellule. Mais la durée nécessaire de l'observation pour arriver à cette propriété – appelé le *temps de mélange* du processus – dépend de la fréquence de passage d'un état à l'autre. Si ces événements sont rares, l'observation doit être très longue et le système est alors qualifié de faiblement ergodique. D'un point de vue plus général le temps de mélange d'un niveau d'expression correspond à la durée de corrélation la plus longue observée dans ses fluctuations et peut

être interprété comme la « mémoire » de la cellule. Cette valeur décrit en effet à quel point une cellule ayant un niveau très différent de la moyenne va conserver cette caractéristique au cours du temps.

Diverses études expérimentales suggèrent que ce temps de mélange peut parfois être très long. Par exemple, les fluctuations du niveau d'expression observées par Rosenfeld *et al.* (2005) chez *E. coli* par vidéo-microscopie ont un temps de corrélation de 40 min, donc de l'ordre du cycle cellulaire. Remarquablement, la part gène-intrinsèque de ces fluctuations étant sensiblement plus rapide (inférieure à 10 min), leur lenteur a pour origine des facteurs globaux. Sur des cellules humaines, Sigal *et al.* (2006) montrent que le temps de mélange du niveau d'expression de 20 protéines différentes est supérieur à 40 h – soit plusieurs cycles cellulaires – mais que chaque cellule finit par perdre la mémoire de ses niveaux d'expression. De même, après avoir trié par cytométrie de flux une population de cellules de souris en trois sous-populations exprimant une protéine donnée à des niveaux de fluorescence faibles, intermédiaires et élevés, Chang *et al.* (2008) observent que la distribution initiale peut être retrouvée à partir de chaque sous-population, mais que ce processus prend de l'ordre de 10 jours[9]. De plus, les auteurs montrent que les profils transcriptomiques[10] de ces trois sous-populations sont très différents après le tri mais convergent lentement vers un même profil. Cette observation indique alors que l'hétérogénéité observée ainsi que son temps de mélange font intervenir un grand nombre de gènes, suggérant alors des mécanismes de transition stochastique entre de nombreux états.

Ces résultats expérimentaux suggèrent également qu'il n'y a pas nécessairement de différence *a priori* entre les cellules d'une population et que l'individualité observée à un instant donné ne persiste pas indéfiniment. Mais la durée de ces fluctuations étant largement supérieure à la durée de la plupart des observations expérimentales, elles peuvent alors apparaître comme des différences de capacité d'expression entre les différentes cellules d'une population (Colman-Lerner *et al.*, 2005). De fait, des hétérogénéités dans l'expression de protéines impliquées, de façon générale, dans le processus d'expression génique (par exemple des sous-unités de la polymérase ou du ribosome, de FT généraux, des cofacteurs, ...) affectent l'expression de nombreux gènes de façon non-spécifique. Notons, de plus, que ces différences de capacité d'expression d'une cellule à l'autre affectent non seulement les niveaux moyens de l'expression des gènes, mais également les caractéristiques de leur stochasticité intrinsèque (Shahrezaei *et al.*, 2008). On observe alors une hétérogénéité des propriétés stochastiques de l'expression des gènes.

Le caractère ergodique de l'activité des cellules dépend non seulement de la composante la plus lente dans leur comportement spontané mais également, à une échelle supérieure, de l'existence d'interactions cellule-cellule au sein de la population. En effet, dans l'exemple simple d'un système bistable évoqué précédemment, si l'on considère que l'état d'une cellule dépend en partie de l'état de ses voisines, les cellules ne sont alors plus indépendantes et le comportement individuel n'est plus représentatif de la population. On rencontre typiquement ce phénomène dans de nombreux tissus au cours du développement. En effet, les cellules peuvent s'inter-stabiliser dans des phénotypes différents et le comportement

[9] Le même type d'expériences, réalisées sur des cellules humaines, a même montré des temps de mélange allant jusqu'à plusieurs semaines (G. Corre et A. Paldi, données non-publiées).

[10] C'est-à-dire les niveaux d'ARN de l'ensemble ou d'un grande partie des gènes.

individuel dépend alors fortement du contexte provoqué par la population (Snijder *et al.*, 2009; Boettiger et Levine, 2009).

*
* *

Ainsi, la stochasticité de l'expression génique est un phénomène faisant intervenir de nombreuses échelles d'organisation : de la molécule au tissu. Les fluctuations des niveaux protéiques prennent en effet leurs origines dans divers événements moléculaires propres au processus d'expression génique et sont alors transmises, modifiées et structurées par l'interaction des différents gènes du réseau et en relation avec le contexte de la cellule au sein du tissu. On constate, au regard de la littérature exposée ici, que deux aspects essentiels de l'expression génique ne sont généralement pas pris en compte dans l'étude de ses propriétés stochastiques. Comme nous l'avons vu à la section 2.7, le déplacement de certaines molécules dans l'espace peut être un aspect déterminant des caractéristiques stochastiques de l'expression génique. Mais ces phénomènes, encore peu étudiés, le sont presque exclusivement dans le contexte très simple d'un espace homogène et en ne considérant que peu de molécules (typiquement quelques facteurs de transcription). Il existe cependant une littérature expérimentale témoignant d'une organisation très hétérogène du noyau cellulaire eucaryote influençant le déplacement des molécules à de nombreux niveaux (aussi bien les facteurs de transcription que les transcrits ou les gènes eux-mêmes). Dans la section suivante, nous décrivons ces mécanismes et tentons de mettre en évidence leurs potentielles implications sur l'expression stochastique des gènes. Un second aspect très étudié expérimentalement mais rarement pris en compte dans les modèles d'expression stochastique est la dynamique de la régulation transcriptionnelle qui, comme nous le verrons à la section 5, fait intervenir de très nombreuses molécules. Cette dynamique présente en effet des propriétés très singulières à plusieurs échelles temporelles et nous verrons que des phénomènes stochastiques peuvent intervenir à divers niveaux.

4 Organisation spatio-temporelle du noyau cellulaire

L'expression génique est un processus intrinsèquement spatialisé dans lequel la rencontre et le déplacement des molécules et des structures qu'elles forment jouent un rôle central. Ces mouvements ont lieu au sein d'un paysage nucléaire structuré et en constante modification, susceptible d'influencer les propriétés dynamiques et stochastiques de l'expression des gènes.

4.1 Un espace hétérogène et dynamique

Le noyau cellulaire, observé à l'aide de techniques de microscopie classique, est longtemps apparu comme une structure homogène, un espace dans lequel l'ADN et les nombreuses protéines impliquées dans l'expression génique étaient réparties de façon à peu près uniforme. Seules quelques structures comme le nucléole et les corps de Cajal (sur lesquels nous reviendrons par la suite) étaient détectables (Ramón y Cajal, 1903). Les techniques récentes de fluorescence et de microscopie (Shav-Tal *et al.*, 2004b) ont depuis permis de montrer que le noyau est au contraire un système hétérogène, structuré et dynamique. En effet, la répartition spatiale d'une ou plusieurs espèces protéiques rendues fluorescentes par l'association de tags dérivés de la GFP (Green Fluorescent Protein) peut être observée en 3 dimensions et au cours du temps grace aux techniques de microscopie moderne. De telles expériences ont alors permis de mettre en évidence une organisation du noyau à diverses échelles (figure I.4). En particulier, différentes structures, appelées collectivement *corps nucléaires*, y ont été identifiées, correspondant à des domaines de l'espace nucléaire de composition particulière (figure I.4a) (Misteli, 2001, 2007; Handwerger et Gall, 2006; Spector, 2006; Shav-Tal *et al.*, 2006; Matera *et al.*, 2009; Darzacq *et al.*, 2009). On compte, parmi ces corps nucléaires, le nucléole, les corps de Cajal (parfois appelés coiled bodies), les speckles, les PML bodies, les usines à transcription, ... Le ou les rôles de toutes ces structures n'ont, pour la plupart, pas encore été clairement identifiés. Alors que certains corps nucléaires ont été montrés comme étant des sites préférentiels pour des réactions enzymatiques, d'autres pourraient ne servir que de régions de stockage de certaines molécules, ou encore n'avoir aucun rôle biologique et n'être qu'un résultat sans conséquence d'un autre processus (Misteli, 2001, 2005, 2007; Handwerger et Gall, 2006; Matera *et al.*, 2009). Toutefois, comme nous allons l'illustrer ici, il apparaît clairement que plusieurs de ces structures sont impliquées à de nombreux niveaux dans le processus d'expression génique. Par ailleurs, les techniques récentes d'imagerie permettent également d'observer l'organisation spatio-temporelle des chromosomes à l'échelle globale du noyau. On observe alors que, pendant l'interphase[11], les différents chromosomes ne sont pas répartis uniformément dans l'espace mais occupent des emplacements privilégiés au sein du noyau et les uns par rapport aux autres, définissant alors des régions appelées *territoires chromosomiques* (Misteli, 2001, 2005, 2007; Cremer et Cremer, 2001; Cremer *et al.*, 2006; Lanctôt *et al.*, 2007; Fraser et Bickmore, 2007). Ces régions n'ont cependant pas de frontière nette et, à l'interface de deux territoires voisins, l'ADN des deux chromosomes peut s'intermêler, permettant alors à des loci très distants en séquence d'interagir (Branco et Pombo,

[11] L'interphase est la phase du cycle cellulaire pendant laquelle les chromosomes sont décondensés.

2006). C'est en particulier dans ces régions inter-chromosomiques que l'on trouve une grande quantité de chromatine décondensée, l'euchromatine, et qu'ont lieu de nombreux phénomènes liés à l'expression génique, en relation avec différents corps nucléaires (Fraser et Bickmore, 2007).

Les questions relatives à l'origine et au maintien de cette organisation nucléaire et à son influence sur l'expression génique constituent un domaine de recherche relativement récent et encore essentiellement expérimental. De nombreuses questions restent encore ouvertes et, comme nous le verrons, différents points de vue coexistent à ce jour dans la littérature. Il y a toutefois des liens évidents, à de nombreux niveaux, entre cette organisation, sa dynamique et le processus d'expression génique, avec des relations d'influence mutuelle ouvrant la possibilité à de nombreux phénomènes d'auto-organisation (Misteli, 2001). On peut alors s'attendre à ce que le caractère stochastique de l'expression génique soit une composante essentielle de cette relation. En effet, de nombreux phénomènes, ayant lieu à des échelles spatiales et temporelles très variées, présentent un caractère intrinsèquement probabiliste : de la diffusion des molécules constituant un corps nucléaire, à la position relative des territoires chromosomiques, ou encore au mouvement d'un locus chromosomique ou d'un corps nucléaire. Il convient alors de comprendre comment ces différents niveaux d'organisation s'influencent mutuellement et, en particulier, comment ils affectent et sont affectés par l'expression génique. Nous commençons ici par décrire le comportement et les propriétés de déplacement des molécules dans le noyau cellulaire et leur influence sur la dynamique des corps nucléaires. Nous décrivons ensuite comment le processus d'expression génique s'inscrit dans le cadre de cette dynamique spatio-temporelle.

4.2 Déplacements moléculaires

Les propriétés de déplacement de nombreuses protéines dans le noyau sont caractéristiques d'un simple processus de diffusion en milieu hétérogène et sans transport actif (Seksek *et al.*, 1997; Phair et Misteli, 2000; Pederson, 2001; Misteli, 2001, 2007; Shav-Tal *et al.*, 2006; Darzacq *et al.*, 2009). Les coefficients de diffusion peuvent être très variés d'une région à l'autre et pour les différentes molécules mais sont de façon générale particulièrement élevés. De fait, certaines molécules peuvent traverser le noyau cellulaire en quelques secondes et d'imposants complexes tels que le spliceosome ou le ribosome, en quelques minutes (Misteli, 2001, 2007). Cette propriété est relativement étonnante puisque l'espace nucléaire est très encombré par de nombreuses biomolécules et que les molécules d'eau, qui ne représentent qu'une fraction relativement faible de l'espace, ont alors un comportement très différent de celui auquel nous sommes habitués lorsqu'elles sont en grandes quantités. En effet, elles sont organisées sous forme de couches d'hydratation autour des protéines adoptant alors des positions et des orientations préférentielles (Pal *et al.*, 2002; Bernini *et al.*, 2004). L'encombrement de l'espace nucléaire par des biomolécules, phénomène qualifié de *macromolecular crowding*, a pour effet de réduire l'espace accessible aux différentes molécules (*excluded volume effect*) et, par conséquence, de modifier les temps de rencontre entre molécules, les taux de réaction et les équilibres chimiques (Ellis, 2001; Saxton, 2001; Berry, 2002; Hancock, 2004; Misteli, 2007; Echevería *et al.*, 2007; Guigas et Weiss, 2008). L'espace accessible à une molécule donnée a en effet une morphologie bien

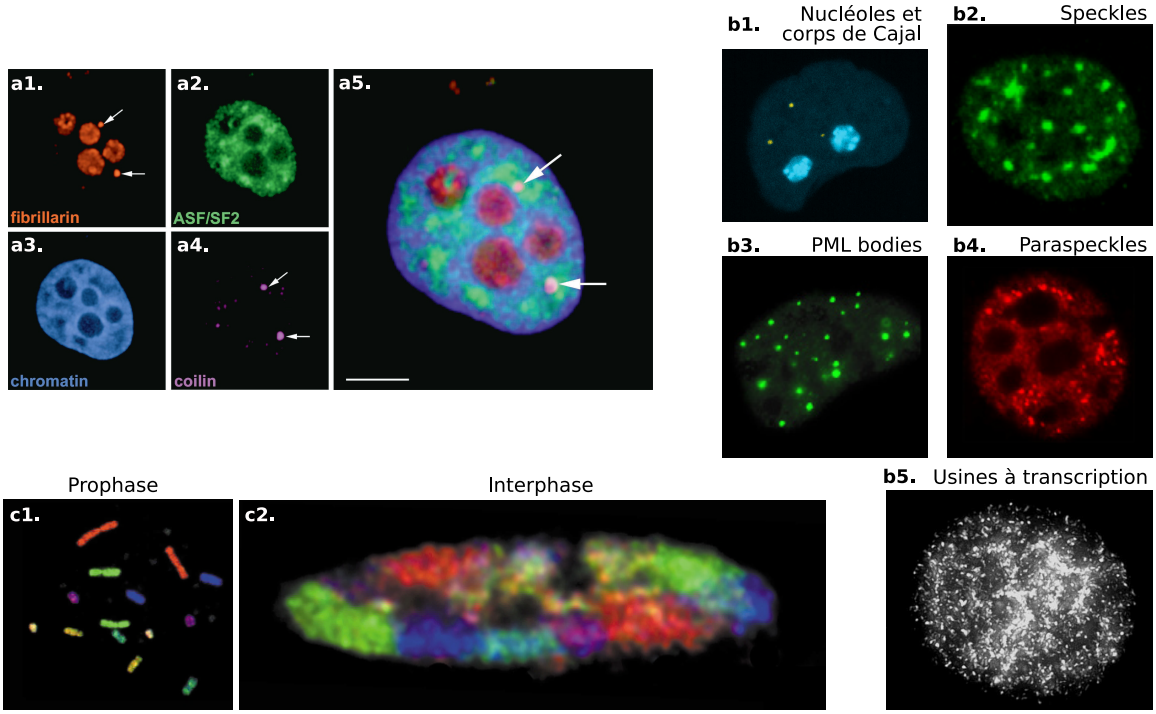


Fig. I.4: Observation expérimentale de l'organisation nucléaire. (a1-4) Le marquage par différents tags fluorescents de plusieurs protéines permet d'observer, par des techniques de microscopie confocale, leur distribution spatiale au sein du noyau cellulaire. (a5) On observe alors une forte hétérogénéité dans cette répartition, démontrant l'existence de régions de composition particulière, appelées collectivement corps nucléaires. Par exemple, les deux régions indiquées par des flèches contiennent une importante concentration de fibrillarine et de coiline, deux protéines caractéristiques des corps de Cajal. (b) De nombreuses structures de taille, de forme et de répartition variées ont alors été mises en évidence par cette technique. (c) La répartition des chromosomes dans l'espace nucléaire témoigne également d'une organisation particulière. Observés depuis longtemps sous forme condensée lors de la prophase (c1), les chromosomes ne sont pas répartis de façon uniforme dans l'espace nucléaire lors de l'interphase (c2). Ils occupent des zones aux frontières floues, appelées territoires chromosomiques, dont l'emplacement au sein du noyau et dont les positions respectives montre un certain caractère de reproductibilité. (a1-5 : (Swedlow et Lamond, 2001), b1 : (Misteli, 2001), b2-4 : (Spector, 2006), b5 : (Osborne *et al.*, 2004), c1-2 : (Cremer et Cremer, 2001)).

particulière et la présence d'obstacles provoque des propriétés de diffusion anormale[12] (Gorski *et al.*, 2006). Ceux-ci forment en effet, au sein de l'espace nucléaire, un réseau de cavités avec des tailles caractéristiques très variées, constituant alors un milieu poreux aux propriétés fractales[13] (Misteli, 2005; Bancaud *et al.*, 2009). On observe en effet qu'un corps nucléaire peut être plus ou moins perméable à différentes molécules en fonction de

[12]La diffusion anormale consiste en un mouvement d'une particule diffusive selon lequel la distance quadratique moyenne parcourue n'augmente pas de façon linéaire avec le temps (comme c'est le cas pour la diffusion normale). Cela ne signifie pas que la diffusion est plus ou moins rapide, mais simplement que ses propriétés sont différentes.

[13]Un objet fractal est une structure présentant des caractéristiques similaires à toutes les échelles (ou sur une grande plage de tailles). L'espace accessible à une particule dans un milieu poreux constitué d'un grand nombre d'obstacles disposés aléatoirement peut avoir naturellement une structure fractale.

leur taille (Handwerger *et al.*, 2005; Bancaud *et al.*, 2009). À une échelle supérieure, la forme des régions inter-chromatiniennes au sein des territoires chromosomiques, visualisée récemment en trois dimensions (Rouquette *et al.*, 2009), paraît susceptible de provoquer le même type de propriétés. Mais ces structures, vues comme des obstacles à la diffusion, ne sont en réalité pas figées et évoluent également dynamiquement. Cela provoque alors des mouvements singuliers reflétant à la fois la diffusion en milieu confiné de la particule et le déplacement à plus grande échelle de la structure. On peut retrouver ce type de mouvement à différents niveaux : du déplacement des molécules (Gorski *et al.*, 2006; Bancaud *et al.*, 2009) à celui du corps nucléaire (Görisch *et al.*, 2004; Shav-Tal *et al.*, 2006). Ces propriétés sont particulièrement intéressantes car les études théoriques de la diffusion en milieu poreux fractal montrent, non seulement que les propriétés de diffusion anormale influencent les temps de rencontre entre particules et les cinétiques de réactions (Saxton, 2001; Berry, 2002; Grima et Schnell, 2006; Echevería *et al.*, 2007; Guigas et Weiss, 2008), mais également que, lorsque les obstacles ne sont pas fixes, ces propriétés peuvent être très différentes (Grima et Schnell, 2006) et dépendent en particulier du type de déplacement des obstacles[14].

De plus, un grand nombre de protéines régulatrices, telles que les facteurs de transcription, sont capables d'interagir avec la chromatine et l'ADN et de s'y fixer brièvement et fréquemment. Le déplacement de ces molécules suit ainsi un processus qualifié de *hit-and-run* ou de *hopping* car elles alternent entre de courtes périodes d'immobilité et de diffusion rapide (figure I.3a2) (Phair *et al.*, 2004; Gorski *et al.*, 2006; Misteli, 2001, 2007). Ce mécanisme entraîne alors une diffusion apparente plus lente, mais augmente la capacité du régulateur à trouver son site de fixation. Les phénomènes de *sliding* libre sur le brin d'ADN, évoqués à la section 2.7, paraissent difficiles dans un contexte où l'ADN est très encombré par de nombreuses molécules (Gorski *et al.*, 2006). Mais des phénomènes de propagation le long du chromosome pourraient être à l'œuvre tels que le *sliding* assisté par le déplacement de la machinerie de transcription ou de réplication, ou l'assemblage progressif de proche en proche – qualifié de *oozing* – sur la chromatine (Talbert et Henikoff, 2006).

Ainsi, dans le noyau, les déplacements à l'échelle moléculaire témoignent de propriétés susceptibles de contraindre différents aspects des interactions entre molécules, en agissant sur les probabilités de rencontre, les temps de recherche, les durées d'interaction ... Mais à une échelle supérieure, ces propriétés peuvent également influencer le comportement des différentes structures, voire l'organisation globale du noyau.

4.3 Dynamique des corps nucléaires

Les corps nucléaires sont des structures très différentes de celles que l'on peut trouver dans le cytoplasme. Elles ont en effet la particularité d'être dépourvues de membrane, une caractéristique ouvrant alors beaucoup de questions quant à leur formation, leur maintien et leur disparition (Misteli, 2001, 2005, 2007; Matera *et al.*, 2009). De plus, même si ce sont des régions de l'espace où l'on trouve certaines espèces moléculaires en sur-concentration,

[14] H. Berry, communication personnelle.

elles ne sont pour autant pas particulièrement plus denses que le reste du noyau mais présentent simplement une composition différente. Les techniques de photobleaching, permettant de mesurer les mouvements des constituants de ces structures ainsi que la façon dont ils sont échangés dynamiquement avec l'extérieur (Shav-Tal *et al.*, 2004b), indiquent que les corps nucléaires sont très dynamiques et en renouvellement permanent : les molécules se déplacent rapidement au sein de ces structures et sont continuellement en échange avec l'extérieur du corps nucléaire (Phair et Misteli, 2000; Misteli, 2001, 2005, 2007; Shav-Tal *et al.*, 2006). En particulier, des expériences de FLIP[15] montrent que des protéines caractéristiques d'un corps nucléaire peuvent très facilement en sortir, passer dans d'autres corps nucléaires, puis y revenir et ce en un temps relativement court (Misteli, 2005, 2007).

Le nombre d'instances d'un type donné de corps nucléaire est très variable et dépend non seulement du corps nucléaire considéré, mais également du type cellulaire, de l'étape dans le cycle, des conditions de stress ... Par exemple, le nucléole est une structure que l'on trouve en très peu d'exemplaires et qui disparaît totalement lors de la mitose (Dundr *et al.*, 2000), alors que les speckles ou les usines à transcription sont très nombreux, la quantité de ces dernières allant de quelques centaines à quelques milliers (Sutherland et Bickmore, 2009; Osborne *et al.*, 2004). Plusieurs hypothèses concernant le processus d'apparition/disparition de ces structures ont été proposées et il semble, d'après les données expérimentales, que différents mécanismes puissent être à l'œuvre. Ces structures peuvent en effet aussi bien apparaître et disparaître par un processus de scission et d'agglomération (Eskiw *et al.*, 2003) que par un processus de nucléation et de dissolution (Handwerger *et al.*, 2002). Le processus de nucléation, consistant en la formation *de novo* d'un corps nucléaire par condensation locale de ses constituants, peut se produire en des emplacements aléatoires ou, dans certain cas, sur des séquences particulières d'ADN. Par exemple, le nucléole se forme sur les régions des chromosomes contenant des clusters de gènes d'ARN ribosomique (Misteli, 2005), ou encore, il semblerait que les séquences LCR (Locus Control Region) et certains promoteurs forts soient capables de nucléer une usine à transcription (Sutherland et Bickmore, 2009), augmentant alors localement la probabilité des gènes avoisinants d'être exprimés.

Ainsi, les propriétés dynamiques et structurelles des corps nucléaires sont très similaires à ce que l'on observe dans des systèmes multiphasiques liquides, c'est-à-dire des mélanges constitués d'espèces chimiques partiellement miscibles. On rencontre de tels systèmes dans des contextes très variés (Grillo, 2003; Scholten *et al.*, 2008) et les phénomènes de transition de phase, de nucléation, de coalescence, de peptisation ... font l'objet de la chimie des colloïdes et sont très étudiés – y compris dans le cas de biomolécules (Galkin *et al.*, 2002). Pour un système à deux composants A et B partiellement miscibles par exemple, une très légère augmentation de B dans une phase riche en A peut faire apparaître une seconde phase se traduisant par la nucléation de gouttelettes très concentrées en B. Dans un système biologique, on imagine aisément que l'apparition de telles structures permette la catalyse de certaines réactions (conduisant par exemple à l'expression de certains gènes) qui vont alors modifier la composition du milieu et influencer la formation/déformation

[15]Fluorescence Loss In Photobleaching : technique consistant à bleacher (faire perdre la caractère fluorescent de manière irréversible) de façon continue les molécules dans une région et d'observer la diminution de fluorescence dans une autre région. Cette diminution témoigne alors des échanges de matière entre les deux régions.

d'autres phases ... Toutes ces interactions réciproques entre les différentes structures, faisant intervenir leurs propriétés dynamiques et leurs fonctions, peuvent alors donner lieu à des comportements complexes auto-organisés dans lesquels l'expression génique joue un rôle central, à la fois comme cause et comme effet (Misteli, 2001, 2007).

4.4 Architecture nucléaire et expression génique

Une grande partie du processus d'expression génique a lieu au sein de cet espace structuré qu'est le noyau cellulaire. Comme le montrent de nombreuses observations expérimentales, il est couplé en de nombreux niveaux à cette organisation spatiale dynamique.

Localisation spatiale des gènes

La position d'un gène au sein du noyau cellulaire peut constituer un aspect essentiel de son expression. On observe en effet dans de nombreux cas que l'état d'activation d'un gène est corrélé au déplacement de son locus chromosomique (Shav-Tal *et al.*, 2006; Fraser et Bickmore, 2007; Misteli, 2007; Darzacq *et al.*, 2009). On a d'abord associé la position d'un gène à la périphérie du noyau à sa répression du fait de l'observation d'une grande quantité d'hétérochromatine dans ces régions. Réciproquement, une position centrale dans le noyau était associée à l'expression du gène. Mais, d'une part, la vision simple selon laquelle l'hétérochromatine correspondrait à un état réprimé des gènes tandis que l'euchromatine permettrait leur expression ne semble pas être toujours valide (Spector, 2004; Misteli, 2005), d'autre part, l'hétérochromatine ne se situe pas nécessairement à la périphérie du noyau et l'euchromatine au centre (Shav-Tal *et al.*, 2006; Fraser et Bickmore, 2007; Sexton *et al.*, 2007; Darzacq *et al.*, 2009). En effet, on observe dans de nombreux cas le phénomène inverse, à savoir l'activation d'un gène suite à son déplacement à la périphérie du noyau et *vice versa* (Lanctôt *et al.*, 2007; Darzacq *et al.*, 2009).

À un niveau de description plus fin, de nombreuses études ont montré une association entre l'expression d'un gène et la relocalisation de son locus hors du territoire chromosomique, ou même jusque dans un territoire chromosomique voisin (Fraser et Bickmore, 2007; Misteli, 2007). Ce mécanisme semble être relativement général, même si certains gènes peuvent toutefois être exprimés à l'intérieur de leur propre territoire chromosomique (Fraser et Bickmore, 2007; Misteli, 2007). Dans les régions inter-chromosomiques, la chromatine se décondense localement et s'intermêle avec l'ADN des chromosomes voisins (Branco et Pombo, 2006; Fraser et Bickmore, 2007). On peut alors observer une colocalisation de gènes du même chromosome même si ceux-ci sont relativement distants en séquence (interaction en *cis*) ou appartiennent à des chromosomes différents (interaction en *trans*[16]) (figure I.5b) (Shav-Tal *et al.*, 2006; Fraser et Bickmore, 2007; Misteli, 2007). Ce phénomène, qualifié de *gene kissing* (Lanctôt *et al.*, 2007), implique temporairement un couplage entre l'expression de différents gènes (Misteli, 2005; Sutherland et

[16]Les notions de *cis* et de *trans* peuvent avoir des significations assez différentes d'un contexte à l'autre. Par exemple, dans le cadre d'un réseau de régulation, une mutation en *cis* affecte la séquence d'un site de fixation pour un FT et donc (de façon directe) l'expression d'un seul gène, alors qu'une mutation en *trans* affecte la partie codante d'un FT et donc potentiellement tous les gènes que ce dernier régule.

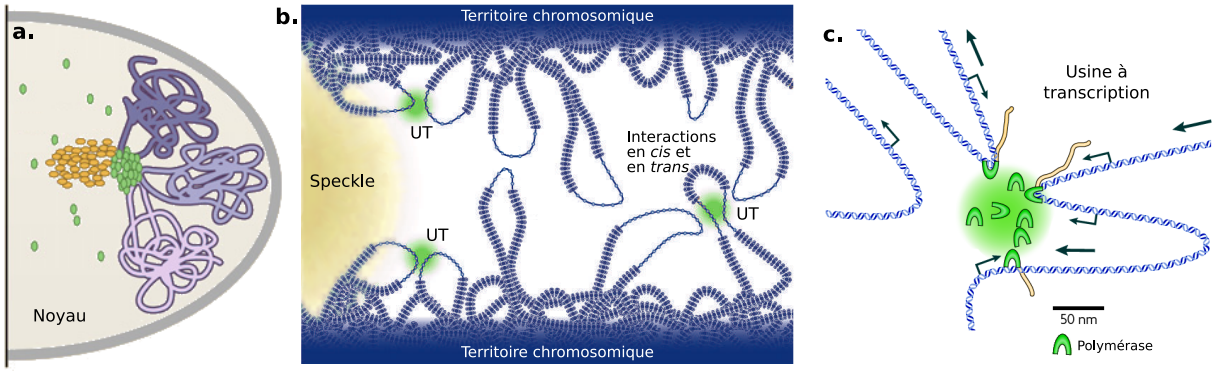


Fig. I.5: Organisation spatiale de l'expression génique. (a) Au sein du noyau des cellules eucaryotes, de nombreux gènes actifs pouvant appartenir à des chromosomes distincts colocalisent dans des zones actives de transcription (vert), hors des territoires chromosomiques (bleu). (b) Dans ces régions inter-chromosomiques où s'intermêle l'ADN des différents chromosomes, se trouvent différents corps nucléaires impliqués dans l'expression génique tels que les speckles (enrichis en enzymes de maturation de l'ARN : épissage, export, ...) ou les usines à transcription (UT, contenant en particulier l'ARN polymérase). Au sein de ces dernières peuvent colocaliser des gènes appartenant au même chromosome (interaction en *cis*) ou à deux chromosomes différents (interaction en *trans*). (c) Les contraintes mécaniques exercées sur l'ADN par l'activité de transcription et la quantité limitante de polymérase par usine à transcription sont alors susceptibles de provoquer des dépendances complexes entre l'expression de ces différents gènes, sur la seule base de leur proximité spatiale. (illustrations d'après Misteli (2007), Fraser et Bickmore (2007) et Sutherland et Bickmore (2009)).

Bickmore, 2009) et l'on peut alors s'attendre à certaines corrélations de leurs propriétés stochastiques. De plus, même si la colocalisation de deux gènes peut être relativement reproductible, elle n'est cependant pas observée systématiquement (Lanctôt *et al.*, 2007), montrant le caractère intrinsèquement probabiliste de ce phénomène.

Usines à transcription

Dans les régions inter-chromosomiques, les différents gènes semblent se rencontrer et être transcrits dans des *usines à transcription*. Il s'agit de très petites zones présentant une forte concentration en protéines impliquées dans la transcription, en particulier la polymérase (figures I.4b5 et I.5c) (Shav-Tal *et al.*, 2006; Fraser et Bickmore, 2007; Darzacq *et al.*, 2009; Sutherland et Bickmore, 2009). Il a longtemps été supposé que la transcription s'effectuait par le déplacement de la polymérase le long de l'ADN[17]. Il apparaît aujourd'hui que, dans de nombreux cas, c'est au contraire l'ADN qui se déplace tandis que la polymérase resterait plutôt fixe (Shav-Tal *et al.*, 2006; Sutherland et Bickmore, 2009). Cette dernière est en effet un moteur moléculaire suffisamment puissant pour faire se déplacer un locus chromosomique hors de sa position d'origine (Shav-Tal *et al.*, 2006). Ce changement de point de vue place alors la polymérase – et donc l'usine à transcription – comme un élément acteur du processus, déplaçant le matériel génétique du fait même de sa lecture. De fait, le déplacement d'un gène est non seulement nécessaire à

[17]Une vision probablement héritée du dogme central de la biologie moléculaire plaçant l'ADN au centre du fonctionnement de tout système biologique.

son expression (relocalisation hors du territoire chromosomique, recrutement dans l'usine à transcription), mais est également une conséquence de sa transcription. On peut alors facilement imaginer que les mouvements induits par l'expression d'un gène donné puissent influencer l'expression d'autres gènes, provoquant des phénomènes de régulation exclusivement spatiaux (Sutherland et Bickmore, 2009).

Les usines à transcription sont des structures très nombreuses mais très petites : on estime à 8 le nombre moyen de polymérases par usines à transcription (Sutherland et Bickmore, 2009). On peut alors s'attendre à ce que cette petite taille provoque certains phénomènes particuliers. En effet, si la présence d'une usine à transcription à proximité des plusieurs gènes augmente collectivement leur probabilité d'être transcrits, il y a probablement également des phénomènes de compétition entre gènes au sein de la même usine. Si un gène au promoteur fort monopolise l'essentiel des polymérases, l'expression des autres gènes sera alors réduite. De plus, la transcription déplaçant un gène hors de l'usine à transcription, on remarque que pour que celui-ci soit transcrit de nombreuses fois à la suite, il peut être nécessaire qu'il forme une boucle pour revenir dans l'usine à transcription (Sutherland et Bickmore, 2009). Tous ces phénomènes putatifs sont alors susceptibles d'avoir une influence particulière sur le caractère stochastique de l'expression génique, avec de portentielles signatures reconnaissables, en particulier dans la corrélation entre l'activité des différents gènes partageant la même usine à transcription.

Maturation et export des transcrits

Une fois transcrit, et avant de pouvoir être traduit en protéines dans le cytoplasme, un ARN doit subir un certain nombre de modifications et être exporté hors du noyau (Darzacq *et al.*, 2005). Comme nous l'avons souligné dans la section 2.6, le délai qu'implique ces différents processus – ou plus précisément, la distribution de ce délai – peut jouer significativement sur les propriétés stochastiques de l'expression du gène. L'organisation spatiale de ces processus au sein du noyau est alors susceptible d'influencer, voire de réguler ces propriétés.

On trouve souvent à proximité des usines à transcription des corps nucléaires appelés les nuclear speckles (parfois également dénommés inter-chromatin granule clusters ; figure I.4b2). Cette structure, sensiblement plus grosse que les usines à transcription, contient des protéines impliquées dans la maturation et l'export de l'ARN (facteurs d'épissage, ...) (Lamond et Spector, 2003; Shav-Tal *et al.*, 2006; Fraser et Bickmore, 2007; Sutherland et Bickmore, 2009). On observe en effet que les speckles sont fréquemment associés à des régions activement transcrites (Sutherland et Bickmore, 2009) et que les étapes de maturation de l'ARN sont souvent couplées à sa transcription (Darzacq *et al.*, 2005, 2009). Elle ne semble cependant pas être le lieu même de maturation des transcrits et pourraient essentiellement servir de zone de stockage ou de séquestration de différentes enzymes, en régulant spatialement leur disponibilité dans le noyau (Shav-Tal *et al.*, 2006; Fraser et Bickmore, 2007). Ainsi, en agissant sur le processus de maturation et d'export des transcrits, ces structures influencent le délai entre la synthèse d'un ARN et le moment où il peut être traduit hors du noyau. En modifiant, plus précisément, la distribution de ce délai, elles pourraient donc de constituer un moyen de régulation des propriétés stochas-

tiques de l'expression du gène (cf section 2.6). En effet, certaines données expérimentales suggèrent des mécanismes de rétention régulée des transcrits (Gorski *et al.*, 2006; Shav-Tal *et al.*, 2006).

Le délai d'export des transcrits dépend également fortement de leur mode de déplacement depuis le site de transcription jusqu'aux pores nucléaires. Ce mécanisme est particulièrement important pour les propriétés stochastiques de l'expression génique puisqu'il joue également sur la distribution du délai d'export (cf section 2.6). En effet, le mode de déplacement passif ou actif d'un particule diffusive peut conduire à des distributions de temps de capture très différents (Cranck, 1975; Berg, 1993). Mais, s'il est clair que ce délai peut être relativement court en moyenne dans de nombreux cas (quelques minutes; Gorski *et al.*, 2006), ce mécanisme reste encore mal compris. En effet, différentes études expérimentales rapportent des phénomènes contradictoires : alors que de nombreuses mesures suggèrent depuis longtemps que les transcrits se déplacent simplement par diffusion et qu'aucun déplacement actif des transcrits n'ait jamais été observé, on constate pourtant une modification significative de leurs mouvements dans le cas de mutants d'actine et de myosine[18] ou d'une déplétion en ATP (Shav-Tal *et al.*, 2004a, 2006; Gorski *et al.*, 2006). Mais il semblerait que ce phénomène soit plutôt dû à une modification des propriétés dynamiques de la chromatine se répercutant indirectement sur le déplacement des transcrits qu'à un mécanisme de transport dirigé de ces derniers (Shav-Tal *et al.*, 2004a; Gorski *et al.*, 2006). Un autre mécanisme pouvant jouer un rôle fondamental sur ce délai est la localisation du gène par rapport au pore nucléaire. En effet, certains gènes sont exprimés à proximité de l'enveloppe nucléaire, voire en interaction physique avec un pore nucléaire, permettant alors un export rapide du messager vers le cytoplasme (Sexton *et al.*, 2007). Il a toutefois été observé que des transcrits synthétisés près de la membrane nucléaire n'étaient pas nécessairement exportés par les pores les plus proches (Gorski *et al.*, 2006).

4.5 Le noyau, un système auto-organisé ?

Les mécanismes de contrôle et de régulation de l'organisation globale du noyau reste très mal connus. Deux points de vue différents coexistent dans la littérature (Fraser et Bickmore, 2007; Misteli, 2005, 2007; Matera *et al.*, 2009; Sutherland et Bickmore, 2009). Certains auteurs voient par exemple l'organisation en territoires chromosomiques et le déplacement des gènes en corrélation avec leur activité comme une simple conséquence du processus d'expression n'ayant pas d'impact fonctionnel. En effet, dans de nombreux cas si l'on perturbe cette organisation, le système peut quand même continuer à fonctionner, mais généralement moins efficacement (Misteli, 2005; Sutherland et Bickmore, 2009). D'autres auteurs considèrent au contraire que cette organisation et cette dynamique sont actrices dans la façon dont le génome est exprimé (Misteli, 2001, 2005, 2007; Fraser et Bickmore, 2007; Sutherland et Bickmore, 2009). Le principe, selon lequel l'organisation spatiale influence l'expression génique et les réactions enzymatiques, elles-mêmes modifiant cette organisation en retour, serait alors à même de provoquer des phénomènes d'auto-organisation complexes, sous-tendant l'ensemble du fonctionnement et de la structuration du noyau.

[18]Des protéines généralement impliquées dans des mécanismes de transport actif.

5 Dynamique de la régulation transcriptionnelle

Les événements moléculaires au niveau du promoteur d'un gène peuvent être une source significative de stochasticité dans son expression. Comme nous l'avons évoqué précédemment (cf section 2.4), la plupart des modèles d'expression stochastique prenant en compte la dynamique du promoteur considèrent celui-ci comme un système à deux états, basculant de l'un à l'autre avec des délais dont la distribution est exponentielle. Il existe cependant une littérature expérimentale très vaste sur la dynamique de la régulation transcriptionnelle, d'après laquelle il apparaît clairement que le promoteur est un système bien plus complexe que ne le suppose cette description. Nous explorons ici cette littérature en décrivant tout d'abord les caractéristiques générales des promoteurs eucaryotes, puis en nous focalisant plus particulièrement sur leur dynamique à différentes échelles temporelles.

5.1 Promoteurs eucaryotes

Le promoteur d'un gène est une séquence d'ADN sur laquelle s'associent de nombreuses molécules dont les interactions régulent l'expression du gène en affectant le recrutement de la machinerie transcriptionnelle et sa capacité à initier la transcription du gène. La notion de promoteur n'a cependant pas de définition claire et peut aussi bien désigner une région de quelques centaines de nucléotides autour du site d'initiation de la transcription (TSS ; transcription start site), appelé *core promoter* et contenant quelques motifs caractéristiques (tels que la TATA box, figure I.6b2), qu'une région de plusieurs milliers de nucléotides contenant des sites d'assemblage d'autres complexes tels que les enhancers (Wray *et al.*, 2003). De plus, ce terme désigne souvent à la fois la séquence d'ADN et les structures macromoléculaires qui lui sont associées dont, en particulier, la chromatine. Les figures I.6b1 et I.6b2 représentent le core promoter du gène pS2, régulé par le récepteur des œstrogènes (ER) chez l'homme (Métivier *et al.*, 2003; Reid *et al.*, 2009), sur lequel nous reviendrons dans ce chapitre et dans le chapitre III.

Chromatine, nucléosomes, histones

L'ADN des cellules eucaryotes est organisé au sein d'une structure, appelée la *chromatine*, dont la conformation tridimensionnelle, pouvant être très variée, est régulée par l'action de nombreuses protéines (figure I.6a1). Elle est composée d'unités, les *nucléosomes* (formés chacun de quatre paires de protéines : les *histones* H2A, H2B, H3 et H4), autour desquelles l'ADN est enroulé en moyenne 1.7 fois (figure I.6a2) et stabilisé par d'autres protéines (les histones H1 et H5) placées entre les nucléosomes (Fischle *et al.*, 2003; Mellor, 2006; Li *et al.*, 2007; Jiang et Pugh, 2009). L'état de compaction de la chromatine au niveau d'un promoteur influence la capacité des différentes protéines à s'associer à leurs sites de fixation sur l'ADN et constitue donc un facteur déterminant dans la régulation de l'expression génique. La conformation de la chromatine dépend cependant de nombreux autres facteurs et est elle-même régulée localement par l'action de divers complexes (Perissi et Rosenfeld, 2005; Mellor, 2006; Li *et al.*, 2007; Henikoff, 2008; Cairns, 2009; Segal et Widom, 2009; Jiang et Pugh, 2009; Raveh-Sadka *et al.*, 2009).

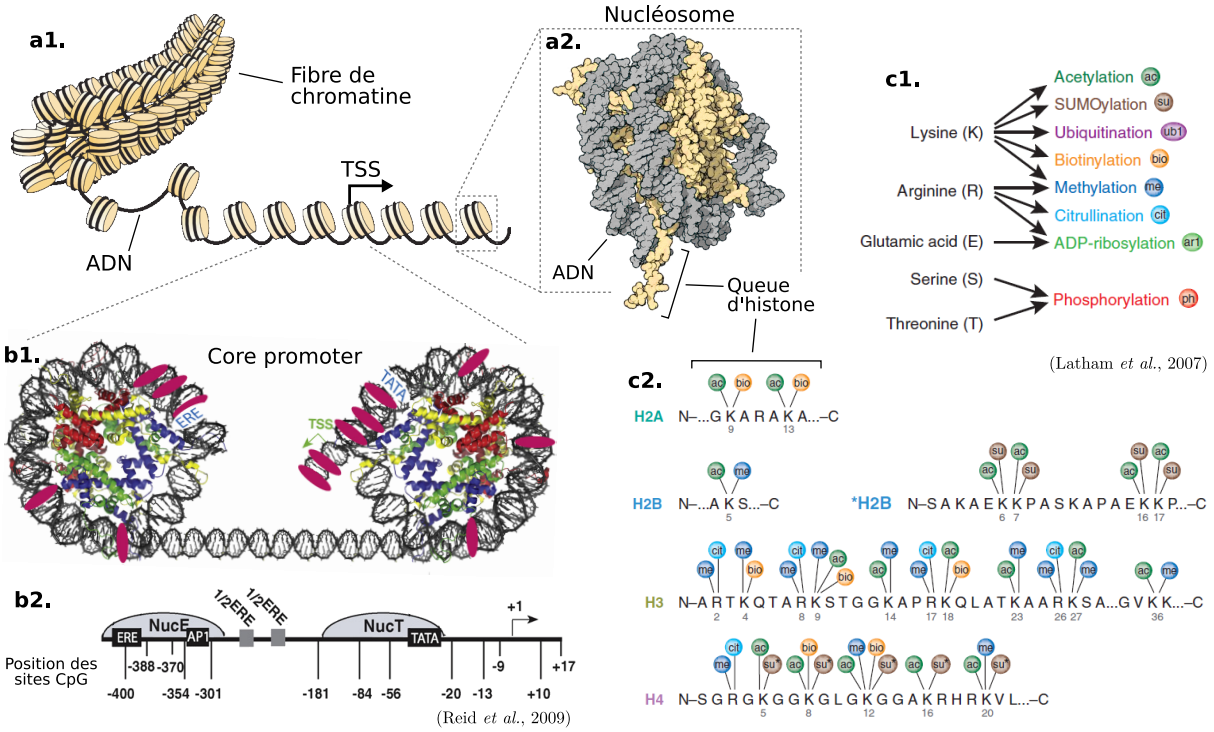


Fig. I.6: Rôle de la chromatine dans la régulation transcriptionnelle. (a1) La chromatine est une structure au sein de laquelle se trouve organisé l'ADN des cellules eucaryotes. Différents aspects de sa conformation, par exemple son niveau de compaction, peuvent affecter la régulation génique. (a2) Cette structure est composée d'unités, appelées les nucléosomes, constituées chacune de 8 protéines appelées histones. (b1) Le promoteur d'un gène (ici le gène pS2 (Reid *et al.*, 2009)) est enroulé autour de plusieurs nucléosomes. (b2) Leurs positions peut alors influencer la capacité de différentes molécules de s'associer à leurs sites de fixation (ERE : site de fixation de ER, TATA box : motif sur lequel s'assemble le PIC ; cf figure I.7). De plus, plusieurs sites CpG sont présents sur le promoteur et leur état de méthylation participe à la régulation du gène. (c1) Chacune des 8 histones composant un nucléosome possède une queue N-terminale pouvant subir des modifications covalentes post-traductionnelles (les différentes modifications étant spécifiques à certains acides aminés) ouvrant alors la possibilité à de nombreuses combinatoires (c2). (a2 : PDB, entrée 1AOI, molecule of the month 07/2000. b1, b2 : d'après Reid *et al.* (2009). c1, c2 : d'après Latham et Dent (2007)).

Chacune des huit histones composant le nucléosome possède une longue queue N-terminale dont certains acides aminés sont sujets à différentes modifications covalentes telles que l'acétylation, la méthylation, la phosphorylation, ... (figures I.6c1 et I.6c2). Ces modifications, effectuées par des complexes spécifiques appelés collectivement *enzymes de modification d'histones*, altèrent les propriétés d'interaction des nucléosomes entre eux, avec l'ADN et avec les protéines pouvant s'y associer (Fischle *et al.*, 2003; Verdone *et al.*, 2005; Mellor, 2006; Li *et al.*, 2007; Kouzarides, 2007; Latham et Dent, 2007). Il est par exemple souvent admis que l'activation d'un gène passe par le recrutement d'une histone acétyltransférase (HAT) au niveau du promoteur et que l'acétylation des histones qui en résulte provoque une décompaction des nucléosomes permettant alors l'association d'autres molécules, même si cette vision simple ne s'applique pas systématiquement (Verdone *et al.*,

2005). L'ADN subit également des modifications covalentes – à savoir la méthylation des sites CpG[19] (figure I.6b2) – qui influencent l'expression génique (Henikoff, 2008; Cairns, 2009).

Lorsque la chromatine est ouverte, la position d'un nucléosome le long de l'ADN est à la fois déterminée par la séquence (interaction nucléosome-ADN et contraintes mécaniques de déformation du brin) et par des phénomènes d'exclusion stérique avec les nucléosomes voisins (Jiang et Pugh, 2009; Segal et Widom, 2009; Cairns, 2009; Raveh-Sadka *et al.*, 2009). La contribution relative de ces deux mécanismes est sujet à débat et il semblerait que dans de nombreux cas un positionnement fortement séquence-dépendant au niveau du promoteur et du terminateur du gène structure ensuite la répartition des nucléosomes sur la partie codante du gène essentiellement via un principe d'interaction stérique (Chevereau *et al.*, 2009; Jiang et Pugh, 2009; Raveh-Sadka *et al.*, 2009). De plus, différents *complexes de remodelage chromatinien* sont capables de modifier cette organisation au niveau du promoteur par des mécanismes ATP-dépendants, en expulsant ou en déplaçant un nucléosome par dessus une barrière énergétique (Mellor, 2006; Henikoff, 2008; Jiang et Pugh, 2009; Segal et Widom, 2009; Cairns, 2009). Ces remodelages peuvent alors mettre à jour ou occulter des sites de fixation pour d'autres molécules. Par exemple, dans l'activation transcriptionnelle du gène pS2 (figure I.6b2), un remodelage du nucléosome NucT par le complexe SWI/SNF permet de mettre à jour la TATA box, site d'assemblage de la machinerie transcriptionnelle (Métivier *et al.*, 2003).

Les modifications conformationnelles et covalentes de la chromatine et de l'ADN influencent donc la capacité des différentes molécules à s'associer au promoteur, et en particulier des enzymes responsables de ces mêmes modifications. Elles peuvent ainsi constituer des marques potentiellement durables qui sont à même de modifier l'activité de transcription d'un gène. Ces marques sont qualifiées d'*épigénétiques* car elles constituent une information (potentiellement héréditaire) affectant l'expression des gènes sans toutefois être portées par la séquence d'ADN. Il existe par ailleurs plusieurs variantes d'histones (H3.3, H2A.Z, ...) qui sont impliquées dans l'activité de nombreux gènes et dont la présence au sein d'un nucléosome modifie les propriétés de ce dernier (Li *et al.*, 2007; Henikoff, 2008; Orsi *et al.*, 2009; Cairns, 2009; Jiang et Pugh, 2009). On observe de plus de nombreux aspects combinatoires dans l'influence mutuelle entre ces facteurs épigénétiques (Fischle *et al.*, 2003; Mellor, 2006; Latham et Dent, 2007). On parle alors d'un *code des histones* susceptible d'encoder de nombreux comportements différents (Jenuwein et Allis, 2001; Wang *et al.*, 2008), même si la spécificité de ce code, voire la notion même d'un code, reste sujet à débat (Henikoff, 2005; Dion *et al.*, 2005).

Facteurs de transcription, cofacteurs et complexe de pré-initiation

Les molécules qui s'associent au promoteur sont de nature très variées (figure I.7). On distingue souvent des catégories telles que les facteurs de transcription généraux et spécifiques, les coactivateurs, les corépresseurs, ... (Thomas et Chiang, 2006; Latham et Dent, 2007).

[19]Un site CpG est une suite de deux nucléotides (C et G) dans la séquence d'ADN, par opposition à

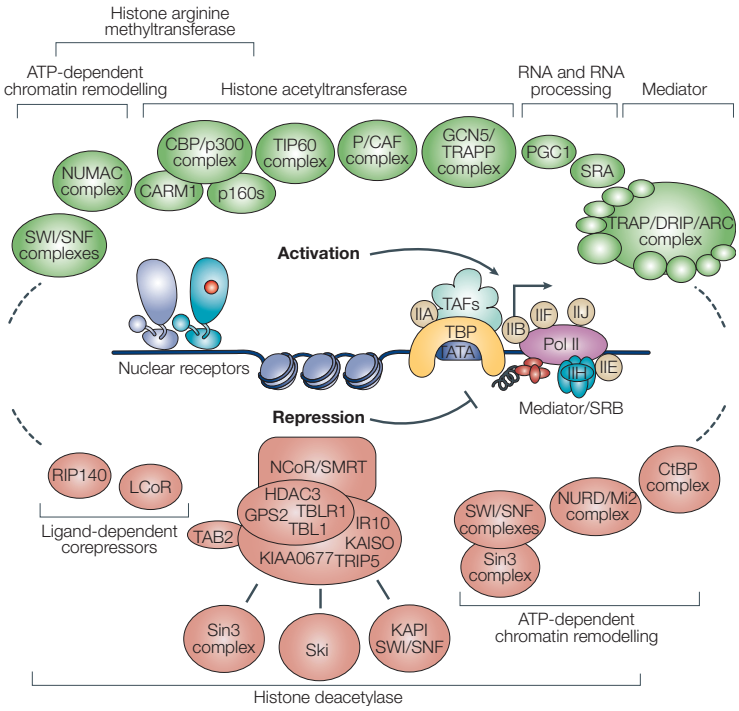


Fig. I.7: Molécules et complexes impliqués dans la régulation transcriptionnelle. De nombreux complexes, appelés cofacteurs, ont des activités très variées sur le promoteur (remodelage de la chromatine, modification covalente des histones) pouvant aussi bien être activatrices que répressives. Leur recrutement et leur activité influencent et sont influencés par des FT spécifiques au gène considéré, tels que les récepteurs nucléaires (GR ou ER par exemple). La formation du complexe de pré-initiation (PIC) contenant divers complexes dont les FT généraux (IIA, IIB, TBP, ...) et l'ARN polymérase est une étape critique pour l'initiation de la transcription et donc pour l'expression du gène. (d'après Perissi et Rosenfeld (2005))

Les facteurs de transcription (FT) sont des molécules qui s'associent au promoteur et facilitent le recrutement d'autres molécules. Les récepteurs nucléaires sont un exemple de FT spécifiques qui régulent l'activité de certains gènes particuliers. À l'inverse, les FT généraux (IIA, IIB, TBP, ...) participent à l'activation de presque tous les gènes de classe II[20] et forment, avec d'autres complexes dont la polymérase, le complexe de pré-initiation (PIC), nécessaire pour que le gène soit transcrit. Les cofacteurs sont des molécules ou des complexes pouvant avoir des rôles très diverses (figure I.7). Par exemple les enzymes de modification d'histones et les complexes ATP-dépendants de remodelage chromatinien évoqués précédemment en font partie.

Notons toutefois que cette classification peut être floue car élaborée à partir de mécanismes de fonctionnement supposés à un instant donné et qui ont parfois été remis en cause par la suite. De plus, certaines molécules peuvent en réalité avoir plusieurs fonctions, rentrant alors dans plusieurs catégories (figure I.7).

5.2 Mobilité des molécules sur le promoteur

Pendant de nombreuses années, les études de la régulation de la transcription ont donné une vision du promoteur correspondant à un système relativement statique et ordonné où les molécules s'associent de façon stable sur un site de fixation et selon des schémas

la paire de nucléotides complémentaires CG. Le « p » représente le phosphate liant les deux nucléotides.

[20]Gènes transcrits par l'ARN polymérase II, représentant la majeure partie des gènes chez les eucaryotes.

prédéfinis. Les méthodes d'observation modernes basées sur des techniques de microscopie à fluorescence et de photobleaching sur cellules uniques (FRAP, FRET) ainsi que sur des techniques de cinétique d'immunoprécipitation de chromatine (cinétique de ChIP), permettant désormais d'étudier les aspects dynamiques de ces structures de différents points de vue (cf Box 2). Au cours des dernières années, ces outils ont révélé que le promoteur et les complexes de régulation de la transcription sont en réalité des structures très dynamiques, démontrant de comportements singuliers à des échelles temporelles très variées basés sur le comportement stochastique des molécules.

Box 2 : Techniques expérimentales pour l'étude de la dynamique du promoteur

- Le **FRAP** (*Fluorescence Recovery After Photobleaching*) est une technique de vidéo-microscopie consistant à estimer la mobilité d'une protéine d'intérêt rendue fluorescente en observant la cinétique de récupération de la fluorescence dans une zone du milieu intra-cellulaire préalablement photoblanchie (le photoblanchiment est la perte irréversible des propriétés de fluorescence d'une protéine suite à un pulse d'excitation). Cette cinétique de récupération peut refléter à la fois la diffusion de la molécule d'intérêt et son association/dissociation sur des structures (considérées fixes) telles que la chromatine (Shav-Tal *et al.*, 2004b; Sprague *et al.*, 2004; Sprague et McNally, 2005). Ces deux composantes peuvent souvent être identifiées en faisant varier la taille de la zone de blanchiment. En régime non-lié à la diffusion, cette récupération présente une ou plusieurs composantes exponentielles, décrivant chacune un état d'association de la molécule d'intérêt et dont on peut estimer les durées moyennes par un fit multi-exponentiel (Phair *et al.*, 2004; Darzacq *et al.*, 2007).
- Le **FRET** (*Förster Resonance Energy Transfer* ou *Fluorescence Resonance Energy Transfer*) est une technique de vidéo-microscopie à florescence permettant de mesurer la proximité spatiale entre deux espèces moléculaires, un donneur et un accepteur (Truong et Ikura, 2001). Plus les deux molécules sont proches, plus le donneur transmet son excitation à l'accepteur. Ainsi la fluorescence du donneur augmente (et celle de l'accepteur diminue) avec la distance entre les deux molécules. Dans le cadre de l'étude la dynamique des complexes de régulation, cette technique permet d'observer au cours du temps les changements conformationnels du promoteur (Li *et al.*, 2005; Poirier *et al.*, 2009; Kelbauskas *et al.*, 2009). On calcule en particulier souvent l'auto-covariance du signal du donneur pour identifier les temps caractéristiques des différents changements conformationnels.
- La **cinétique de ChIP** (*Chromatin ImmunoPrecipitation*) permet de mesurer, au cours du temps, le taux d'occupation, par différentes molécules, du promoteur d'un gène d'intérêt lors de son activation transcriptionnelle (Shang *et al.*, 2000; Métivier *et al.*, 2003; Degenhardt *et al.*, 2009). Cette technique se base sur la synchronisation d'une population de cellules à l'intant 0 de la cinétique (par l'ajout ou le retrait d'une substance dans le milieu) et sur la réalisation, à intervalles réguliers (par exemple 5 min pour Métivier *et al.* (2003)) et pour chaque anti-corps utilisé, d'une immunoprécipitation de la chromatine. Les étapes essentielles sont alors de fixer une fraction de la population de cellules au formaldéhyde (provoquant des liaisons covalentes en particulier entre l'ADN et les protéines s'y trouvant associées), de fragmenter leur ADN, d'isoler les complexes protéine-ADN reconnus par l'anti-corps utilisé, puis de quantifier parmi l'ADN ainsi récolté la quantité correspondant au promoteur d'intérêt, en l'amplifiant par une technique de PCR. Des données de ChIP sont toujours renormalisées par la valeur obtenue en l'absence d'anti-corps, appelé l'*input*. Une expérience de Re-ChIP consiste à utiliser successivement deux anti-corps afin de ne garder que les complexes protéine-ADN comprenant les deux protéines d'intérêt.

Dynamiques d'association/dissociation

Les expériences de photobleaching telles que le FRAP, permettent de mesurer le temps de résidence des molécules sur leurs sites de fixation (figure I.8 et Box 2). Lorsque le FRAP est effectué sur une zone quelconque du noyau, on observe souvent une composante lente et une composante rapide dans la récupération de la fluorescence, la première étant interprétée comme des associations spécifiques de la molécule d'intérêt sur un site cible et la seconde comme des associations non-spécifiques sur la chromatine (Sprague *et al.*, 2004; Phair *et al.*, 2004). En effet, lorsque le FRAP est effectué sur une zone riche en sites de fixation de cette molécule, la fraction lente est prépondérante (Karpova *et al.*, 2008; Johnson *et al.*, 2008; Hager *et al.*, 2009). Pour de très nombreuses molécules, la fraction « lente » indique des associations particulièrement transitoires. Le temps de résidence moyen dépend de nombreux facteurs (la molécule considérée, le promoteur, la présence de cofacteurs ou de ligands, l'état actif ou non du complexe, les conditions expérimentales, ...), mais s'étend typiquement de la dizaine de secondes à la minute[21] pour de nombreuses molécules (FT, cofacteurs, complexes de remodelage chromatinien, sous-unités de la polymérase, ...) (McNally *et al.*, 2000; Stenoien *et al.*, 2001; Phair *et al.*, 2004; Becker *et al.*, 2002; Hager *et al.*, 2002; Dundr *et al.*, 2002; Stavreva *et al.*, 2004; Elbi *et al.*, 2004; Agresti *et al.*, 2005; Rayasam *et al.*, 2005; Sharp *et al.*, 2006; Misteli, 2007; Darzacq *et al.*, 2007, 2009; Karpova *et al.*, 2004, 2008; Johnson *et al.*, 2008; Gorski *et al.*, 2008). Certaines molécules telles que les histones sont quant à elles associées à l'ADN pendant des périodes sensiblement plus longues (Hager *et al.*, 2002; Phair *et al.*, 2004; Kimura, 2005; Shav-Tal *et al.*, 2006; Gorski *et al.*, 2006; Wiesmeijer *et al.*, 2008). Le taux de renouvellement des histones est assez variable et peut être régulé – il dépend de la séquence d'ADN, du type d'histone, de la présence de certains complexes de remodelage, ... (Cairns, 2009; Henikoff, 2008). Leur durée de résidence est typiquement de l'ordre de plusieurs minutes ou dizaines de minutes.

Changements conformationnels spontanés

Alors que certaines modifications conformationnelles de la chromatine requièrent l'action spécifique de complexes de remodelage (passage au dessus d'une barrière énergétique), d'autres peuvent se produire spontanément du fait de l'agitation thermique du milieu (déplacements locaux dans un bassin énergétique). Différentes études, basées entre autres sur des techniques de FRET (cf Box 2), montrent que cette dynamique peut être également très rapide : par exemple quelques dizaines de millisecondes pour l'enroulement/déroulement de l'ADN autour d'un nucléosome (Li *et al.*, 2005; Kelbauskas *et al.*, 2009) ou quelques secondes pour les modifications les plus lentes observées dans la compaction/décompaction spontanée d'une fibre de chromatine (Poirier *et al.*, 2009). Ces mouvements peuvent être plus ou moins contraints d'une situation à l'autre et contrôlent alors l'accessibilité de certains sites de fixations (Métivier *et al.*, 2003; Nagaich *et al.*, 2004;

[21] Notons que certains auteurs utilisent le temps de demi-récupération $t_{0.5}$, d'autres le temps pour atteindre par exemple 90% de la récupération finale $t_{0.9}$ ou d'autres encore le temps caractéristique de l'exponentielle t_e (obtenue par un fit). La durée moyenne de résidence de la molécule correspond à t_e et doit être corrigée dans les deux autres cas : $t_e = t_{0.5}/\log 2 \simeq 1.44\, t_{0.5}$ et $t_e = t_{0.9}/\log(1/0.1) \simeq 0.434\, t_{0.9}$.

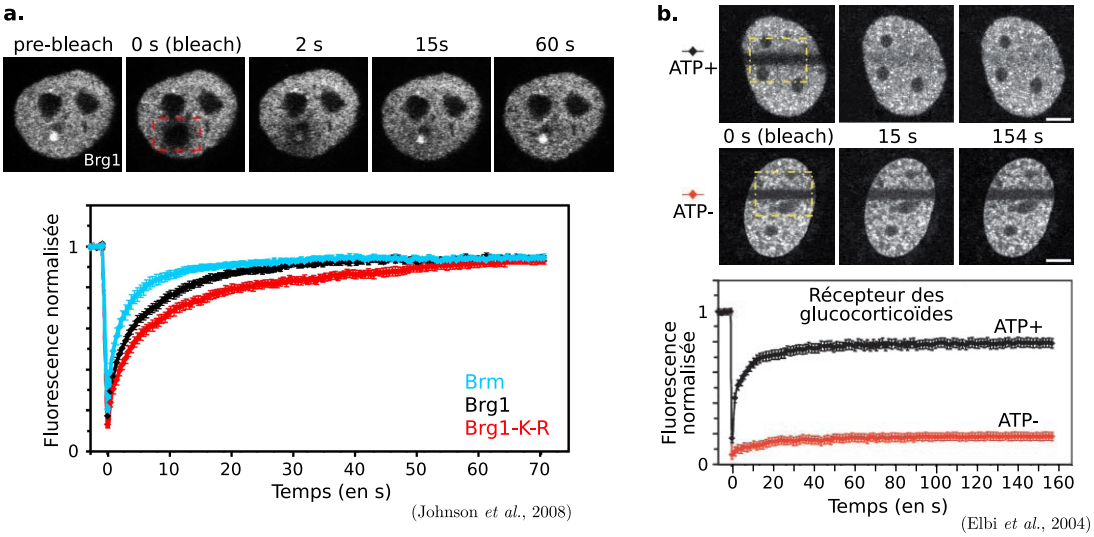


Fig. I.8: Mesure expérimentale de la mobilité des molécules sur le promoteur. (a) La récupération de fluorescence observée après le photoblanchiment (photobleaching) d'une région du noyau contenant une grande quantité de sites de fixation pour une molécule donnée rendue fluorescente (FRAP, cf Box 2) reflète le temps de résidence de cette dernière. Comme de nombreuses molécules, Brm et Brg1 (sous-unités du complexe de remodelage SWI/SNF) ont des temps de résidence très courts, de l'ordre de quelques secondes. La protéine Brg1-K-R, un mutant ATPase-déficient de Brg1, montre une récupération de fluorescence différente, présentant une fraction lente (cf Box 2) et suggérant donc un rôle actif de l'activité de remodelage sur la mobilité de ces molécules (d'après Johnson *et al.*, 2008). (b) De façon beaucoup plus flagrante, la mobilité des molécules sur la chromatine est très fortement altérée par l'inhibition de toute activité de remodelage de la chromatine aussi bien par à l'absence d'ATP (présenté ici pour le récepteur des glucocorticoïdes (GR) ; d'après Elbi *et al.* (2004)) que par l'absence des complexes de remodelage eux-mêmes.

Perissi et Rosenfeld, 2005; Segal et Widom, 2009; Raveh-Sadka *et al.*, 2009).

La mobilité, un processus actif

Plusieurs études expérimentales ont montré que la mobilité de nombreuses molécules était liée aux mécanismes de remodelage de la chromatine (aussi bien par les complexes ATP-dépendants que par les chaperonnes d'histones). En effet, la cinétique de récupération des courbes de FRAP est significativement ralentie lorsque la cellule est déplétée en ATP ou lorsque les complexes de remodelage son absents (Fletcher *et al.*, 2002; Elbi *et al.*, 2004; Karpova *et al.*, 2004; Rayasam *et al.*, 2005; Agresti *et al.*, 2005). La figure I.8b montre un exemple où la mobilité du récepteur des glucocorticoïdes (GR) est très largement affectée par la présence d'ATP. Ce phénomène explique pourquoi les mesures obtenues pendant des années par des techniques *in vitro* (Perlmann *et al.*, 1990) – et donc en l'absence des complexes de remodelage – ont rapporté des temps de résidence beaucoup plus longs que les techniques modernes de microscopie sur cellules vivantes. La mobilité des molécules sur la chromatine apparaît alors comme un phénomène actif et assisté par la machinerie de régulation de la transcription.

Ainsi, les complexes présents sur un promoteur sont des structures très dynamiques, en perpétuel renouvellement, et fortement soumises au comportement stochastique des molécules. Mais ces phénomènes probabilistes sont cependant contraints par des mécanismes d'interactions mutuelles et combinatoires à de nombreux niveaux.

5.3 Interactions combinatoires

Les différentes molécules s'associant au promoteur peuvent interagir entre elles et modifier mutuellement leurs propriétés dynamiques. En effet, dans de nombreux cas, on observe que l'affinité d'une molécule donnée à son site de fixation dépend de la présence de certaines autres molécules, et ce aussi bien chez les procaryotes (Ackers *et al.*, 1982; Scholl et Nixon, 1996; Dodd *et al.*, 2004) que chez les eucaryotes (Adams et Workman, 1995; Tanaka, 1996; Stenoien *et al.*, 2001; Miller et Widom, 2003; Agresti *et al.*, 2005; Rayasam *et al.*, 2005). Ces mécanismes de coopération et de compétition entre molécules sont souvent décrits en termes d'énergie d'interaction (Ackers *et al.*, 1982; Scholl et Nixon, 1996; Lehman *et al.*, 1998; Bintu *et al.*, 2005a) indiquant à quel point le contact physique entre plusieurs molécules rend plus ou moins favorable leur présence simultanée sur le promoteur. Ces phénomènes combinatoires agissent également sur les aspects dynamiques. On observe en effet non seulement que l'activité de remodelage, nécessitant la présence de certains complexes sur le promoteur, influence la mobilité des molécules des FT (figure I.8b), mais également que le temps de résidence des récepteurs nucléaires[22] dépend par exemple de la présence de leur ligand (Stenoien *et al.*, 2001; Sharp *et al.*, 2006). L'affinité du ligand avec le récepteur modifie également significativement le temps de résidence de ce dernier sur le promoteur (Rayasam *et al.*, 2005; Darzacq *et al.*, 2009).

Par ailleurs, comme nous l'avons évoqué précédemment, les modifications conformationnelles et covalentes de la chromatine permettent d'importantes propriétés combinatoires, conceptualisées par la notion de code des histones. Par exemple, l'acétylation d'un site d'une queue d'histone est effectuée par une HAT, et nécessite donc le recrutement sur le promoteur des différentes sous-unités de cette enzyme, lui-même assisté par d'autres molécules telles que des FT. Ainsi, la combinatoire des molécules sur le promoteur affecte le status épigénétique de la chromatine. Mais ce dernier est également capable d'influencer en retour l'affinité des molécules sur le promoteur et leurs potentielles activités enzymatiques ou de remodelage.

Ainsi, comme nous l'avons décrit dans la section 2.5, le rapport entre les échelles temporelles de dégradation des transcrits ou des protéines et celles des événements au niveau du promoteur détermine à quel point ces derniers sont en mesure d'influencer la variabilité du niveau d'expression (figure I.2e). On peut ainsi s'attendre à ce que le caractère stochastique des associations/dissociations des molécules sur le promoteur soit masqué par leur rapidité. Cependant, les promoteurs eucaryotes peuvent témoigner d'une dynamique lente, à la base de laquelle se trouvent ces événements stochastiques rapides et où les contraintes d'interactions combinatoires jouent un rôle central.

[22]Les récepteurs nucléaires sont des facteurs de transcription eucaryotes dont l'action est régulée par l'association d'une petite molécule appelée ligand.

5.4 Dynamique lente du promoteur

L'immunoprécipitation de la chromatine (ChIP) est une technique expérimentale permettant de mesurer, pour une zone d'intérêt du génome telle qu'un promoteur par exemple, le taux d'occupation par différentes molécules et des modifications conformationnelles et covalentes de la chromatine et de l'ADN au sein d'une population de cellules. Réaliser une cinétique de ChIP après activation du gène permet alors de suivre l'évolution au cours du temps de la probabilité de présence des différentes molécules et des modifications de la chromatine (cf Box 2).

Oscillations et activité périodique

Comme l'illustre la figure I.9a, des cinétiques de ChIP, réalisées sur le promoteur du gène pS2 (figure I.6b) régulé par le récepteur des œstrogènes (ER) dans des cellules humaines de cancer du sein, révèlent une activité périodique lente très caractéristique (Shang *et al.*, 2000; Métivier *et al.*, 2003). Chaque molécule et chaque modification de la chromatine témoigne en effet d'un profil spécifique, oscillant avec une période de 40 min. Ils décrivent une dynamique cyclique des différents événements dont l'ordre précis peut être observé sur des cinétiques à une résolution fine de 1 min (figure I.9a6). Il a de plus été montré récemment que la méthylation des sites CpG de ce promoteur (figure I.6b2) démontrait également un caractère périodique au cours de l'activation transcriptionnelle (Métivier *et al.*, 2008).

Plusieurs auteurs ont également observé un tel comportement périodique sur d'autres promoteurs et chez divers organismes eucaryotes (Kang *et al.*, 2002; Nagaich *et al.*, 2004; Väisänen *et al.*, 2005; Karpova *et al.*, 2008; Sun *et al.*, 2008; Degenhardt *et al.*, 2009). En particulier, Sun *et al.* (2008) rapportent des oscillations sur plus de 5000 gènes chez la souris, révélant alors que ces observations, loin d'être des cas particuliers, reflètent un phénomène relativement général. Il peut par ailleurs être observé aussi bien sur un système minimal[23] *in vitro* (figure I.9b, Nagaich *et al.*, 2004) que sur cellule unique vivante (figure I.9c1, Karpova *et al.*, 2008). De plus, parmi ces études, celles où l'évolution du niveau d'ARN du gène a également été mesurée démontrent ce même caractère périodique dans l'activité de transcription (Karpova *et al.*, 2008; Sun *et al.*, 2008; Degenhardt *et al.*, 2009). La dynamique globale du promoteur peut ainsi provoquer des hétérogénéités temporelles lentes dans le processus de transcription.

Redondance fonctionnelle et états alternatifs

Bien que l'on puisse identifier différentes étapes dans l'évolution du promoteur – par exemple Degenhardt *et al.* (2009) identifient des phases d'activation, d'initiation et de

[23] L'approche de Nagaich *et al.* (2004) consiste à reconstituer *in vitro* un promoteur démontrant un comportement périodique avec le minimum de composants : L'ADN du promoteur MMTV, des nucléosomes, le récepteur des glucocorticoïdes et son ligand, le complexe de remodelage SWI/SNF et de l'ATP. Les oscillations disparaissent totalement en l'absence de l'un de ces constituants (en particulier de l'ATP).

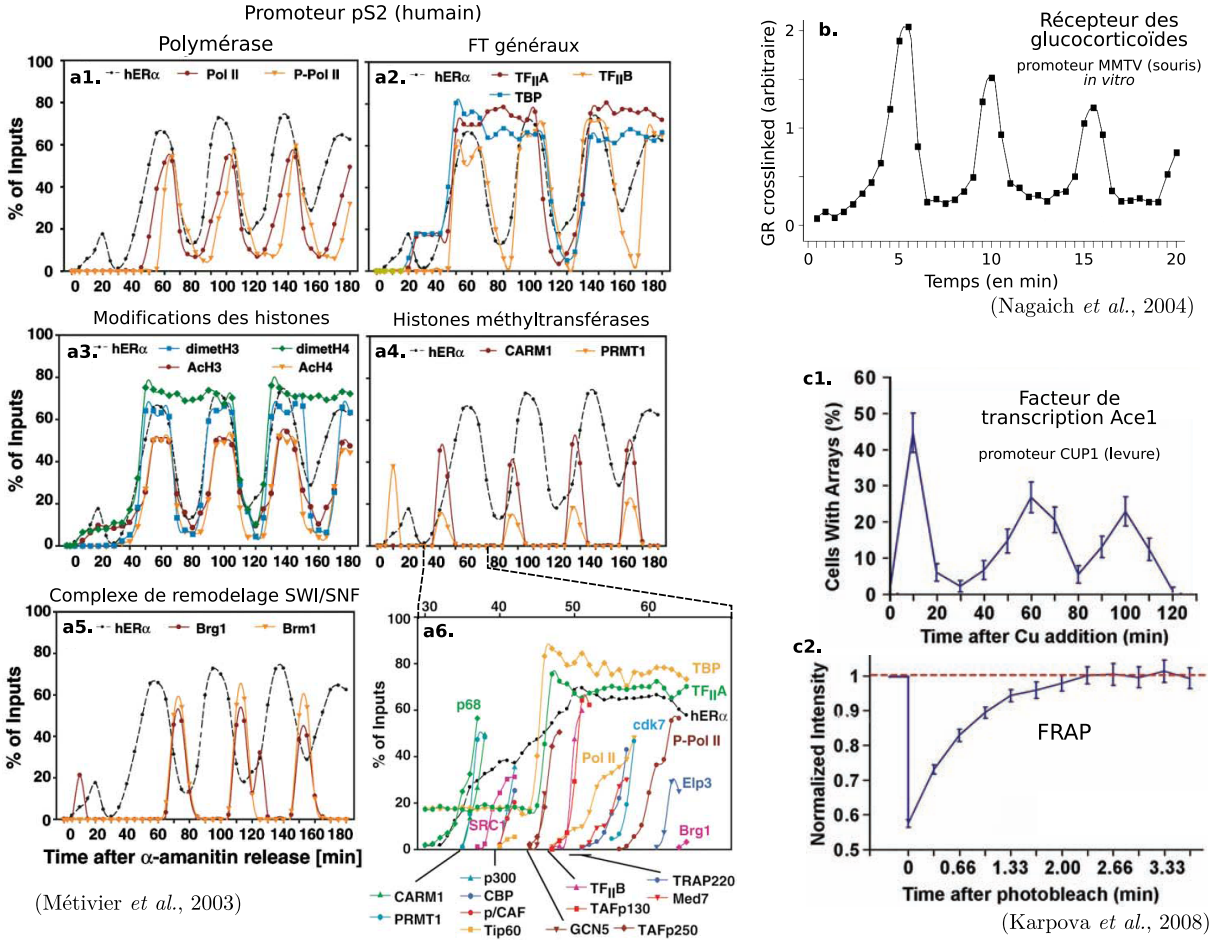


Fig. I.9: Dynamique périodique lente des promoteurs eucaryotes. (a1-a5) Les expériences de cinétique de ChIP (cf Box 2) montrent que, lors de l'activation transcriptionnelle d'un gène, le profil d'occupation de son promoteur par différentes molécules ainsi que les modifications conformationnelles et covalentes de sa chromatine suivent un comportement fortement périodique à une échelle temporelle de l'ordre de l'heure. Les données présentées ici ont été obtenues par Métivier *et al.* (2003) sur le promoteur pS2 (cf figures I.6b1 et I.6b2). Elles décrivent des relations dynamiques complexes entre les différents événements sur le promoteur, révélant des fenêtres temporelles distinctes pour le recrutement des diverses molécules et dont le détail précis peut être observé sur des cinétiques à une résolution temporelle fine de 1 min (a6). (hERα : human estrogen receptor-α). Des oscillations similaires peuvent être observées sur d'autres promoteurs eucaryotes, par exemple : (b) sur un système minimaliste[23] étudié *in vitro* par une technique similaire au ChIP (d'après Nagaich *et al.*, 2004) et (c1) sur un promoteur de levure par vidéo-microscopie (d'après Karpova *et al.*, 2008). Dans ce second cas, une expérience de FRAP (c2) montre qu'il y a bien co-occurrence sur un même promoteur de la dynamique périodique lente et de la mobilité rapides des molécules.

désactivation – il apparaît que de nombreux états alternatifs du promoteur puissent co-exister au sein de la population à un instant donné d'une cinétique de ChIP. En effet, les expériences de re-ChIP[24] montrent que certains cofacteurs ayant des fonctions similaires

[24]Re-ChIP : expérience de ChIP effectuée en immunoprécipitant successivement avec deux anticorps différents afin de ne garder que les fragments d'ADN sur lesquels les deux antigènes sont présents.

et dont la présence est observée simultanément au sein de la population (par exemple CARM1 et PRMT1 ; figure I.9a4) ne sont en réalité jamais présents simultanément sur un même promoteur (Métivier *et al.*, 2003). Ainsi, plusieurs fractions distinctes peuvent coexister au même instant au sein de la population, montrant que, même si l'activité semble structurée et ordonnée au niveau global, la dynamique individuelle de chaque promoteur peut être différente et ne consiste pas en une unique séquence d'événements prédéterminée. Notons cependant qu'il n'est pas possible, à partir de ces données, de savoir si ces complexes s'excluent mutuellement mais peuvent s'associer à tour de rôle sur un même promoteur, ou bien si les différents promoteurs sont temporairement engagés sur des voies différentes ne permettant chacune l'association que d'un seul de ces complexes.

La dynamique globale du promoteur au niveau d'une population de cellules apparaît donc comme le reflet d'un comportement individuel stochastique dans lequel de nombreux chemins alternatifs fonctionnellement redondants peuvent être empruntés.

Dynamique apériodique lente

Il a été observé sur le promoteur MMTV une dynamique lente ne présentant pas le caractère périodique évoqué précédemment (Becker *et al.*, 2002; Hager *et al.*, 2006; Voss *et al.*, 2009). Ce gène, régulé par le récepteur des glucocorticoïdes (GR) chez la souris, est présent en 200 copies répétées en tandem dans une lignée cellulaire. Une telle quantité de sites de fixation, concentrés à un locus donné, permet alors de détecter en vidéo-microscopie sur cellule unique si cette région est occupée par le FT rendu fluorescent. En observant la fraction parmi plusieurs cellules où une occupation du site est détectable, les auteurs constatent un maximum au bout 20 à 30 min après induction du gène, suivi d'une valeur intermédiaire de la probabilité d'occupation du site à partir de 50 min. Cela ne signifie pas que le site est occupé de façon intermédiaire mais plutôt que seule une fraction de la population présente une occupation du site (les données présentées par Becker *et al.* (2002) restent cependant quelque peu ambiguës sur ce point). Comme nous le verrons plus en détail dans le chapitre III, l'amortissement des oscillations observées au niveau d'une population est d'autant plus important que la durée individuelle du cycle pour chaque cellule est variable (Degenhardt *et al.*, 2009). Il est à noter que le promoteur MMTV, étudié *in vitro*, peut démontrer un caractère clairement périodique (figure I.9b, Nagaich *et al.*, 2004). On peut alors imaginer que ce promoteur puisse avoir naturellement une activité périodique, mais que le profil apériodique observé par Becker *et al.* (2002) sur la population provienne d'une forte variabilité inter-individuelle dans la période du cycle. En effet, les auteurs montrent que l'instant où le site apparaît occupé est très différent d'une cellule à l'autre : 20 min après activation du gène pour certaines cellules et jusqu'à 30∼40 min pour d'autres. Ainsi, les 200 promoteurs de certaines cellules sont tous (ou en grande majorité) occupés plus tôt que ceux de certaines autres cellules. Ce comportement pourrait par exemple être expliqué par des différences inter-cellules dans la capacité globale d'expression (Colman-Lerner *et al.*, 2005) telles qu'évoquées à la section 3.4.

Il est intéressant des mettre ce comportement en relation avec les observations de Chubb *et al.* (2006) montrant une dynamique d'activation/désactivation d'un promoteur eucaryote (le gène dscA chez *Dictyostelium*) dont les distributions des durées sont proches de

lois exponentielles. Le promoteur MMTV (Becker *et al.*, 2002) peut alors être vu comme un intermédiaire entre les deux cas extrêmes de l'activation/inactivation très régulière du promoteur pS2 (Métivier *et al.*, 2003) et de celle très irrégulière du promoteur dscA (Chubb *et al.*, 2006).

5.5 Progression de l'état chromatinien

Les expériences de FRAP et de ChIP ont souvent été perçues comme contradictoires car démontrant respectivement des temps de résidence des molécules de la dizaine de secondes à la minute et une activité périodique de la structure de l'ordre de l'heure. Il a été montré récemment que ces deux échelles temporelles co-existaient bien sur un même promoteur (figures I.9c1 et I.9c2, Karpova *et al.*, 2008). La réconciliation de ces échelles de temps et la compréhension de l'émergence d'une dynamique globale lente à partir d'un comportement local rapide ouvrent un certain nombre de questions et constituent un champ de recherche très actuel.

Il a été proposé par plusieurs auteurs que de nombreux événements stochastiques rapides aient lieu sur le promoteur et que seule une faible fraction des ces événements soit « productive ». Il en résulterait une étape irréversible faisant alors progresser le promoteur dans un nouvel état servant de substrat pour la transition suivante. Il s'agit du concept de *transcriptional ratchet* (Métivier *et al.*, 2006) ou du *return-to-template* (Hager *et al.*, 2006). Ces mêmes auteurs ont également proposé que ces événements rares puissent correspondre à des associations spécifiques et durables de molécules qui, la plupart du temps, s'associent de façon non-spécifique et transitoire. Ils postulent alors que ce second type d'associations ne serait pas détectable par les techniques de ChIP alors qu'il constituerait l'essentiel des observations de FRAP. Il est cependant important de noter que la durée d'un motif sur un profil de ChIP ne doit pas être interprété comme reflétant l'association d'une même molécule pendant une longue période. Comme nous le montrerons plus en détail dans le chapitre III, il s'agit de façon plus générale d'un changement dans la probabilité de présence de la molécule provenant de l'évolution au cours du temps de sa cinétique d'association/dissociation, celle-ci pouvant rester dans tous les cas très rapide. Il n'est alors nullement besoin de supposer une telle différence de sensibilité entre les deux techniques.

De nombreuses modifications de l'état de la chromatine, comme la position des nucléosomes et le statut des queues d'histones, peuvent être qualifiées de *méta-stables* car elles nécessitent l'action de complexes de remodelage de la chromatine ou d'enzymes de modification d'histones pour avoir lieu. Elles sont donc de fait moins fréquentes que les associations dissociations des molécules et sont localement irréversibles[25] car elles requièrent, pour être inversées, le recrutement d'un complexe différent de celui qui les a provoquées. Elles peuvent ainsi correspondre à ces réactions rares qui sont à la base du concept de *transcriptional ratchet* (Métivier *et al.*, 2006). La possibilité pour une molécule donnée de s'associer au promoteur est alors déterminée par l'état courant de la chromatine, définis-

[25]Notons qu'il s'agit en réalité de réactions très orientées (et non irréversibles) car ces réactions font intervenir l'hydrolyse de cofacteurs fortement énergétiques (ATP, acetyl-CoA, ...). Ce point se révèlera particulièrement important par la suite.

sant pour les différentes molécules des fenêtres de temps pendant lesquelles elles s'associent un grand nombre de fois et de manière transitoire (Karpova *et al.*, 2008; Reid *et al.*, 2009; Degenhardt *et al.*, 2009). C'est donc la progression lente de l'état chromatinien qui définit le spectre des molécules pouvant être présentes à un instant donné du cycle. Mais ce sont ces mêmes molécules qui provoquent les modifications de l'état de la chromatine. C'est ainsi les interactions combinatoires entre la chromatine et les nombreuses molécules pouvant s'y associer qui définissent la structure du comportement cyclique du promoteur. Il semble alors évident que les dynamiques locales des complexes vont se répercuter sur les propriétés temporelles globales du promoteur. Nous examinons au chapitre III différents modes d'interaction évoqués dans la littérature – tels que le principe du *return-to-template* (Hager *et al.*, 2006), la différence entre *pre-assembly* et *neo-assembly* (Gorski *et al.*, 2008; Darzacq et Singer, 2008; Darzacq *et al.*, 2009), ou entre *random assembly*, *regulated assembly* (Hager *et al.*, 2009) et *preferentially random assembly* (Degenhardt *et al.*, 2009) – au regard de leur influence sur le comportement global. Nous proposons alors certaines hypothèses quant à la nature des interactions entre molécules et avec la chromatine et aux propriétés d'assemblage/désassemblage des complexes.

Ainsi, le promoteur témoigne d'une activité globale complexe et multi-échelle, impliquant toute la machinerie moléculaire et provoquant des hétérogénéités temporelles dans la synthèse des transcrits. Cette activité comporte de nombreux phénomènes probabilistes à différents niveaux – de l'association/dissociation des molécules à l'activation/désactivation du promoteur – dont l'influence collective sur la stochasticité de l'expression génique reste à ce jour très peu étudiée.

Chapitre II

Organisation nucléaire dynamique et expression stochastique des gènes

L'expression génique est un processus intrinsèquement spatialisé. Comme nous l'avons décrit à la section 4 du chapitre I, chez les eucaryotes, une grande partie de ce processus a lieu dans le noyau cellulaire, un espace hétérogène et structuré, et dont la dynamique est elle-même intimement liée au processus d'expression des gènes. Ainsi, le caractère stochastique de l'expression génique a probablement d'étroites relations avec de nombreux mécanismes spatiaux tels que la diffusion des FT dans l'espace nucléaire hétérogène, le déplacement des loci chromosomiques au sein des territoires chromosomiques et, à la fois comme cause et comme conséquence de leur transcription, le recrutement de loci distants dans une même usine à transcription, les propriétés de déplacement des transcrits jusqu'aux pores nucléaires, ... Tous ces mécanismes sont susceptibles de constituer des sources de stochasticité ou d'influencer les propriétés d'autres phénomènes provoquant une variabilité de l'expression. Par exemple, de simples mécanismes de diffusion ou des phénomènes locaux entre gènes situés à des loci potentiellement très distants en séquence (par exemple le partage d'une même usine à transcription) peuvent produire des propriétés particulières dans les corrélations temporelles de l'expression d'un ou de plusieurs gènes.

Alors que le temps constitue un point central de la plupart des modèles représentant des mécanismes intracellulaires et est représenté de façon précise et explicite, l'espace a longtemps été ignoré ou pris en compte de manière très simplifiée. Comme le montrent cependant de plus en plus d'études expérimentales (voir, par exemple, la section 4 du chapitre I), la localisation spatiale des molécules constitue un aspect central du fonctionnement de la cellule. Cet aspect doit alors être pris en compte explicitement dans les études de modélisation, donnant alors lieu à des « modèles 4D » – intégrant 3 dimensions spatiales et une temporelle – (Amar *et al.*, 2002; Bork et Serrano, 2005; Takahashi *et al.*,

2005; Lemerle *et al.*, 2005; Andrews et Arkin, 2006). La plupart des modèles d'expression stochastique des gènes ne tiennent cependant pas compte de ces aspects spatiaux dans leur description du processus d'expression génique. Seuls quelques travaux se sont intéressés à l'influence de la diffusion des facteurs de transcription dans le cadre de l'expression stochastique et ont montré l'importance de ces phénomènes (van Zon *et al.*, 2006; Tkačik et Bialek, 2009, cf chapitre I, section 2.7).

Nous souhaitons ici proposer un modèle permettant d'étudier l'influence de divers mécanismes spatiaux sur le caractère stochastique de l'expression génique. Cette dernière étant potentiellement à la fois cause et conséquence de la structuration dynamique du noyau (par des phénomènes supposés d'auto-organisation ; cf section 4.5, chapitre I), cela implique nécessairement de considérer un questionnement scientifique plus vaste : celui de l'organisation dynamique du noyau cellulaire, dans laquel l'expression génique tient une place centrale. Ainsi, notre but est ici de définir un modèle – et un logiciel de simulation – pour l'étude des phénomènes de structuration et d'auto-organisation de l'espace nucléaire. En outre, ce modèle devra permettre de représenter les différentes étapes du processus d'expression génique, afin de comprendre, à terme, les relations entre ces différents niveaux d'organisation. Un tel projet constitue cependant un travail de recherche de longue haleine et les travaux présentés dans ce chapitre ne constituent que les premières étapes de la démarche de modélisation. Il s'agit essentiellement ici des phases de formulation de la problématique et des questions d'intérêt, d'identification des formalismes et des représentations appropriées et de conception même du modèle de simulation (ainsi que son implémentation). Les avancées réalisées sur cette thématique restent donc encore relativement exploratoires mais nous permettent d'ores et déjà d'entrevoir les possibilités qu'offriront ces travaux ainsi que leurs potentielles limitations.

Nous commençons, dans la section 1, par décrire différents modèles de simulation spatiale à l'échelle cellulaire de la littérature, nous amenant à définir les caractéristiques principales de notre modèle en relation avec l'échelle spatio-temporelle de notre questionnement. Dans la section 2, nous décrivons alors le modèle que nous proposons, en le définissant tout d'abord d'un point de vue fonctionnel, puis en décrivant son implémentation sous la forme d'un logiciel de simulation numérique. Enfin, dans la section 3, nous proposons un certain nombre d'hypothèses ou de mécanismes relatifs à l'expression stochastique des gènes qu'un tel outil permet de tester et d'étudier.

1 Modélisations spatiales intracellulaires

On pourrait être tenté de penser, de prime abord, qu'une simulation décrivant chaque atome de chaque molécule d'un noyau cellulaire constituerait un cas idéal et permettrait de comprendre l'organisation de ce dernier. Il est cependant évident qu'une telle simulation est absolument inaccessible d'un point de vue computationnel. De plus, il est capital de garder à l'esprit que le but d'un modèle n'est pas de reproduire la réalité de la manière la plus fidèle et la plus complète possible, mais, au contraire, de l'abstraire et de la simplifier autant que possible dans le but de comprendre un phénomène particulier et de répondre à un questionnement scientifique donné (Beslon, 2008). Ainsi, dans le contexte de notre

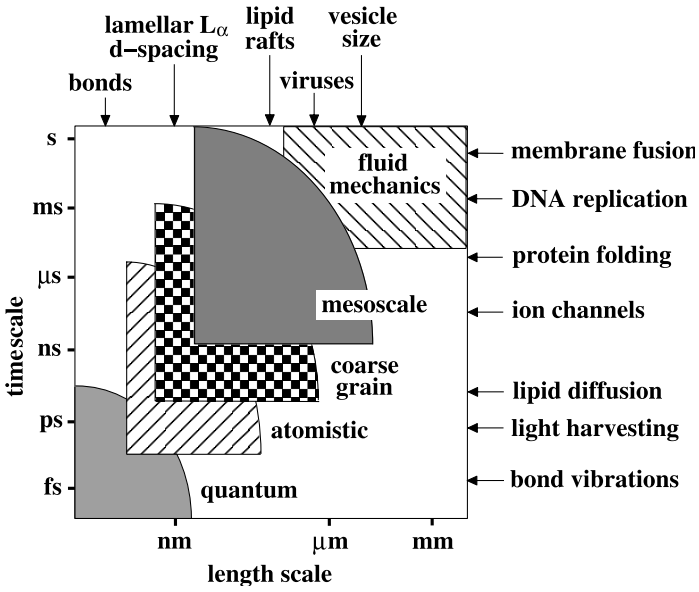


Fig. II.1: Échelles spatio-temporelles et formalismes de simulation intracellulaire. En fonction du formalisme qu'ils emploient pour représenter la matière, les nombreux modèles de simulation spatiale intracellulaire permettent de reproduire des phénomènes à des échelles spatio-temporelles très diverses. De nombreux mécanismes peuvent de plus être reproduits par plusieurs modèles considérant des niveaux d'abstraction différents. Ces recouvrements entre les échelles de description jouent un rôle particulièrement important pour la définition et la paramétrisation de chacun de ces formalismes et dans le cadre des approches multi-échelles. Illustration d'après Nielsen *et al.* (2004).

étude, il est important de différencier les aspects de la réalité qui doivent être pris en compte de façon explicite de ceux qui peuvent – ou doivent – être simplifiés.

Dans tout modèle, l'échelle spatio-temporelle de l'objet et/ou du phénomène d'intérêt détermine ce qui doit être pris en compte explicitement et ce qui doit être admis et représenté de manière phénoménologique, voire ignoré (figure II.1). Par exemple, à une échelle *macroscopique*, la matière est considérée comme continue et l'on peut se baser sur des notions moyennes, telles que la densité, la température, les taux de réactions ... Des cas typiques de modélisations à cette échelle considèrent par exemple des réseaux de régulation génique ou des réseaux métaboliques, décrivant l'évolution temporelle de concentrations (supposées homogènes dans la cellule et décrites par une valeur réelle) gouvernées par un ensemble de réactions biochimiques. De plus, même si ce formalisme n'est pas employé pour représenter des phénomènes intracellulaires, il s'agit également l'échelle d'étude de la mécanique des fluides, décrivant le comportement de la matière de façon continue par des champs scalaires et vectoriels (pression, vitesse, ...). À l'autre extrême, l'échelle *atomique* ou *nanoscopique* considère quant à elle les différents atomes des molécules individuellement et de manière explicite et permet d'étudier le comportement d'une ou de quelques molécules. À cette échelle, les notions moyennes évoquées précédemment n'ont pas de sens. Cette échelle est par exemple celle de l'étude des changements conformationnels d'une protéine et de l'interaction de deux macromolécules. Entre ces deux échelles se trouve un niveau de description intermédiaire, qualifié de *mésoscopique* (du grec *mesos* : « au milieu »), même si la définition précise de cette échelle est assez variable (figure II.1). Il s'agit de l'échelle où des concepts moyens tels que la température et la densité ont un sens, mais où l'on doit néanmoins considérer certaines molécules individuelles de façon explicite (figure II.1). On observe alors le comportement collectif de grandes structures multiprotéiques, typiquement de plusieurs centaines d'entités.

De nombreux phénomènes observés au sein du noyau cellulaire ont lieu à cette échelle

mésoscopique. En effet, ils font souvent intervenir des entités uniques nécessitant une représentation au moins partiellement discrète de la matière (nucléation d'une structure, rencontre d'un FT et d'un locus chromosomique ...) et se manifestent à un niveau global (niveaux d'expression, dynamique d'un corps nucléaire, organisation globale du noyau). La capacité du modèle à exprimer des phénomènes de structuration à l'échelle mésoscopique devra donc être une de nos priorités centrales. Afin de déterminer ses caractéristiques, nous explorons dans cette section différentes approches de simulation spatiale de la littérature employées pour représenter des phénomènes intracellulaires, et les analysons au regard de notre objet d'étude et de notre questionnement.

Comme nous le verrons, la plupart des modèles mésoscopiques intracellulaires proches de nos questionnements ne prennent pas en compte les interactions physiques entre molécules de façon réaliste (forces électromagnétiques, liaisons souples, ...). De nombreux phénomènes mésoscopiques nucléaires se basent pourtant vraisemblablement sur de telles propriétés. Par exemple, comme nous le verrons, des comportements de phase (analogues à ceux observés dans le noyau ; cf chapitre I, section 4.3) ne peuvent être reproduits que par des modèles considérant explicitement les forces d'attraction/répulsion entre molécules. De plus, il paraît probable que la flexibilité des protéines et la souplesse de leurs interactions – une caractéristique essentielle de ces molécules (Zaccai, 2000) – soient déterminantes dans la dynamique et la fluidité des structures nucléaires. Ainsi, après une brève description des formalismes macroscopiques de modélisation spatiale et de leurs limites dans le cadre de notre étude (section 1.1), nous commencerons par décrire les techniques de simulation physique aux échelles atomiques et moléculaires (sections 1.2 et 1.3), avant de décrire enfin les modèles spécifiques à des questionnements mésoscopiques les plus proches des nôtres (section 1.4). Le modèle que nous proposons dans ce chapitre s'inspire fortement des méthodes et des formalismes employés aux échelles inférieures.

1.1 Modèles macroscopiques

Les réactions biochimiques au sein d'une cellule sont classiquement représentées par des équations différentielles couplées décrivant l'évolution au cours du temps des concentrations des différentes espèces chimiques. Certains modèles ont alors ajouté un caractère spatial à une telle description en partageant l'espace en compartiments au sein desquels évoluent les concentrations des différentes espèces moléculaires et entre lesquels ont lieu des échanges de matière (Schaff *et al.*, 1997; Takahashi *et al.*, 2005; Lemerle *et al.*, 2005; Andrews et Arkin, 2006). Pour notre étude, de tels modèles pêchent par l'absence d'une définition explicite de l'espace et la présupposition *a priori* d'une structuration globale interdisant sa modification en retour par l'activité du système. Les mécanismes d'auto-organisation au sein du noyau cellulaire ne peuvent de toute évidence pas être étudiés par des modèles présupposant une telle structure globale puisqu'il s'agit au contraire d'observer son émergence spontanée à partir d'une définition locale de la matière.

Paradoxalement, un modèle vieux de près de 60 ans, le modèle de morphogenèse d'Alan Turing (1952), répond en partie aux problèmes des modèles compartimentaux, sans pour autant se baser sur une description discrète de la matière. En effet, ce modèle considère un espace explicite sans structuration préétablie et permettant l'apparition et le maintien

d'hétérogénéités spatiales. Mais ce type de modèle ne prend en compte qu'une partie des mécanismes pouvant provoquer une structuration de l'espace : il ne considère que la diffusion et la réaction des espèces chimiques et non l'interaction physique des molécules. Ainsi, son application à l'étude de corps nucléaires nécessite la supposition de structures globales préexistantes (Carrero *et al.*, 2006). Bien qu'utiles pour rendre compte de nombreux phénomènes d'organisation spatiale spontanée, les modèles de réaction-diffusion purs restent sur une description continue de la matière et ne permettent donc pas d'étudier les phénomènes stochastiques dus au caractère discret des entités en présence, comme c'est le cas pour l'expression des gènes. Il est en effet nécessaire de considérer des entités discrètes, se déplaçant dans un espace explicite et interagissant physiquement les unes avec les autres (Bork et Serrano, 2005; Andrews et Arkin, 2006). C'est en particulier ce que considèrent, à un niveau beaucoup plus fin, de nombreux modèles de simulation moléculaire.

1.2 Simulation à l'échelle atomique

À une échelle très inférieure, les modèles *tout-atomes* sont fréquemment employés pour prédire des comportements très précis des molécules ou des complexes, tels que l'assemblage ou la modification conformationnelle d'un complexe, ou encore la dynamique de vibration d'une protéine. Dans ces approches, qualifiées de dynamique moléculaire (MD), chaque atome de chaque molécule est considéré individuellement[1] dans un espace 3D continu. En associant une énergie potentielle à chaque paire d'atomes (en fonction de leur distance ; cf figure II.2) et à chaque liaison covalente (en fonction sa longueur, de sa torsion ...), on peut déduire la force appliquée à chacun des atomes et déterminer leurs mouvements en appliquant la seconde loi de Newton[2]. L'allure de cette énergie potentielle pour une paire d'atomes non liés dépend de la nature des deux atomes considérés. Elle est,

[1]Les atomes d'hydrogène sont parfois ignorés et implicitement associés à l'atome auquel ils sont liés.

[2]La seconde loi de Newton permet de déduire le déplacement d'un objet en fonction de sa masse et de la somme des forces qui lui sont appliquées.

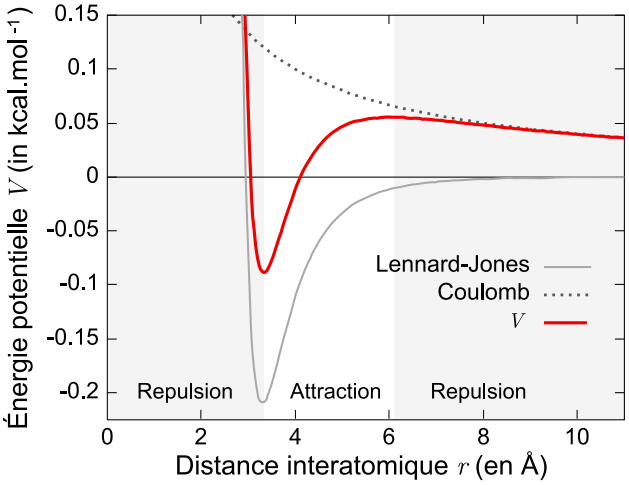


Fig. II.2: Énergie potentielle d'interaction entre deux atomes d'oxygène. La force que deux atomes non-liés exercent l'un sur l'autre peut être décrite par l'allure de leur énergie potentielle d'interaction V en fonction de la distance r séparant leurs centres. Ce potentiel de paire est souvent décrit par l'expression empirique (II.1) dont les deux premiers termes constituent un potentiel de Lennard-Jones (représentant les interactions de van der Waals) et dont le dernier décrit la force de Coulomb. Comme l'illustre la figure II.3, la force entre les deux atomes correspond à la dérivée de l'énergie potentielle. Celle-ci décrit donc ici une répulsion à courtes et à longues distances et une attraction à moyennes distances. (Valeurs paramétriques d'après Cornell *et al.* (1995)).

en outre, très liée au mode d'interaction utilisé. Ainsi, les nombreux logiciels de simulation appliquant un modèle de MD (dont on peut citer parmi les plus connus CHARMM, AMBER, GROMOS et GROMACS) emploient chacun des énergies potentielles légèrement différentes (Hobza *et al.*, 1997), celles-ci étant soit empiriques, soit obtenues par des calculs de physique quantique. Une représentation souvent employée du potentiel de paire entre deux atomes non liés est donnée par la formule empirique suivante[3] :

$$V(r) = \epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - 2\left(\frac{\sigma}{r}\right)^{6}\right] + \frac{Q}{r} \tag{II.1}$$

où r représente la distance entre les centres des deux atomes considérés et ϵ, σ et Q sont des paramètres dépendant du type d'atomes (voir par exemple (Cornell *et al.*, 1995) ou (MacKerell Jr. *et al.*, 1995) pour une description détaillée du paramétrage de ce potentiel). La figure II.2 représente un exemple de cette fonction de potentiel pour deux atomes d'oxygène. La force entre les deux atomes étant la dérivée de leur énergie d'interaction en fonction de r, une énergie décroissante correspond à une répulsion et une énergie croissante, à une attraction (figure II.3a). Ainsi, un potentiel d'interaction s'interprète intuitivement par le fait que les deux particules[4] tendent spontanément à aller vers des minima locaux d'énergie. De plus, de nombreux aspects de leur interaction sont déterminés par diverses caractéristiques du profil de ce potentiel, permettant alors une grande expressivité dans la description de l'interaction physique de deux particules (figures II.3b et II.3c1-5).

Les deux premiers termes de l'expression (II.1) correspondent à un potentiel de Lennard-Jones (figure II.2) et représentent les forces de van der Waals entre les deux atomes : le terme en r^{-12} est très répulsif à courtes distances, décrivant le principe Pauli empêchant les cortèges électroniques des deux atomes de s'interpénétrer, et le terme en r^{-6} est attractif à moyennes distances. Les paramètres σ et ϵ décrivent respectivement la distance et l'énergie du point où le potentiel de Lennard-Jones est minimal, correspondant au contact entre les deux atomes. Le troisième terme de l'expression (II.1) correspond aux interactions électrostatiques entre les deux atomes, décrites par la loi de Coulomb. Ce terme en r^{-1} décrit des forces de portées beaucoup plus longues que celles de van der Waals. Le paramètre Q correspond (à un facteur près) au produit des charges des deux atomes et est donc positif si ces dernières sont de même signe et négatif sinon, décrivant alors respectivement une force répulsive ou attractive.

En décrivant les énergies d'interaction entre atomes non liés par des potentiels d'énergie et les liaisons covalentes par des liaisons élastiques (sur les longueurs et sur les angles), il est alors possible de déterminer l'énergie potentielle globale du système et d'en déduire la probabilité des différentes conformations sans avoir recours à une simulation dynamique. Mais de ces énergies potentielles peuvent également être déduites des forces permettant de décrire l'évolution de la vitesse de chaque atome au cours du temps et d'en déduire toute la dynamique du système. Il s'agit là du principe de base des simulations de MD (Allen et

[3]Il existe différentes façons de paramétrer ce potentiel, mais il s'agit toujours de la somme de trois termes, respectivement en r^{-12}, r^{-6} et r^{-1}.

[4]Dans ce chapitre, le mot « particule » est employé pour désigner une entité matérielle considérée comme élémentaire. Il pourra s'agir, en fonction du contexte, d'un atome, d'un groupe d'atome, d'une molécule, voire d'un groupe de molécules.

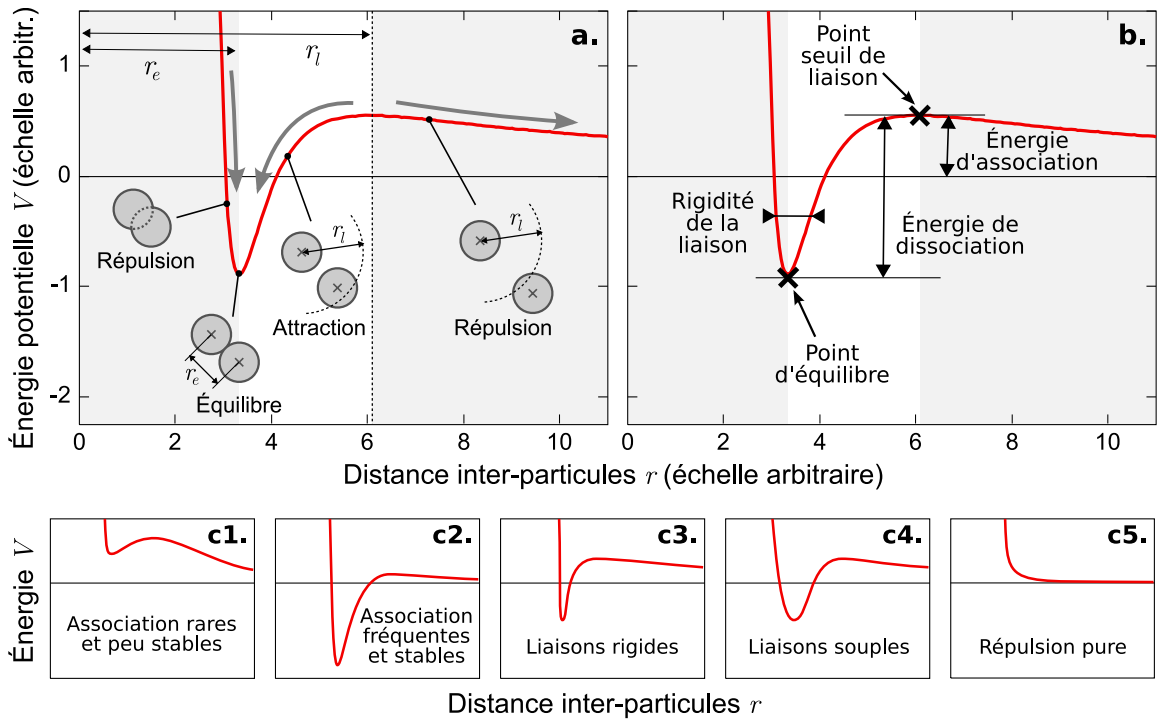


Fig. II.3: Interprétation et pouvoir d'expression d'un potentiel d'interaction de paire. Le profil de l'énergie potentielle V représentant l'interaction entre deux particules en fonction de leur distance r permet de rendre compte de nombreux aspects physiques de leur association/dissociation. Le principe illustré ici s'applique aussi bien à l'interaction entre atomes (comme dans la figure II.2), qu'à l'interaction entre grains ou entre macromolécules entières (cf section 1.3), le potentiel ayant par ailleurs souvent un profil qualitativement similaire dans ces différents cas. (a) La force qu'exercent les deux particules l'une sur l'autre est déterminée par la dérivée de l'énergie potentielle V par rapport à r. Le potentiel fortement décroissant pour $r < r_e$ décrit donc ici une répulsion importante des deux particules (correspondant à la force de contact), la partie croissante pour $r_e \leq r < r_l$ décrit une force d'attraction et la partie décroissante pour $r_l \leq r$, à nouveau une répulsion. Ainsi, dès lors que la distance inter-particules r est inférieure à r_l, les deux particules tendent spontanément vers l'équilibre de la liaison $r = r_e$ où le potentiel admet une tangente horizontale convexe. (b) Non seulement le point d'équilibre et le point seuil de liaison définissent les distances r_e et r_l, mais ils déterminent également l'affinité des deux particules. La hauteur de la barrière de potentiel décrit l'énergie à fournir pour leur association (depuis une distance théoriquement infinie où le potentiel est nul) et la profondeur du puits correspond à l'énergie à fournir pour leur dissociation, permettant alors de représenter simplement différentes affinités (c1,c2). De plus, la largeur du puits de potentiel détermine à quel point deux particules associées peuvent facilement s'écarter de leur distance d'équilibre r_e, permettant alors de décrire des liaisons plus ou moins souples (c3,c4). Enfin, certaines particules peuvent n'interagir que par une simple exclusion stérique, auquel cas le potentiel reste décroissant et ne représente qu'une force de contact (c5).

Tildesley, 1991; Frenkel et Smit, 2002; Steinbach, 2005; Vattulainen et Karttunen, 2005; Karplus et Kuriyan, 2005).

Le coût computationnel que représente une telle approche ne permet cependant que de simuler des systèmes relativement restreints (quelques macromolécules) et sur des temps

très courts (quelques dizaines ou centaines de nanosecondes) (Nielsen *et al.*, 2004; Karplus et Kuriyan, 2005; Venturoli *et al.*, 2006; Tozzini, 2010, et figure II.1). Bien que ce formalisme permette de prendre en compte l'interaction physique des molécules dans l'espace, il semble évident que ce niveau de description très précis et les limitations en termes de calculabilité qui y sont liées ne conviennent pas pour notre questionnement. Il existe cependant plusieurs variantes de ce modèle de MD permettant d'accélérer les calculs et donc de simuler des systèmes plus volumineux et sur des durées plus longues. Mais il s'agit dans tous les cas de simplifier la réalité au prix d'une perte en réalisme de la simulation. En particulier, il est possible, comme nous allons le décrire maintenant, de considérer les molécules de solvant de façon implicite et/ou de simplifier la description structurelle des macromolécules.

Représentation implicite du solvant

Une possibilité souvent employée pour rendre plus efficaces les simulations moléculaires est de ne pas représenter explicitement les molécules de solvant (eau, ions ...), mais de ne considérer que leur effet sur les atomes des molécules décrites explicitement. Une telle représentation implicite doit alors être prise en compte à plusieurs niveaux.

D'un point de vue énergétique, la présence d'un solvant modifie les propriétés d'interaction entre atomes. Il existe différents modèles de solvants implicites prenant en compte le fait que, malgré l'absence d'une description explicite des molécules qui le peuplent, le milieu n'est pas vide et possède donc des propriétés particulières (constante diélectrique différente de 1, présence d'ions répartis non-uniformément en fonction de la charge des atomes avoisinants, ...). Parmi les modèles permettant de décrire de façon précise l'énergie de solvatation d'une macromolécule, on peut citer par exemple l'équation de Poisson-Boltzmann (Wagoner et Baker, 2004), l'approximation Born généralisée (Tsui et Case, 2000) et les modèles SASA (Solvant Accessible Surface Area) (Ferrara *et al.*, 2002). Mais ces modèles ne permettent pas une description des énergies d'interaction comme étant additives par paires d'atomes, ce qui rend particulièrement difficile leur application pour la simulation dynamique. Une méthode simple, bien que sa justification physique soit relativement limitée, est de modifier le dernier terme de l'expression (II.1) décrivant les forces de Coulomb, afin de représenter les différents phénomènes liés au solvant (Steinbach, 2005).

Outre son effet sur les potentiels d'interaction entre atomes, le solvant est responsable de deux forces qu'il convient de prendre en compte dans une simulation dynamique s'il est décrit de façon implicite : une force dite *fluctuative* et une force dite *dissipative*. La première est due au choc des molécules de solvant sur les macromolécules et correspond à une force aléatoire appliquée à chaque atome et dont l'amplitude moyenne dépend de la température. La seconde reflète la viscosité du milieu faisant opposition au déplacement des molécules et se traduisant par une force opposée et proportionnelle à la vitesse de chaque atome. Ces deux forces sont responsables du mouvement brownien des molécules dans le solvant. Lorsqu'elles sont prises en compte explicitement, l'expression de la loi du mouvement telle qu'utilisée en MD devient alors une équation différentielle stochastique et l'on parle de dynamique de Langevin (LD). De plus, on peut faire une approximation

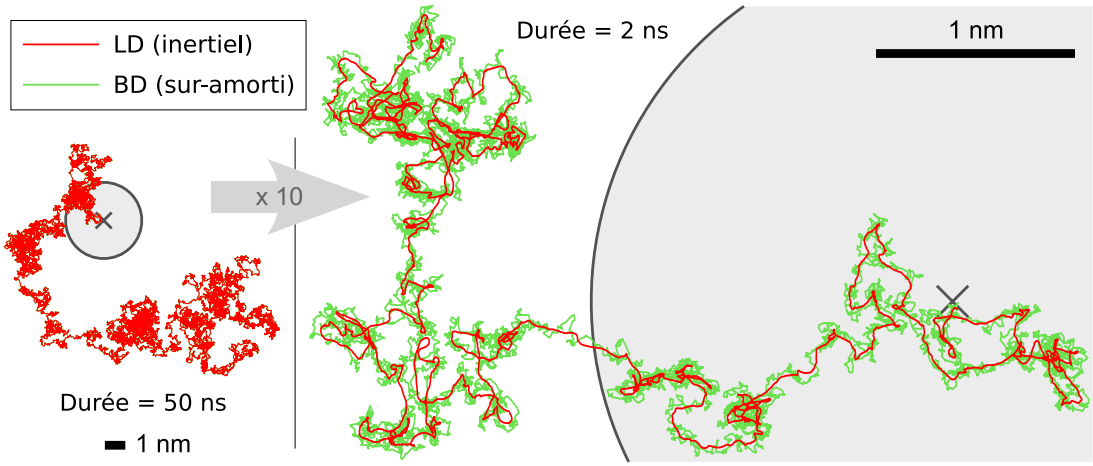


Fig. II.4: Hypothèse de sur-amortissement pour la diffusion d'une particule. Cette figure représente la diffusion d'une protéine (disque gris), observée à deux échelles spatio-temporelles différentes (sur des durées de 50 ns et de 2 ns respectivement), et simulée avec deux lois du mouvement différentes : une équation de Langevin (LD ; Langevin Dynamics), prenant en compte la nature inertielle du déplacement, et une équation de Langevin sur-amortie (parfois qualifiée de BD ; Brownian Dynamics), supposant l'effet de la viscosité du milieu très supérieur à l'inertie de la particule. Comme l'illustre cette figure, la dynamique sur-amortie tend vers le comportement inertiel dans la limite des temps longs (à gauche). Mais sur des durées et des distances courtes (à droite), les deux trajectoires diffèrent significativement l'une de l'autre et seule l'équation de Langevin complète (LD) reproduit de façon réaliste le mouvement de la particule. Les valeurs paramétriques utilisées ici sont celles d'une protéine de rayon = 1.8 nm et de masse = 43 kDa, dans de l'eau à 310 K.

supplémentaire, dite de *sur-amortissement*, si l'on ignore le caractère inertiel de mouvement[5]. On parle alors de dynamique brownienne[6] (BD) et cette approximation affecte essentiellement le mouvement de la particule sur les temps courts (figure II.4).

La représentation implicite des molécules du solvant permet de réduire sensiblement le temps de calcul et autorise donc à simuler des systèmes plus importants. Mais, afin de réellement passer à une échelle supérieure en termes de temps simulé et de taille du système, il est nécessaire de faire des approximations supplémentaires sur la description des molécules d'intérêt.

[5]Cela se traduit par la supression d'un terme dans la loi du mouvement. Nous reviendrons sur ce point dans la section 2.

[6]Bien que, dans la plupart des cas, les termes *Brownian dynamics* (BD) et *Langevin dynamics* (LD) sont employés pour désigner respectivement le cas sur-amorti et le cas non sur-amorti (Takahashi *et al.*, 2005; Grima et Schnell, 2006; Nielsen *et al.*, 2004), on note que BD peut parfois désigner le cas non sur-amorti (Allen et Tildesley, 1991; Chen et Kim, 2004).

1.3 Coarse graining : structures moléculaires simplifiées

De nombreux modèles, dits de *coarse-graining*, font l'approximation qu'un groupe d'atomes voisins d'une même macromolécule peut être considéré comme une entité unique, appelée *grain*. Cette technique permet de simplifier significativement la description des molécules, aussi bien dans leur morphologie que dans leur comportement, en ne gardant que les propriétés physiques essentielles pour l'étude d'un phénomène donné et améliorant alors sensiblement les possibilités de calcul. Ces modèles sont employés dans des contextes différents. Ils sont en particulier très utilisés pour l'étude de l'organisation des membranes phospholipidiques (faisant intervenir de l'ordre d'une ou plusieurs centaines de phospholipides, représentés chacun par une dizaine de grains, et simulées sur des durées allant jusqu'à la milliseconde) (Nielsen *et al.*, 2004; Venturoli *et al.*, 2006) ou pour l'étude du repliement et des propriétés mécaniques d'une protéine (considérant souvent une unique protéine et un grain par acide aminé) (Head-Gordon et Brown, 2003; Tozzini, 2005).

De la même façon que dans les modèles de MD, les interactions entre les grains sont représentées par des liaisons élastiques pour les grains liés et par des énergies potentielles pour les paires de grains non liés (sauf exception que nous préciserons ultérieurement). Le potentiel entre deux grains dépend du type d'atomes que ceux-ci représentent. En fonction du formalisme employé, il est parfois préférable, pour un bon réalisme physique, que les grains représentent des atomes de même nature. Cependant la détermination du potentiel entre paires de grains est plus délicate que pour des modèles de MD et dépend en particulier du but du modèle.

En ce qui concerne les modèles de repliement de protéines, il faut distinguer d'une part ceux utilisés pour étudier le chemin de repliement et/ou certaines propriétés mécaniques d'une protéine dont on connaît déjà la conformation native (par cristallographie par exemple), et d'autre part ceux servant au contraire à déterminer cette conformation repliée de la protéine, uniquement à partir de sa séquence d'acides aminés. Il existe différents modèles pour ces deux cas de figure et les potentiels de paire entre grains peuvent y être choisis de manières très différentes.

Chemin de repliement et propriétés mécaniques d'une protéine de conformation native connue

Il existe principalement deux familles de modèles se basant sur la connaissance de la conformation repliée de la protéine d'intérêt : ceux dérivés du modèle de Gō et les modèles de réseau élastique (ENM pour Elastic Network Models). Dans ces deux cas les interactions entre grains sont biaiséés vers la conformation native.

Le modèle de Gō est une description minimaliste d'une protéine reproduisant son chemin de repliement en la décrivant comme une chaîne articulée où chaque grain représente un acide aminé et où les interactions entre grains sont indépendantes de la séquence (Head-Gordon et Brown, 2003; Tozzini, 2005, 2010). Les grains liés interagissent selon des liaisons élastiques similaires à celles utilisées en MD et les grains non liés, par des potentiels de Lennard-Jones où seuls les paires de grains en contact dans la conformation native ont

une composante attractive[7]. Le paysage énergétique du système est alors très régulier et la chaîne se replie spontanément vers la conformation native. Il est alors possible d'étudier le ou les chemins de repliement de la protéine dans différents contextes ou en fonction de certaines propriétés de la protéine (pour des exemples de telles études, voir (Koga et Takada, 2001; Takagi *et al.*, 2003; Levy *et al.*, 2004)). Ce modèle permet également de reproduire certaines propriétés mécaniques de la protéine repliée. Sur ce plan, le modèle de Gō est cependant supplanté par un autre modèle, l'ENM, qui permet d'étudier ces propriétés de manière plus précise.

L'ENM est un modèle particulièrement simple décrivant la protéine d'intérêt comme un ensemble de grains, représentant chacun un acide aminé. Ceux-ci sont reliés par un réseau de liaisons élastiques de mêmes coefficients d'élasticité, reliant chaque paire de grains proches (relativement à une distance seuil donnée), que les acides aminés correspondants soient liés ou non, et maintenant l'ensemble dans la conformation native (Tozzini, 2005, 2010). Ce modèle, se basant sur de simples liaisons élastiques, constitue un approximateur harmonique de la dynamique de la protéine, c'est-à-dire qu'il n'est valable que pour des mouvements autour de l'état d'équilibre. Mais, bien que n'ayant que deux paramètres (la distance seuil et le coefficient d'élasticité), il est cependant capable de reproduire de façon remarquablement précise ces propriétés mécaniques essentielles de la protéine telles que sa flexibilité et sa dynamique vibratoire lente, et ce, en accord avec les observations expérimentales (Doruker *et al.*, 2002; Delarue et Dumas, 2004; Kurkcuoglu *et al.*, 2009; Kim *et al.*, 2009; Tozzini, 2005, 2010). Un avantage de ce formalisme est également la possibilité d'en faire un traitement analytique en décomposant la dynamique du système en modes normaux de vibration (des composantes élémentaires simples dans la dynamique spontanée de la molécule) donnant alors une description bien plus précise et générale de la dynamique que ne le permet la simulation (Doruker *et al.*, 2002; Delarue et Dumas, 2004; Kurkcuoglu *et al.*, 2009; Kim *et al.*, 2009).

Ce formalisme peut de plus être étendu en utilisant des grains de différentes tailles, ouvrant alors la possibilité aux approches multi-échelles. Il est en effet possible de décrire la structure de la protéine ainsi que sa dynamique avec une précision arbitraire : plus le découpage est grossier, plus le système ne reproduit fidèlement que les mouvements lents de la protéine (Doruker *et al.*, 2002). Dans ce cadre, plusieurs méthodes (dont certaines se basant sur un ENM à l'échelle des acides aminés) ont été récemment proposées pour identifier les grands domaines rigides au sein des protéines et automatiser ainsi le découpage de celles-ci en quelques blocs (Kundu *et al.*, 2004; Potestio *et al.*, 2009; Abyzov *et al.*, 2010). Certains travaux emploient de plus une description multi-échelles, soit en décrivant les différentes parties de la protéine avec différents niveaux de résolution (Kurkcuoglu *et al.*, 2009), soit en déduisant la dynamique globale d'une protéine – voire d'un complexe multi-protéique – à partir de la dynamique individuelle des différents domaines (Kim *et al.*, 2009).

Ainsi, l'ENM est un modèle très simple aussi bien dans sa description que dans sa paramétrisation et peut être employé pour représenter des macromolécules avec un niveau de précision arbitraire afin de capturer leurs propriétés mécaniques globales. Il paraît donc

[7]Notons que dans les modèles de Gō, le terme attractif du potentiel de Lennard-Jones est souvent en r^{-10} et non r^{-6}, impliquant des forces de portée moins longue.

tout à fait pertinent pour répondre à un questionnement comme le nôtre dans lequel les propriétés physiques des molécules doivent être prises en compte, mais suffisamment simplifiées pour permettre une simulation efficace et une paramétrisation simple. Dans ce cadre, il a en particulier été fait récemment usage d'un modèle hybride combinant une description par ENM et par potentiels de paires pour étudier le processus d'association protéine-protéine et l'assemblage de grands complexes multiprotéiques (Periole *et al.*, 2009).

Prédiction de conformations repliées et interactions de plusieurs molécules : vers des interactions intra- et inter-moléculaires plus génériques ?

Les modèles décrits précédemment ne permettent pas de prédire la conformation repliée d'une protéine puisqu'ils se basent précisément sur cette information et ignorent la séquence d'acides aminés. Les protéines dont la conformation native a été identifiée expérimentalement ne représentent pourtant qu'une fraction très faible de l'ensemble des protéines connues (et, en particulier, une fraction très faible de celles dont on connaît la séquence codante d'ADN et donc d'acides aminés). Il existe ainsi des modèles dont le but est de prédire la conformation repliée d'une protéine à partir de sa séquence, en reproduisant les propriétés d'interactions entre acides aminés à l'aide de potentiels de paires. Les divers modèles diffèrent alors essentiellement par la méthode employée pour déterminer les fonctions de potentiel (Head-Gordon et Brown, 2003; Tozzini, 2005, 2010). De plus, certaines de ces méthodes sont assez similaires à celle utilisées pour déterminer les potentiels entre paires de grains dans les simulations coarse-grained de membranes phospolipidiques, même si les deux champs bibliographiques sont assez distincts (Nielsen *et al.*, 2004; Lyubartsev et Laaksonen, 2004; Lyubartsev, 2005; Vattulainen et Karttunen, 2005). Dans tous les cas, les potentiels entre paires de grains ne sont pas décrits par une expression mathématique (comme c'est le cas de l'expression (II.1) par exemple) mais ont un profil arbitraire (on parle également de potentiels tabulés) et sont déterminés de manière numérique.

On distingue deux grandes catégories de méthodes pour obtenir ces potentiels : les méthodes – dites statistiques ou empiriques – basées sur des connaissances et celles basées sur la reproduction de propriétés physiques (Head-Gordon et Brown, 2003; Nielsen *et al.*, 2004; Tozzini, 2005, 2010). Dans le premier cas, il s'agit de déterminer un potentiel entre paires d'acides aminés à partir d'un ensemble statistiquement représentatif d'exemples de conformations connues de protéines. La seconde famille de méthodes vise quant à elle à déterminer le potentiel de paire reproduisant certaines données physiques déterminées soit expérimentalement, soit par une simulation en MD (on parle également de potentiel effectif). Par exemple, on peut simuler l'organisation d'une membrane de phospholipides de très petite taille afin de déterminer les fonctions de distribution radiale ou les profils de densité. Il existe alors différentes méthodes d'inversion (particulièrement développées dans le cas de membranes phospholipidiques) pour en déduire les potentiels entre grains (Shelley *et al.*, 2001a; Nielsen *et al.*, 2004; Lyubartsev et Laaksonen, 2004; Lyubartsev, 2005). Notons finalement que certains modèles de coarse graining emploient une description plus précise qu'un simple potentiel de paire entre grains, comme par exemple des

potentiels anisotropes (c'est-à-dire dépendant de l'orientation des grains l'un par rapport à l'autre) (Buchete *et al.*, 2004; Mukherjee *et al.*, 2005). Mais une telle description, bien qu'améliorant le réalisme de la simulation, est particulièrement délicate à employer et à paramétrer et reste encore peu exploitée.

Bien que les potentiels obtenus par ces différentes méthodes se révèlent être relativement efficaces, il est essentiel de remarquer qu'ils dépendent fortement de l'ensemble d'exemples d'apprentissage ou du contexte de la simulation en MD et ne reflètent pas de manière objective et générale les propriétés d'interactions entre grains (Nielsen *et al.*, 2004; Tozzini, 2005, 2010). Ainsi, plus la structure des macromolécules est simplifiée, plus le paramétrage des interactions doit être spécifique à la simulation et ne peut être déterminé simplement à partir de la composition des grains. Il est cependant important de remarquer que, quelle que soit la méthode employée pour déterminer les potentiels d'interaction, ceux-ci ont une allure qualitativement similaire à ceux des figures II.2 et II.3 et de l'expression (II.1). En effet, on observe systématiquement une barrière fortement répulsive pour les distances courtes et souvent un bassin pour les distances moyennes suivi d'une barrière énergétique (impliquant alors une répulsion à des distances plus élevées) (Shelley *et al.*, 2001a; Buchete *et al.*, 2004; Lyubartsev et Laaksonen, 2004; Lyubartsev, 2005; Mukherjee *et al.*, 2005). Notons que le potentiel peut parfois présenter plusieurs bassins énergétiques.

Modèles réduits : vers une paramétrisation plus simple

Contrairement aux travaux présentés précédemment utilisant des potentiels d'interaction arbitraire pouvant avoir un profil très détaillé, d'autres approches de simulation coarse grain tendent au contraire à réduire l'espace paramétrique et à n'utiliser qu'une description très simplifiée des interactions entre grains. On peut en particulier mentionner l'existence de deux formalismes, relativement différents l'un de l'autre, mais prenant une telle approche.

Le modèle de Dissipative Particle Dynamics (DPD), essentiellement employé pour l'étude de phénomènes concernant les bio-membranes, consiste à décrire de manière explicite toutes les espèces moléculaires du système (y compris le solvant), mais sous forme de grains dits dissipatifs, c'est-à-dire dont on considère de façon implicite la dynamique interne (Groot et Warren, 1997; Vattulainen et Karttunen, 2005; Laradji et Sunil Kumar, 2005; Venturoli *et al.*, 2006). En effet, l'interaction entre chaque paire de grains fait intervenir trois forces : une force conservative, une force fluctuative et une force dissipative. La première est identique à celle décrite dans les modèles précédents et ne dépend que de la distance entre les deux grains. Les deux autres forces sont similaires à celles que nous évoquions pour représenter implicitement un solvant. Mais au lieu de s'appliquer de la même façon à toutes les particules, elle sont définies ici pour chaque paires de grains en interaction (qu'il s'agisse du solvant ou non) et ont la direction de vecteur reliant ces deux grains. La force dissipative est alors proportionnelle et opposée à la vitesse relative des deux particules et la force fluctuative a une amplitude aléatoire dont la valeur moyenne dépend de la température. Un intérêt particulier de ce modèle est qu'il vérifie un grand nombre des propriétés physiques importantes (Venturoli *et al.*, 2006). De plus, dans ce modèle, les forces conservatives entre grains sont particulièrement simples à paramétrer. Il

s'agit très souvent d'une simple force de répulsion n'agissant qu'en dessous d'une distance seuil et augmentant linéairement lorsque l'on rapproche les deux particules. Ainsi, le seul paramètre à définir est la force maximale de répulsion pour chaque paire de particules. Malgré la très grande simplification de ce profil d'interaction, les modèles de DPD sont capables de reproduire un grand nombre de phénomènes physiques et sont couramment utilisés (Vattulainen et Karttunen, 2005; Laradji et Sunil Kumar, 2005; Venturoli *et al.*, 2006; Füchslin *et al.*, 2009).

Le modèle de Discrete Molecular Dynamics (DMD), principalement employé pour étudier des propriétés de repliement, de reconnaissance stérique et d'agglomération de protéines, est relativement différent de tous les modèles présentés précédemment car il s'agit d'un algorithme de simulation événementielle (Nguyen et Hall, 2004, 2006; Sharma *et al.*, 2007). Ce formalisme utilise des potentiels de paire discontinus et constants par morceau pour décrire les interactions entre grains (aussi bien intra- que inter-moléculaires), ce qui conduit alors à approximer le mouvement des particules par des trajectoires linéaires par morceau. En effet, d'après la première loi de Newton, tant que le système reste à une même valeur d'énergie potentielle, chaque particule se déplace en ligne droite à une vitesse constante. Dans ces conditions, il est possible, à partir d'une configuration donnée du système (positions et vitesses des grains), de déterminer l'instant de la prochaine rencontre d'une discontinuité de potentiel dans l'ensemble du système. À cet instant, les vecteurs vitesse des particules concernées sont modifiés et, les particules avancent alors de nouveau en ligne droite jusqu'au prochain événement. L'itération de ces calculs constitue l'algorithme événementiel à la base de la DMD. Cependant, une telle approche ne permet pas de représenter certaines molécules de façon implicite. En effet, dans ces approches, les molécules de solvant ne sont pas représentées explicitement et les forces de fluctuation et de dissipation qu'elles provoquent ne sont alors pas prises en compte (Nguyen et Hall, 2004, 2006; Sharma *et al.*, 2007). Les mouvements des molécules sont alors relativement peu réalistes, et ce en particulier, lorsqu'une molécule se déplace seule dans l'espace sans rencontrer d'autres molécules. Pour étudier les propriétés de repliement, les interactions stériques et l'assemblage des protéines, ces mouvements non réalistes n'ont qu'un impact mineur. Mais ils ne conviennent pas pour nos besoins (à savoir par exemple la diffusion d'un facteur de transcription ou d'une molécule d'un transcrit dans le noyau).

***

Ainsi, en adoptant une description simplifiée de la structure des macromolécules, il est possible de décrire les interactions au sein d'une simulation de manière relativement synthétique et avec un niveau de précision choisi arbitrairement. Une représentation simple par des potentiels de paires permet toutefois de conserver une description physique des molécules et de leurs interactions (volume d'exclusion stérique, flexibilité, propriétés d'association/dissociation, ...). Les modèles de coarse graining utilisant des potentiels de paire et une représentation implicite du solvant sont donc particulièrement pertinentes dans le cadre de notre étude et serviront de base pour la définition de notre modèle.

1.4 Approches mésoscopiques

Le concept d'échelle *mésoscopique* n'est pas défini de façon précise et objective, si bien que la plupart des travaux basés sur des modèles de coarse graining peuvent être qualifiés d'approches mésoscopiques (puisqu'il s'agit d'étudier le comportement collectif d'au moins plusieurs dizaines ou centaines d'entités discrètes non atomiques). Par exemple, dans un ENM, les modes normaux sont bien des comportements observés au niveau de la dynamique collective de l'ensemble des grains. Cependant, il ne s'agit là que d'un phénomène intra-moléculaire local et non de l'organisation spatio-temporelle de nombreuses molécules à un niveau plus global. De nombreux modèles de coarse graining sont toutefois couramment employés pour répondre à des questions d'ordre clairement mésoscopique. En particulier, dans le cadre de l'étude des membranes phospholipidiques, de nombreux modèles de coarse graining (qu'il s'agisse de potentiels arbitraires ou de modèles de DPD) sont employés pour étudier des phénomènes physiques typiques de l'échelle mésoscopique. On peut citer, par exemple, l'étude des transitions de phase dans l'arrangement des queues hydrophobes au sein d'une bicouche (en fonction de paramètres tels que la température, la longueur des chaînes carbonées, ...) (Venturoli *et al.*, 2006), des phénomènes de séparation de phase entre différents composants (Shelley *et al.*, 2001b; Laradji et Sunil Kumar, 2005; Füchslin *et al.*, 2009), ou encore des mécanismes de formation/fission/fusion de vésicules (Nielsen *et al.*, 2004; Lyubartsev, 2005; Laradji et Sunil Kumar, 2005; Lales *et al.*, 2005; Venturoli *et al.*, 2006; Füchslin *et al.*, 2009).

Par ailleurs, de nombreux autres phénomènes mésoscopiques intracellulaires sont couramment abordés par des modèles basés sur des formalismes très divers, allant parfois jusqu'à des niveaux d'abstraction importants. Ces modèles, nés de la nécessité de prendre en compte l'espace et le caractère discret des molécules pour rendre compte de nombreux phénomènes à l'échelle cellulaire (Amar *et al.*, 2002; Bork et Serrano, 2005; Andrews et Arkin, 2006), sont souvent plus centrés sur l'influence de la spatialité dans les réactions biochimiques entre molécules que sur des mécanismes d'interaction physique réaliste comme dans les modèles présentés précédemment. Ces aspects sont en effet souvent très simplifiés, voire ignorés, permettant alors d'atteindre des tailles et des durées de simulation bien supérieures à celles des modèles présentés précédemment, mais au détriment du réalisme physique des interactions entre molécules. Ces modèles – et, le cas échéant, les logiciels de simulations associés – sont basés sur des formalismes variés, faisant chacun des approximations différentes (Takahashi *et al.*, 2005; Lemerle *et al.*, 2005). Nous en décrivons ici les principales caractéristiques.

Diffusion et réaction de molécules ponctuelles

L'approche proposée par Andrews et Bray (2004); Andrews *et al.* (2010), constituant la base du logiciel de simulation Smoldyn, consiste à représenter les molécules d'intérêt comme des entités ponctuelles diffusant dans un espace 3D continu. Les molécules se déplacent selon un mouvement brownien sur-amorti (représentant donc le solvant et la plupart des protéines de façon implicite) et sans aucune force d'interaction, mais peuvent cependant provoquer des réactions biochimiques. Pour ce faire, la rencontre de deux molé-

cules est décrite par une distance seuil en dessous de laquelle elles peuvent réagir ensemble. Lorsqu'une réaction a lieu, les produits de réaction sont alors placés différemment en fonction du type des molécules et du type de réaction (soit au centre de masse des réactants, soit séparés d'une distance prédéterminée pour la réaction, soit placés aléatoirement et uniformément dans l'espace de la simulation). Ce modèle permet de montrer que la simple considération de la localisation spatiale de molécules modifie significativement la dynamique du système et est susceptible de créer spontanément des hétérogénéités (Andrews et Bray, 2004). Une telle approche, n'intégrant pas de forces entre molécules, permet alors de simuler des systèmes beaucoup plus grands et des durées beaucoup plus longues que les modèles physiques décrits dans les sections précédentes (Andrews *et al.*, 2010). Les auteurs avancent de plus que des effets d'exclusion stérique entre molécules peuvent être représentés sur la base des réactions chimiques qu'inclut le modèle. Mais il est important de noter qu'il n'y a pas de réelle exclusion stérique par des forces (comme ce serait le cas avec des potentiels ou un modèle de sphères dures) et que le comportement d'un tel modèle pour représenter un système fortement condensé est probablement assez artéfactuel[8].

Un modèle aux capacités relativement similaires, le modèle GFRD (Green's Function Reaction Dynamics) introduit par van Zon et ten Wolde (2005a,b), fait appel à un principe de simulation sensiblement différent, basé sur un algorithme événementiel. Il s'agit, pour une configuration donnée du système (ensemble des positions des molécules), de déterminer la durée maximale du pas de temps permettant aux molécules (diffusant librement selon un mouvement brownien sans forces) de ne se rencontrer que deux par deux. Les nouvelles positions sont alors tirées aléatoirement selon une distribution en accord avec la durée du pas de temps et les paires de molécules potentiellement en interaction pendant cet intervalle de temps réagissent éventuellement entre elles. Cet algorithme permet d'accélérer très significativement les calculs et ce, en particulier lorsque les molécules ne sont que rarement proches les unes des autres. Les auteurs ont appliqué ce modèle à l'étude de l'influence de la diffusion d'un facteur de transcription sur sa dynamique d'association/dissociation sur son site de fixation et ses effets sur la stochasticité de l'expression d'un gène régulé par ce facteur (van Zon *et al.*, 2006). Cependant, d'une part le formalisme utilisé ne permet pas en l'état d'inclure des forces entre molécules et, d'autre par une telle approche événementielle perd tout son intérêt dès lors que les molécules sont souvent proches les unes des autres, comme c'est en particulier le cas si l'on souhaite étudier des phénomènes liés à l'encombrement moléculaire.

Ainsi, il est possible d'étudier des questions relatives à l'effet de la spatialité sur l'expression stochastique des gènes avec de tels modèles. Mais l'absence d'une représentation physique des molécules et de bonnes propriétés d'exclusion stérique dans les modèles GFRD et Smoldyn ne permet pas d'étudier de nombreux phénomènes nucléaires comme l'effet de l'encombrement moléculaire sur les propriétés de réaction entre molécules.

[8] L'apparition d'un produit de réaction dans la zone d'exclusion d'une autre molécule va immédiatement déclencher une autre réaction, conduisant probablement à une nouvelle réaction ... etc. Cette série de déplacements brusques et très artificiels ne correspond pas aux propriétés de diffusion et d'interaction que devraient avoir les molécules et est susceptible de provoquer d'importantes instabilités numériques, imposant alors probablement un pas de temps relativement court.

Encombrement moléculaire : diffusion anormale et cinétiques de réactions

Plusieurs travaux montrent que la présence de nombreux obstacles à la diffusion des molécules est susceptible d'influencer significativement les cinétiques de réactions et les temps de rencontre entre deux molécules ou entre un facteur de transcription et son site de fixation (Berry, 2002; Grima et Schnell, 2006; Echevería *et al.*, 2007; Guigas et Weiss, 2008). Certains de ces travaux se basent en particulier sur une représentation discrète de l'espace sous la forme d'une lattice, une grille régulière (2D ou 3D) décrivant la position des molécules selon un ensemble fini d'emplacements discrets (Berry, 2002; Grima et Schnell, 2006) Des obstacles sont placés aléatoirement et uniformément dans l'espace, conférant alors à l'espace accessible des propriétés fractales et provoquant une diffusion dite anormale des molécules[9]. Il est ainsi possible d'étudier l'influence de cette diffusion sur de nombreux mécanismes tels que la recherche de cibles par un facteur de transcription et la rencontre entre molécules, ainsi que leur influence sur les cinétiques des réactions. La représentation discrète de l'espace sous la forme d'une lattice permet de simuler beaucoup plus efficacement que dans une approche continue et n'a pas, pour ce type de phénomène, d'influence qualitative sur les résultats (Grima et Schnell, 2006). Cependant, dans le cadre plus général de notre questionnement, les modèles basés sur des lattices ne sont pas en mesure de reproduire fidèlement certaines propriétés essentielles (forme et orientation des molécules ...) et présuppose une anisotropie de l'espace.

Il est à noter qu'une approche alternative se basant sur une description continue de l'espace pour étudier l'effet de la diffusion anormale sur les temps de rencontre entre espèces moléculaires consiste à ne pas représenter les obstacles de façon explicite mais à les décrire implicitement. En effet, dans ces approches la marche aléatoire de la particule d'intérêt est obtenue en choisissant une force fluctuative, non pas décorrélée comme dans le cas classique, mais aux propriétés particulières conférant à la diffusion des propriétés anormales (Saxton, 2001; Guigas et Weiss, 2008).

Vers davantage d'abstraction : automates spatiaux

Depuis de nombreuses années, la communauté de la *vie artificielle* développe des modèles, qualifiés de *chimie artificielle*, considérant des molécules abstraites se déplaçant dans l'espace et réagissant les unes avec les autres (Dittrich *et al.*, 2001). Mais cette communauté est fondée sur une problématique beaucoup plus fondamentale que les approches de modélisation de systèmes biologiques que nous considérons ici et cherche davantage à déterminer les principes fondamentaux de la vie qu'à représenter des systèmes biologiques réels (Bedau *et al.*, 2000; Bedau, 2003). Les formalismes employés ont par conséquent souvent un niveau d'abstraction particulièrement élevé. Par exemple, les chimies artificielles sont très souvent basées sur des espaces discrets dans lesquels les molécules ont des comportements relativement stéréotypés (obéissant à un ensemble des règles symboliques) et sont plus vues comme des automates spatiaux que comme des modèles biochimiques.

[9]Un espace fractal possède des régularités et des propriétés d'auto-similarité sur une grande plage d'échelles. Une molécule diffusant dans un tel espace a une diffusion anormale car, contrairement à la diffusion classique, son déplacement quadratique moyen n'augmente pas de façon linéaire avec le temps.

Toutefois, la frontière entre ces deux types de modèles n'est pas toujours franche et les deux communautés se nourrissent l'une l'autre. En particulier, on peut citer un certain nombre d'approches, parfois regroupées sous le nom de Systèmes Multi-Agents[10] (SMA) moléculaires, dans lesquelles des molécules situées dans un espace explicite (discret ou continu) peuvent se déplacer, se lier ou réagir ensemble selon un ensemble de règles reproduisant de façon plus ou moins simplifiée la réalité physique (Amar *et al.*, 2004, 2008; Ballet *et al.*, 2004; Lales *et al.*, 2005; Bernaschi *et al.*, 2007; Ridgway *et al.*, 2008). Ces règles ont pour but de permettre une formalisation et une description algorithmique simple, parfois au prix de leur justification physique.

Les travaux de Soula *et al.* (2005) peuvent être vus comme une approche intermédiaire, combinant à la fois une description partiellement physique des molécules et employant des règles d'interaction relativement simples. Ce modèle considère le noyau cellulaire comme un espace 3D continu dans lequel le déplacement des protéines est décrit par une équation de Langevin et où leurs interactions physiques sont décrites par un modèle de sphères dures (détection de collision et transfert du moment cinétique). Lors d'une collision, les molécules ont de plus la possibilité de s'associer et de former alors des complexes rigides pouvant à tout instant se dissocier. Ce modèle a permis en particulier de reproduire des structures aux caractéristiques morphologiques et dynamiques similaires à celles que l'on observe dans les noyaux cellulaires par des techniques de microscopie à fluorescence.

*
* *

Il existe ainsi de nombreux modèles aux formalismes très différents pour représenter des phénomènes mésoscopiques intracellulaires. Mais ceux-ci se placent souvent dans deux cadres relativement distincts. Le premier considère des interactions physiquement réalistes entre molécules (forces distantes, associations/dissociations souples), nécessaires pour reproduire de nombreux phénomènes de structuration spatiale, mais ne permettant pas d'atteindre des durées de simulation suffisamment longues et ne considérant pas de réactions chimiques entre molécules. Le second se base au contraire sur les réactions entre des molécules diffusant dans l'espace, souvent décrites de façon très simplifiée et ne permet pas de représenter des forces. Il semble pourtant très probable que la compréhension de nombreux phénomènes mésoscopiques puisse nécessiter de prendre en compte ces deux aspects de la réalité. Nous tentons ici de répondre à un tel besoin par le modèle que nous présenterons dans la section 2. Il est toutefois à noter que deux formalismes proposés récemment pourraient, dans ce cadre, s'avérer particulièrement prometteurs.

Développements récents : les approches hybrides physiques/réactives

Dans le formalisme Multiparticle Collision dynamics (MPC dynamics) (Kapral, 2008), les molécules sont représentées sous la forme de particules se déplaçant dans un espace continu, mais leurs interactions sont en partie calculées à l'aide d'une grille 3D. En effet,

[10]La définition d'un SMA est assez ambiguë et est souvent confondue avec la *modélisation individu-centrée*, cadre dont on peut argumenter qu'il regroupe en réalité tous les modèles décrits dans ce chapitre faisant intervenir des entités discrètes interagissant localement les unes avec les autres.

les collisions entre particules ne sont pas calculées explicitement : à chaque pas de temps, les vecteurs vitesse des particules sont modifiés en fonction du contenu de la cellule où elles se trouvent, selon un algorithme dérivé de principes physiques et mécaniques et contenant une part d'aléatoire. Un intérêt de ce formalisme est qu'il préserve un certain nombre de propriétés physiques importantes de la mécanique Newtonnienne classique. De plus, en évitant les calculs de collision, il rend la simulation beaucoup plus efficace et permet de simuler des systèmes de grande taille et sur des durées suffisantes pour reproduire des phénomènes hydrodynamiques et de dynamique des fluides (Kapral, 2008). Ce modèle, computationnellement efficace et physiquement réaliste pour notre échelle d'intérêt, peut de plus être combiné avec des méthodes basées sur des potentiels d'interaction entre molécules et intégrer des réactions chimiques (Kapral, 2008). Il a en particulier été utilisé pour étudier des phénomènes mésoscopiques tels que la diffusion et la réaction en milieu encombré (Echevería *et al.*, 2007) et pour reproduire des phénomènes de comportement de phase et de structuration sous la forme de membranes (Tüzel *et al.*, 2007). Ces méthodes hybrides en sont encore à un stade peu avancé (Kapral, 2008), mais il semble probable que de nombreux phénomènes mésoscopiques ayant lieu dans les noyaux cellulaires puissent être étudiés par un tel formalisme dans les années à venir.

Enfin, le formalisme de DPD (modèle réduit de coarse graining, décrit dans la section 1.3, à la page 73) emploie classiquement une description des macromolécules par plusieurs grains (typiquement une dizaine de grains pour un phospholipide) et décrit leurs interactions par des champs de potentiels (représentant donc des forces entre particules), mais ne permettant pas de réactions chimiques entre molécules. Il a cependant l'avantage, tout comme le MPC dynamics, de respecter de nombreuses propriétés physiques et de permettre des calculs efficaces. Füchslin *et al.* (2009) ont récemment proposé une extension du formalisme de DPD en multipolar reaction DPD (mprDPD). Les particules sont décrites de façon anisotrope, permettant alors de représenter des macromolécules de façon beaucoup plus simplifée (par exemple, un grain pour un phospholipide) tout en conservant certaines propriétés géométriques essentielles permettant par exemple de reproduire des phénomènes de tension de surface d'une membrane. De plus les molécules peuvent ici réagir entre elles, ouvrant alors de nombreuses perspectives d'utilisation pour un tel modèle. Ces propriétés permettent d'accélérer encore les temps de calcul et d'atteindre des durées et des tailles de simulation probablement pertinentes pour étudier des phénomènes d'organisation mésoscopique au sein du noyau cellulaire.

Ainsi, mis à part les formalismes encore très récents de MPC dynamics et de mprDPD, la plupart des modèles mésoscopiques intracellulaires ne permettent pas à la fois de représenter la réalité physique des interactions entre molécules, de décrire des mécanismes de réactions biochimiques et de simuler des systèmes sur des échelles de taille et de temps suffisamment importantes pour aborder des questions relatives à l'organisation nucléaire et à sa relation au caractère stochastique de l'expression des gènes. Nous proposons maintenant un modèle tentant de répondre à ce besoin.

2 Simulateur mésoscopique 4D de phénomènes nucléaires

Le travail présenté ici correspond aux phases de conception et d'implémentation d'un modèle de simulation (cette seconde phase étant par ailleurs encore en cours de réalisation). Il s'intègre dans le cadre d'un projet mettant en œuvre une démarche itérative de modélisation. Il est donc important de tout d'abord mentionner le contexte de cette étude.

La problématique de départ de ce projet[11] concerne en premier lieu la modélisation de phénomènes d'auto-organisation au sein du noyau cellulaire, motivant le développement des « modèles 4D » (trois dimensions spatiales et une temporelle), prenant en compte les aspects essentiels que sont la localisation spatiale et la nature discrète des molécules, ainsi que leurs interactions physiques et biochimiques (Bork et Serrano, 2005; Takahashi *et al.*, 2005; Lemerle *et al.*, 2005; Andrews et Arkin, 2006). Cette problématique a de plus évolué pour y donner une place plus importante au caractère stochastique des processus biochimiques en relation à l'organisation nucléaire, au premier plan desquels l'expression génique.

Versions antérieures

Une première version du modèle et du logiciel de simulation associé (nommé 3DSPI pour *3-dimensional Dynamic Simulator of Protein-Protein Interactions*) a été développée précédemment dans le cadre de ce projet (Soula *et al.*, 2005, travaux évoqués dans la section précédente; page 78). Les protéines y sont représentées comme des particules, dotées d'un rayon et d'une masse, se déplaçant selon un mouvement de LD (cf figure II.4) et n'interagissant que lors de leur collision : soit par un rebond décrit par un modèle de sphères dures (transfert du moment cinétique), soit en s'associant par une liaison rigide et indéformable. Ces deux comportements possibles lors d'une collision sont choisis aléatoirement selon une probabilité préalablement définie. De plus, à chaque pas de temps, chaque liaison entre deux protéines a une probabilité donnée d'être rompue. Dans ce modèle, les associations/dissociations sont donc binaires et rigides et les molécules sont isotropes[12], n'interagissent que lors de leur contact (donc sans forces distantes) et ne réagissent pas ensemble.

Nous avons alors développé un second logiciel de simulation (nommé 3DSPI v2), implémentant un formalisme relativement différent du précédent (Coulon *et al.*, 2007, 2008) et s'inspirant fortement des techniques de simulation aux échelles atomiques et coarse grain décrites dans les sections 1.2 et 1.3. Nous ne le décrirons cependant que brièvement dans le cadre de ce manuscrit afin de nous concentrer sur la troisième version de ce logiciel, la plus récente, qui en constitue une généralisation sur de nombreux aspects. Dans la seconde version, les protéines sont représentées par des particules se déplaçant selon un mouvement de BD et interagissent par des potentiels de paire (figure II.3) dont la forme est décrite par l'expression (II.1), celle-ci étant paramétrée en accord avec l'échelle de la représenta-

[11] Projet mené conjointement par des informaticiens (coordonnés par le Pr. Guillaume Beslon au laboratoire LIRIS) et des biologistes (coordonnés par le Dr. Olivier Gandrillon au laboratoire CGMC).

[12] Notons qu'une extention du logiciel de simulation a par la suite permis d'y représenter les protéines de façon anisotrope par plusieurs particules liées de façon rigide (réalisé par Kevin Frugier).

tion. L'intégration numérique de l'équation du mouvement est faite de façon asynchrone et avec un pas de temps adaptatif[13]. Cependant, un tel algorithme rend particulièrement délicate l'implémentation et n'a pas d'intérêt lorsque le système est fortement condensé (comme c'est le cas si l'on souhaite représenter l'encombrement moléculaire). Il a donc été abandonné dans la nouvelle version du logiciel, au profit d'un intégrateur beaucoup plus efficace que nous décrirons à la section 2.2.

Il est à noter que le logiciel 3DSPI v2 a été étendu afin d'y inclure la possibilité de réactions chimiques entre les molécules[14]. Ces travaux ont permis mettre en avant certaines spécificités relatives aux réactions dans un contexte d'interactions par potentiels de paire et de proposer certains choix de modélisation et d'implémentation.

Nous présentons maintenant un troisième logiciel de simulation réalisé dans ce cadre (nommé 3DSPI v3), s'appuyant sur différents concepts de modélisation décrits dans la section 1.

2.1 Modèle(s) : définition d'un formalisme général

Il est d'usage dans certaines communautés de ne pas faire de différence entre un modèle et son implémentation. C'est en particulier le cas dans la communauté informatique qui considère souvent des modèles dont la description est très proche des choix d'implémentation (voir par exemple les SMA, page 78). Modèle et logiciel de simulation sont alors souvent conçus conjointement. Il est cependant préférable, dans de nombreux cas et comme nous le ferons ici, de séparer ce qui relève du modèle (définition du système et de son comportement de base) et ce qui relève de son implémentation (réalisation d'un logiciel permettant de simuler numériquement le système en question). En effet, idéalement, le comportement d'un modèle doit être indépendant de son implémentation et le fait d'utiliser un certain algorithme plutôt qu'un autre ne doit pas influencer le résultat final d'une simulation.

De plus, nous considérons ici un formalisme volontairement générique, permettant de représenter des modèles relativement différents. Comme nous le verrons, il permet par exemple de décrire la vibration d'une protéine par un ENM, l'interaction par des potentiels arbitraires d'un ensemble de molécules coarse-grain se déplaçant en LD ou en BD, voire la dynamique d'un fluide de particules de Lennard-Jones se déplaçant par des mouvements balistiques[15]. Il est donc important de souligner que, bien que de nombreux choix de modélisation restent ouverts à l'utilisateur du logiciel, ceux-ci doivent être déterminés par la nature du phénomène étudié et non par les possibilités qu'offre l'approche.

[13]Pour chaque particule, le pas de temps maximal ne produisant pas une erreur supérieure à un seuil préalablement fixé est déterminé en fonction de la proximité des autres particules. Cet algorithme permet de simuler rapidement le déplacement des protéines seules et de décrire de façon fine le mouvement des protéines en interaction. Pour davantage de détails, le lecteur pourra se référer à (Coulon *et al.*, 2007).

[14]Travail réalisé par David Parsons.

[15]Un mouvement balistique est celui qu'ont des particules dans le vide. Il s'agit simplement de la loi de Newton classique, ne prenant pas en compte de forces fluctuatives et dissipatives due à un solvant.

Niveau de description et vue globale

Nous employons une description coarse grain dans laquelle les macromolécules sont représentées comme un système de plusieurs particules (notons que l'on pourra parfois utiliser le terme « protéine » pour désigner une macromolécule, même s'il peut en réalité s'agir également d'ADN, de transcrits, ...). Cette représentation permet de décrire, avec un niveau de précision choisi arbitrairement, la structure géométrique des protéines en représentant leurs différents domaines, ceux-ci ayant éventuellement des propriétés d'interaction spécifiques avec les autres molécules (figure II.5). Nous parlerons ainsi souvent de *domaines* pour qualifier ces particules. Une protéine peut alors aussi bien être représentée de façon très simple comme une unique particule isotrope ou de façon plus détaillée comme un ensemble de domaines.

La dynamique intra-moléculaire entre ces différents domaines est régie par un ENM, permettant ainsi de reproduire les propriétés mécaniques de flexibilité et modifications conformationnelles des molécules (cf section 1.3). Dans ce cadre les approches de description hiérarchique des protéines par un ENM (Doruker *et al.*, 2002; Kurkcuoglu *et al.*, 2009; Kim *et al.*, 2009), ainsi que les méthodes permettant d'identifier les différents domaines protéiques comme des blocs rigides (Kundu *et al.*, 2004; Potestio *et al.*, 2009; Abyzov *et al.*, 2010) peuvent être d'un intérêt tout particulier pour paramétrer le modèle. Les interactions inter-moléculaires sont évaluées entre paires de domaines et sont décrites par des potentiels de paire de formes arbitraires. Cela permet alors de représenter aussi bien des interactions de Lennard-Jones et de Coulomb telles que celles décrites par l'expression (II.1) que des potentiels statistiques ou empiriques (cf page 72) ou encore des potentiels au contraire très simplifiés dans une approche de modèle réduit (figure II.5c2).

Enfin, le modèle proposé emploie une représentation implicite du solvant (et des molécules ne nécessitant pas d'être prises en compte individuellement) se traduisant par des forces fluctuatives et dissipatives telles que décrites précédemment. Le déplacement des particules s'effectue de plus dans un espace torique[16] et, les modèles de forces et de réactions entre molécules que nous décrirons par la suite permettent de définir très simplement une frontière réflexive et/ou absorbante[17] de forme sphérique.

Structure des macromolécules

On définit tout d'abord un ensemble de *types de domaines* (figure II.5a), dotés chacun d'une masse et d'un rayon (notons que ce rayon n'est utilisé que pour évaluer la force dissipative, le volume d'exclusion stérique de la particule étant déterminé par ses potentiels de paire). On définit ensuite des *espèces moléculaires* comme des ensembles de domaines, chacun étant de l'un des types de domaines définis précédemment (figure II.5b). Leurs

[16]Dans un espace torique, on définit que toute particule sortant de l'espace en en traversant une frontière (par exemple une des faces du cube si l'espace est cubique) re-rentre par une autre frontière (la face opposée du cube). Il s'agit d'une technique permettant de décrire un espace pseudo-infini sans limite. On parle également de frontières périodiques.

[17]Les particules tentant de traverser une frontière de l'espace rebondissent si celle-ci est réflexive et disparaissent si elle est absorbante.

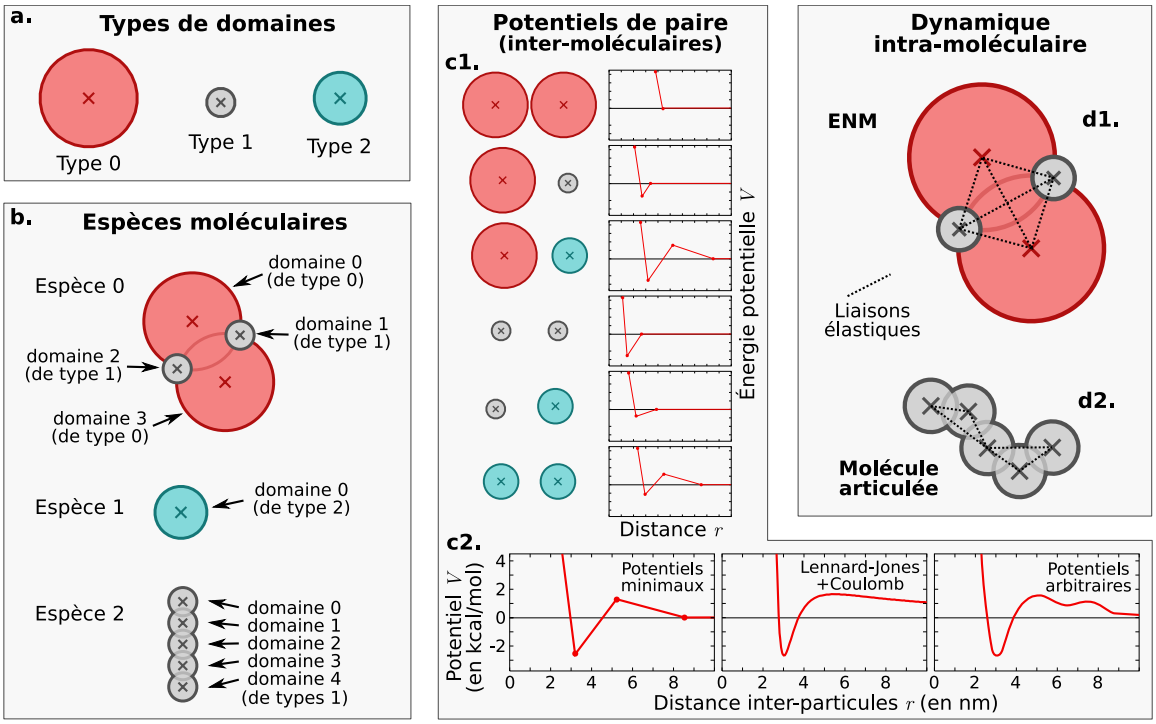


Fig. II.5: Définition des macromolécules et de leurs interactions. Cette figure illustre, par un exemple très simple d'instanciation du modèle, la représentation des molécules et de leurs comportements. L'unité de description du modèle que nous présentons ici est le domaine moléculaire. Un ensemble de *types de domaines* (a) est alors défini, à partir duquel sont construites les différentes *espèces moléculaires* (b). Chacune de ces espèces est constituée d'un nombre arbitraire de domaines de chaque type, placés dans un repère de coordonnées locales, définissant alors la structure de référence de la macromolécule. (c) Les interactions inter-moléculaires sont représentées par des potentiels d'interaction de paire (un pour chaque paire de types de domaines ; c1) et définis de façon arbitraire par des fonctions linéaires par morceau. (c2) Cette représentation permet aussi bien d'employer une approche minimaliste dans la description des interactions que de représenter celles-ci de façon beaucoup plus fine comme dans différentes approches évoquées dans la section 1 (par exemple, par des potentiels de Lennard-Jones ou par des potentiels obtenus selon des méthodes physiques ou statistiques). (d) La dynamique intra-moléculaire est définie par un ensemble de liaisons élastiques dont la longueur au repos est celle de la conformation de référence et dont l'élasticité est un paramètre indépendant défini pour chaque paire de domaine. Cela permet alors de représenter un ENM (d1) ou, de façon plus générale, une molécule articulée avec plus ou moins de degrés de liberté (d2).

positions les uns par rapport aux autres sont définis dans un système de coordonnées locales où $\overline{\mathbf{x}}_\alpha^p$ représente la position du domaine α dans une molécule d'espèce p.

L'espace de la simulation contient différentes macromolécules, chacune étant de l'une des espèces définies précédemment et donc composée de différents domaines, chacun étant de l'un des types définis précédemment. On note $\mathbf{x}_i$ la position dans l'espace de la particule i. De plus, $r_{ij} = \|\mathbf{x}_j - \mathbf{x}_i\|$ représente la distance entre les particules i et j, et $\mathbf{u}_{ij} = (\mathbf{x}_j - \mathbf{x}_i)/r_{ij}$ le vecteur unitaire allant de i vers j.

Potentiels et forces d'interaction

La force entre deux particules i et j dépend du type de domaine qu'elles représentent, ainsi que de leur appartenance ou non à une même molécule. Dans tous les cas, cette force peut être décrite simplement comme l'opposé du gradient d'un potentiel de paire V_{ij} entre les deux particules. De plus, ce potentiel ne dépendant que de la distance r_{ij} entre les deux particules, la force exercée par la particule j sur la particule i vaut alors simplement

$$-\boldsymbol{\nabla}_{\mathbf{x}_i}\big[V_{ij}\big] = \frac{\partial V_{ij}}{\partial r_{ij}}\,\mathbf{u}_{ij} \tag{II.2}$$

(où $\boldsymbol{\nabla}_{\mathbf{x}}\big[V\big]$ représente le gradient du champ scalaire V en fonction de $\mathbf{x}$). Il s'agit ainsi toujours d'une force dont la direction est celle du vecteur $\mathbf{u}_{ij}$ et dont l'amplitude est la dérivée du potentiel V_{ij} en fonction de r_{ij} (cf figure II.3).

Pour chaque paire de types de domaines, un potentiel de paire est défini comme paramètre du modèle et décrit par une fonction linéaire par morceaux (représentée par un ensemble de points ; cf (figure II.5)c1). Notons dès maintenant que, dans l'implémentation du logiciel de simulation, le nombre de points utilisés pour décrire le potentiel n'a aucune influence sur l'efficacité de simulation, si bien qu'il n'est pas limitant de décrire de cette façon des potentiels très précis (figure II.5c2). Dans le cas où les deux domaines i et j appartiennent à deux molécules différentes (interactions inter-moléculaires), le potentiel V_{ij} est simplement celui évoqué précédemment et est déterminé en fonction du type des deux domaines.

Lorsque les domaines i et j appartiennent à une même macromolécule (interactions intra-moléculaires), le potentiel V_{ij} est défini par un ENM (figure II.5d1). Il s'agit d'un réseau de liaisons élastiques reliant les différents domaines d'une même protéine et dont les longueurs au repos sont celles de la conformation de référence (cf section 1.3). Si les particules i et j correspondent respectivement aux domaines α et β d'une même macromolécule d'espèce p (chacun de coordonnées locales $\overline{\mathbf{x}}^p_\alpha$ et $\overline{\mathbf{x}}^p_\beta$ dans la conformation de référence), alors leur interaction est définie par le potentiel

$$V_{ij} = \frac{C^p_{\alpha\beta}}{2}\left(r_{ij} - \|\overline{\mathbf{x}}^p_\beta - \overline{\mathbf{x}}^p_\alpha\|\right)^2 \tag{II.3}$$

où $C^p_{\alpha\beta}$ est le coefficient d'élasticité de la liaison, défini en paramètre. Le facteur $\frac{1}{2}$ provient simplement du fait que la force est obtenue en dérivant le potentiel. En effet,

$$-\boldsymbol{\nabla}_{\mathbf{x}_i}\big[V_{ij}\big] = C^p_{\alpha\beta}\left(r_{ij} - \|\overline{\mathbf{x}}^p_\beta - \overline{\mathbf{x}}^p_\alpha\|\right)\mathbf{u}_{ij} \tag{II.4}$$

Pour une espèce moléculaire p donné, les coefficients d'élasticité $C^p_{\alpha\beta}$ entre chaque paire de domaines peuvent être définis très simplement comme étant tous égaux à une même valeur C^p ou, de manière plus précise, par une matrice $\mathbf{C}^p$ permettant alors de représenter différents systèmes. En particulier, il est possible, comme dans la plupart des modèles d'ENM, de ne considérer que les interactions entre les domaines plus proches qu'une certaine distance seuil r_c dans la conformation de référence :

$$C^p_{\alpha\beta} = C^p\,H(r_c - \|\overline{\mathbf{x}}^p_\beta - \overline{\mathbf{x}}^p_\alpha\|) \tag{II.5}$$

($H(\cdot)$ est la fonction de Heaviside[18]). On peut également, en paramétrant la matrice $\mathbf{C}^p$, décrire une molécule articulée avec des élasticités variées et un nombre arbitraire de degrés de liberté (figure II.5d2). Notons que, bien que ce formalisme ne permette pas de représenter la torsion des liaisons (comme c'est par exemple le cas en MD), il est toutefois souvent possible de représenter un comportement relativement similaire avec des liaisons élastiques.

Loi(s) du mouvement

L'ensemble des forces s'appliquant sur une particule i comprend non seulement les forces provenant de ses interactions avec les autres particules (soit $\sum_{j\neq i} -\boldsymbol{\nabla}_{\mathbf{x}_i}\left[V_{ij}\right]$, où la somme sur j est faite parmi toutes les particules autres que i), mais également les forces dissipatives et fluctuatives dues à la représentation implicite du solvant. Ainsi, la seconde loi de Newton, appliquée à la particule i, s'écrit

$$m_i\ddot{\mathbf{x}}_i = -\lambda_i\dot{\mathbf{x}}_i + \boldsymbol{\xi}_i + \sum_{j\neq i} -\boldsymbol{\nabla}_{\mathbf{x}_i}\left[V_{ij}\right] \tag{II.6}$$

où m_i représente la masse du domaine i (déterminée par son type) et $\dot{\mathbf{x}}_i$ et $\ddot{\mathbf{x}}_i$ sont respectivement sa vitesse et son accélération. La force dissipative $-\lambda_i\dot{\mathbf{x}}_i$ décrit le fait que le milieu fait opposition au déplacement de la particule et est opposée et proportionnelle à sa vitesse. Le coefficient de viscosité λ_i dépend du rayon r_i du domaine (déterminé par son type) selon la formule classique reliant ces deux quantités dans le cas d'une sphère :

$$\lambda_i = 6\pi r_i\eta \tag{II.7}$$

où η est la viscosité dynamique du milieu, un paramètre global du modèle. La force fluctuative $\boldsymbol{\xi}_i$ provient du choc aléatoire des molécules de solvant sur la macromolécule. Il s'agit d'un processus stochastique gaussien à 3 dimensions, décorrélé dans le temps et d'une particule à l'autre. Son autocovariance est

$$\left\langle \boldsymbol{\xi}_i(t)\boldsymbol{\xi}_j(t+\tau)\right\rangle_t = 6\lambda_i \mathrm{k_B} T\delta_{ij}\delta(\tau) \tag{II.8}$$

où $\mathrm{k_B}$ est la constante de Boltzmann, T la température du milieu et δ_{ij} et $\delta(\tau)$ sont les fonctions de Kronecker et de Dirac.

L'équation du mouvement (II.6) est une équation de Langevin décrivant un comportement de type inertiel sur les temps courts et diffusif sur les temps longs (LD sur la figure II.4). Si l'on fait de plus une hypothèse de sur-amortissement consistant à négliger le terme $m_i\ddot{\mathbf{x}}_i$ face au terme $-\lambda_i\dot{\mathbf{x}}_i$, l'équation du mouvement devient alors

$$\lambda_i\dot{\mathbf{x}}_i = \boldsymbol{\xi}_i + \sum_{j\neq i} -\boldsymbol{\nabla}_{\mathbf{x}_i}\left[V_{ij}\right] \tag{II.9}$$

et décrit un comportement uniquement diffusif (BD sur la figure II.4). En effet, la masse m_i et le coefficient de viscosité λ_i d'une particule étant respectivement proportionnels à ${r_i}^3$

[18]Fonction de Heaviside : $H(x) = 0$ si $x < 0$ et 1 sinon.

et à r_i, plus les particules considérées sont de petite taille, plus leur mouvement s'approche du régime sur-amorti, pour une durée d'observation donnée. Mais sur des temps courts, le mouvement n'est décrit de façon réaliste que par l'équation (II.6).

Le logiciel que nous proposons permet de simuler les deux variantes du modèle considérant chacune des deux lois du mouvement (II.6) et (II.9). Il est par ailleurs possible, en employant la loi du mouvement (II.6) avec une viscosité dynamique du milieu nulle $\eta = 0$, de décrire un système sans solvant dans lequel les particules ont alors des mouvements purement balistiques.

Réactions biochimiques

Toutes les fonctionnalités du modèle que nous venons de décrire sont actuellement implémentées dans le logiciel de simulation 3DSPI v3. Ça n'est cependant pas le cas des réactions chimiques entre molécules. Cet aspect du modèle est par ailleurs particulièrement difficile à définir de façon indépendante de l'implémentation car il est fréquent, dans ce type d'approches, que la réalisation du logiciel révèle un comportement inattendu, remettant alors en cause certains choix de modélisation. Nous ne présentons donc ici que certains concepts généraux et non un modèle précis pouvant être directement implémenté.

Dans ce cadre, le travail de David Parsons, ayant consisté à intégrer des réactions biochimiques dans la version antérieure du logiciel 3DSPI, constitue une aide précieuse. Un phénomène particulier mis en évidence par ces travaux est le fait que, dans un contexte où les molécules interagissent par des forces, il est possible qu'un produit de réaction apparaisse dans la région d'exclusion d'une autre particule (région où le potentiel est très élevé et très décroissant et qu'il est très peu probable d'atteindre par simple diffusion). Dans ce cas, les deux particules exercent brutalement l'une sur l'autre une force de répulsion très élevée et se trouvent artificiellement expulsées. Il est alors nécessaire de prendre en compte ce phénomène dans le positionnement spatial des produits de réaction et dans la modélisation de leur apparition. Notons que le modèle considéré dans cette version du logiciel de simulation constitue un cas particulier du modèle présenté ici dans lequel les macromolécules ne sont représentées que par une unique particule[19]. Le contexte multidomaine que nous considérons ici rend significativement plus complexe la description des réactions et impose également de faire certains choix différemment. Nous décrivons donc ici brièvement les contraintes importantes à prendre en compte dans ce cadre.

Le modèle doit permettre de décrire des réactions d'ordre[20] 0, 1 et 2. Le premier cas est décrit par un taux instantané de synthèse des molécules (alors placées de façon uniforme dans l'espace). Les deux autres cas sont décrits par un taux s'appliquant respectivement à chaque molécule de réactant et à chaque paire de molécule de réactant se trouvant plus proches qu'une certaine distance d'interaction préalablement définie. Dans ce dernier cas, on peut définir de façon plus précise une distance seuil pour plusieurs paires de domaines

[19] De plus, dans cette version les molécules interagissent nécessairement par des potentiels de Lennard-Jones et Coulomb et ne peuvent se déplacer que selon un mouvement de BD (Coulon *et al.*, 2007, 2008).

[20] Rappelons que l'ordre d'une réaction est simplement le nombre de molécules de réactant. Par exemple $\varnothing \to A$ est d'ordre 0, $A \to B$, d'ordre 1 et $A + B \to A + C$, d'ordre 2.

entre les deux réactants, permettant alors de contraindre leurs orientations respectives. Il pourrait également être pertinent d'employer une seconde distance seuil en dessous de laquelle deux domaine ne doivent pas être pour que la réaction ait lieu.

Afin non seulement de minimiser les cas risquant de provoquer le phénomène d'expulsion brutale mentionné précédemment, mais également pour permettre un bon niveau de réalisme du modèle, il est important d'identifier au maximum l'origine des produits de réaction parmi les réactants. Par exemple, dans la réaction $A + B \rightarrow A + C$, il est important de pouvoir décrire que la molécule C prend la place de la molécule B et que la molécule A n'est pas transformée et ne doit donc pas être déplacée par la réaction. Les caractéristiques morphologiques précises des molécules doivent alors également être prises en compte. En effet, dans de nombreux cas, les molécules B et C peuvent être morphologiquement très proches. Elle peuvent par exemple être identiques à un domaine près, celui-ci ne voyant que ses propriétés d'interaction avec d'autres molécules modifiées par la réaction (c'est par exemple le cas lors de la méthylation d'un site sur un brin d'ADN ou de la phosphorylation d'un acide aminé sur une protéine). Il est alors capital que la position de tous les domaines de C soit celle des domaines de B afin de ne pas détruire les structures dans lesquelles la macromolécule pourrait être engagée. Dans les cas où il n'est pas possible d'associer chaque domaine des produits avec un domaine des réactants (par exemple les réactions de synthèse et de dégradation), il reste important de maîtriser le placement de ces particules dans l'espace. On pourra par exemple définir un repère local à partir des positions des réactants (en tirant parti de leur anisotropie[21]) et y définir l'emplacement et l'orientation des produits de réaction.

De plus, afin d'éviter l'apparition brutale d'un produit de réaction dans la région d'exclusion d'une autre particule dans le cas où certains produits ne sont pas associés à des réactants et *vice versa*, une solution pourrait consister à ne pas faire disparaître les réactants et apparaître les produits de façon instantanée. On peut en effet définir un courte fenêtre temporelle pendant laquelle les produits et les réactants sont présents simultanément (mais n'interagissent pas ensemble) et où les potentiels d'interaction des produits avec toutes les autres particules sont pondérés par une fonction allant progressivement de 0 à 1 (et ceux des réactants par la fonction complémentaire allant de 1 à 0).

Ces différents concepts ne sont pas encore à un stade suffisamment avancé pour permettre leur implémentation et nécessitent d'être précisés et testés préalablement.

[21] On peut par exemple déterminer un repère local à partir d'un ensemble choisi de domaines. Si cet ensemble de domaine a trop ou trop peu de degrés de liberté, on pourra respectivement y introduire une composante aléatoire ou chercher le repère produisant le meilleur aligement avec les conformations de référence des réactants.

2.2 Implémentation et logiciel de simulation

Le modèle décrit précédemment est implémenté sous la forme du logiciel de simulation 3DSPI v3. Encore en cours de développement, ce logiciel intègre toutes les caractéristiques du modèle décrit précédemment à l'exception des réactions biochimiques[22] et, comme nous le verrons, il n'est encore que partiellement optimisé. Nous décrivons ici brièvement les aspects pratiques relatifs à son implémentation et à son optimisation, ainsi que les facteurs influençant ses performances.

Aspects pratiques

Le cœur du logiciel de calcul est écrit en C et constitue un module Python utilisant des structures de données SciPy[23]. Toutes les entrées/sorties (paramètres, résultats des simulations, ...) sont gérées par des scripts Python, facilitant grandement à la fois le paramétrage des simulations, la définition des conditions initales, la sauvegarde des résultats des calculs et leur analyse. L'implémentation fait intervenir trois logiciels distincts :

- un logiciel destiné exclusivement à la simulation numérique, permettant de tirer parti au maximum des performances de calcul. Il comprend en particulier des fonctionnalités d'exploration paramétrique et de distribution des simulations sur des architectures de types machine multicore, cluster ou grille de calcul.
- un logiciel de visualisation graphique du résultat des simulations, simple lecteur permettant de naviguer dans une scène en 4D à différentes échelles spatio-temporelles.
- une boîte-à-outils pour la manipulation des fichiers d'entrée/sortie (intialisation des simulations, extraction de frames, affichage du log d'une simulation ...).

Ces logiciels permettent à la fois une utilisation simple et flexible (développement de prototypes pour tester des hypothèses, analyse visuelle du résultat, ...) et une approche de calcul scientifique intensif (explorations paramétriques et répétitions, analyse avancée par des outils de calcul numérique et statistique ...).

Optimisations et algorithme d'intégration numérique

Bien que l'implémentation d'un modèle en un simulateur numérique ne doive pas avoir d'influence sur le comportement du système, certains aspects peuvent cependant jouer sur la vitesse des calculs et donc sur les possibilités qu'offre le simulateur en termes de taille et de durée des simulations.

Pour le calcul des forces entre particules, le simulateur utilise les méthodes classiques employées dans la plupart des logiciels de simulation moléculaire (Allen et Tildesley, 1991). Il s'agit de définir pour chaque type d'interaction (donc ici pour chaque paire de types de

[22] Notons que l'architecture du logiciel a toutefois été prévue de façon à permettre l'intégration de ce type de fonctionnalités (quantités variables de molécules au cours du temps, ...).

[23] SciPy est une librairie Python destinée au calcul numérique scientifique (www.scipy.org). Il constitue un équivalent libre du logiciel Matlab et dispose de nombreux outils permettant d'analyser les résultats des simulations.

domaines) une table de référence pré-calculée (*lookup table*) permettant d'éviter certains calculs coûteux (en particulier celui d'une racine carrée). En effet, cette table, notée ici F, renvoie la valeur du module de la force à appliquer (cf équation (II.2)) normalisée par la distance inter-particules et ce, en fonction[24] du carré de cette distance (et non de la distance elle-même). Ainsi, le calcul de la force $-\boldsymbol{\nabla}_{\mathbf{x}_i}\left[V_{ij}\right]$ exercée par j sur i est simplement $F(\mathbf{d}^2)\,\mathbf{d}$ avec $\mathbf{d} = \mathbf{x}_j - \mathbf{x}_i$ (et non un calcul de la forme $F(\sqrt{\mathbf{d}^2})\,\mathbf{d}/\sqrt{\mathbf{d}^2}$) évitant donc le calcul coûteux de $\sqrt{\mathbf{d}^2}$ et la renormalisation du vecteur $\mathbf{d}$.

Le temps d'exécution d'une simulation peut également être sensiblement amélioré en optimisant l'algorithme de recherche des paires de particules en interaction. En effet, la plupart du temps, les potentiels de paire peuvent être considérés comme nuls au delà d'une certaine distance. Afin d'éviter d'évaluer les distances entre chaque paire de particules de la simulation, il existe différentes techniques permettant d'éliminer l'essentiel des paires se trouvant hors de portée l'une de l'autre (Allen et Tildesley, 1991; Frenkel et Smit, 2002). Une technique relativement classique consiste à définir sur l'espace une grille tridimensionnelle permettant d'accéder uniquement aux particules contenues dans une cellule de la grille. Ainsi, pour l'interaction d'une particule donnée avec les particules voisines, seules les cellules suffisamment proches sont considérées. Cette méthode, que nous avions déjà implémentée dans 3DSPI v2, est en cours d'implémentation dans la version 3 du logiciel mais n'est pas totalement fonctionnelle à l'heure actuelle.

L'intégration numérique d'une équation du mouvement consiste en l'approximation d'un processus continu par un ensemble fini d'étapes discrètes, représentant chacune une durée Δt. Il s'agit donc d'un aspect très important de l'implémentation puisqu'en dépendent à la fois la vitesse de calcul et la précision de la simulation. Cette problématique constitue par ailleurs un champ de recherche en soi, s'appliquant à la simulation physique à des échelles très variées (Hairer *et al.*, 2003; Hairer, 2005). Une méthode très classique pouvant être appliquée à n'importe quelle équation différentielle décrivant un processus temporel est l'approximation d'Euler[25]. Bien que très simple et très générale, cette méthode est relativement peu efficace dans notre cas. D'autres méthodes tirent au contraire parti des spécificités de l'équation du mouvement et permettent d'utiliser un pas de temps Δt sensiblement plus long, accélérant ainsi la vitesse de calcul. Il existe ainsi de nombreuses méthodes adaptées aux différentes structures de l'équation du mouvement (Allen et Tildesley, 1991; Groot et Warren, 1997; Paterlini et Ferguson, 1998; Beard et Schlick, 2000; Frenkel et Smit, 2002; Vattulainen et Karttunen, 2005). En particulier, les algorithmes de Verlet et velocity-Verlet (VV) (Allen et Tildesley, 1991; Frenkel et Smit, 2002) sont très couramment employés en MD, tandis que la DPD utilise une variante de l'algorithme velocity-Verlet (Groot et Warren, 1997; Vattulainen et Karttunen, 2005). Il existe également différentes méthodes pour intégrer une équation de Langevin comme celle que nous considérons ici (équation (II.6)) (Allen et Tildesley, 1991; Paterlini et Ferguson, 1998; Beard et Schlick, 2000). Dans le logiciel de simulation décrit ici, nous employons la méthode de Langevin Dynamics Velocity-Verlet (LDVV) (Paterlini et Ferguson, 1998),

[24] Ce tableau est évidemment indexé de façon discrète par des valeurs entières et non de façon continue par des valeurs réelles, mais sa taille (10^4 éléments dans nos simulations) permet de le considérer comme continu sans pour autant occuper un espace mémoire trop important.

[25] La méthode d'intégration d'Euler suppose simplement qu'après une durée très courte Δt, une variable x peut être mise à jour par $x(t+\Delta t) = x(t) + \Delta t\,\dot{x}(t)$ où $\dot{x}(t)$ est la dérivée de $x(t)$ à l'instant t.

adaptant la méthode VV au cas présentant un terme fluctuatif et un terme dissipatif[26], ce qui permet d'augmenter sensiblement la vitesse de calcul en comparaison avec l'intégrateur d'Euler. En revanche, bien qu'il existe très probablement de nombreux intégrateurs plus efficaces, la version sur-amortie de l'équation du mouvement (équation (II.9)) est actuellement intégrée par la méthode d'Euler.

Utilisations et temps de calcul

Le modèle présenté ici permet de représenter des phénomènes très différents. On peut en effet aussi bien reproduire, à une échelle temporelle fine, la dynamique vibratoire d'une protéine par un ENM (telle que décrite par les méthodes analytiques que permettent ce formalisme), que représenter des phénomènes de transition de phase relatifs à la dynamique des corps nucléaires. Lorsque les réactions biochimiques seront intégrées au logiciel de simulation, il sera de plus possible de décrire un système similaire aux approches de réaction/diffusion de particules en l'absence d'interactions physiques (tels que les modèles GFRD et Smoldyn), permettant alors d'atteindre des durées simulés probablement suffisamment longues pour décrire le processus d'expression génique de façon globale. Il est cependant clair que la représentation de processus biologiques sur de telles durées interdit d'utiliser une description fine des macromolécules, même si une telle description est par ailleurs permise par le modèle (représentation multidomaine, ...). Par exemple, l'étude de la dynamique d'association/dissociation, sur une séquence d'ADN, d'un FT représenté par deux domaines ne permet pas d'atteindre des durées de l'ordre de la seconde. À titre d'exemple, la simulation d'un tel système (représenté sur la figure II.7) pendant 2 ms prend de l'ordre de l'heure avec la version actuelle du logiciel. Même si certaines optimisations importantes peuvent encore y être apportées, il paraît peu probable d'atteindre un rapport temps de calcul/temps simulé significativement meilleur. De même, l'exemple de la figure II.6, décrivant un phénomène à l'échelle du corps nucléaire sur une durée de 20 ms et employant la version 2 du logiciel, représente un temps de calcul de l'ordre de la dizaine d'heures.

Ainsi, si l'approche employée ici permet de représenter des phénomènes très différents à des échelles spatio-temporelles très variées, il est important de garder à l'esprit que c'est avant tout le niveau de détail de la description des molécules qui détermine les possibilités de calcul.

2.3 Illustrations à différentes échelles

Afin d'illustrer la capacité du modèle à représenter des phénomènes à des échelles spatio-temporelles différentes, nous présentons ici deux exemples très simples d'utilisation du logiciel de simulation. Il ne s'agit là que de preuves-de-concept et ces simulations ne constituent en aucun cas une étude rigoureuse des phénomènes qu'elles représentent. Malgré cela, elles permettent d'illustrer certains mécanismes et certains concepts ayant de pos-

[26] Cette technique fait en particulier intervenir une distribution normale bivariée pour représenter l'effet du terme fluctuatif en modifiant de façon cohérente à la fois les positions et les vitesses des particules.

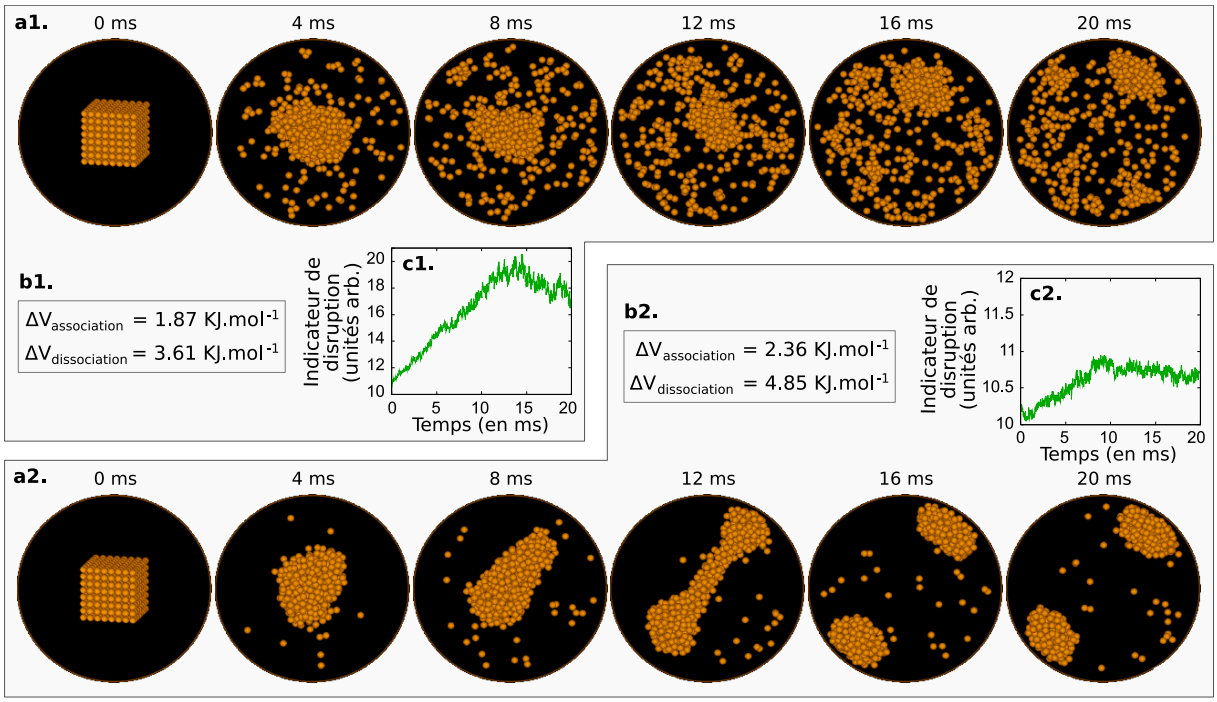


Fig. II.6: Disruption et équilibre de phase d'un corps nucléaire minimal. Cette figure décrit la dynamique spontanée d'un corps nucléaire très simple, constitué de 512 protéines identiques, interagissant par des potentiels de Lennard-Jones et de Coulomb, au sein d'un espace borné par une membrane sphérique. Nous comparons ici le processus d'équilibration de phase depuis une structure compacte (a1-2), pour deux paramétrages différents des énergies d'association et de dissociation (b1-2 ; cf figure II.3b). Comme en témoigne l'indicateur de disruption (c1-2), reflétant la quantité de protéines libres, on observe dans les deux cas qu'un équilibre est atteint au cours de la simulation. En effet, la proportion de protéines libres augmente jusqu'à ce que les quantités de protéine entrant et quittant la structure s'équilibrent. Ce processus prend ici de l'ordre de la dizaine de millisecondes, indiquant que le renouvellement des protéines constituant la structure est très rapide, une propriété observée expérimentalement sur les corps nucléaires. On constate également que la concentration libre de la protéine est très dépendante de ses propriétés d'auto-interaction. De plus, ces deux exemples reflètent différents mécanismes possibles pour la formation et la disparition des corps nucléaires. Dans le premier cas, en plus de la structure initiale qui perdure pendant toute la durée de simulation, on observe la nucléation transitoire de petites régions de phase condensée. À l'inverse, dans le second cas, on observe clairement un phénomène de fission de la structure de départ, phénomène également qualifié de *défloculation* ou de *peptisation* en chimie de colloïdes. (Ces résultats ont été obtenus avec le logiciel 3DSPI v2).

sibles implications pour l'expression génique comme nous l'évoquerons dans la section suivante.

La figure II.6 représente un phénomène de disruption d'un corps nucléaire très simple : partant d'un structure compacte, nous observons l'évolution temporelle de deux systèmes dans lesquels l'unique espèce protéique représentée possède des propriétés d'auto-interaction différentes. On observe alors que la quantité de protéines libres se stabilise rapidement (indiquant un taux élevé de renouvellement des protéines au sein du corps

nucléaire) et ce, à une valeur très différente dans les deux cas représentés. On observe également différents mécanismes possibles pour la formation des corps nucléaires (nucléation, peptisation, ...).

À une échelle inférieure, nous décrivons, sur la figure II.7, une séquence d'ADN d'affinité hétérogène et observons la dynamique d'association/dissociation d'un facteur de transcription (FT). Nous comparons le cas d'une séquence ne présentant qu'un seul site de fixation avec celui d'une séquence en présentant trois. On observe clairement un comportement différent dans ces deux situations. Les caractéristiques dynamiques de la présence du FT sur un site donnée (ou plus précisément, son spectre de puissance) sont en effet influencées par la présence d'autres sites de fixation à proximité, par un phénomène de séquestration du FT. Même si cet exemple ne constitue qu'une simple diffusion essentiellement en 1D le long d'une séquence hétérogène et pourrait donc être étudié par d'autres techniques, l'utilisation de notre modèle pour aborder de telles questions permettrait de prendre en compte certains phénomènes habituellement difficiles à représenter (affinité du FT avec d'autres molécules présentes sur le promoteur, interaction avec des structures hors du promoteur, ...).

3 Hypothèses et mécanismes à étudier

Bien que nous ayons identifié certaines limites de notre simulateur en termes de temps de calcul et de pouvoir descriptif, la démarche de conception du modèle nous a conduit à formuler certaines hypothèses quant à l'influence de la spatialité et de l'organisation nucléaire sur le processus d'expression génique. Nous présentons ici ces hypothèses, certaines pouvant être étudiées à l'aide de notre modèle, d'autres nécessitant probablement l'utilisation de modèles différents et plus spécifiques.

3.1 Structuration de l'expression par des mécanismes de diffusion

Nous proposons ici des hypothèses quant aux corrélations à de nombreux niveaux que peuvent provoquer des phénomènes de diffusion dans l'espace (facteurs de transcription ou autres entités intervenant dans la régulation génique).

Corrélation et anti-corrélation de l'activité de deux gènes

Comme l'illustre la figure II.8, la régulation de deux gènes par un même facteur de transcription (FT) présent en une unique copie, semble être susceptible de provoquer des corrélations particulières entre leurs niveaux d'expression par un phénomène en partie analogue à celui mise en évidence par van Zon *et al.* (2006). Les caractéristiques de cette corrélation dépendraient alors de nombreux paramètres tels que la distance entre les deux gènes, leur emplacement au sein de l'espace nucléaire, le coefficient de diffusion du FT et la durée de vie des transcrits (et des protéines). Ce mécanisme met alors en défaut les notions classiques de stochasticités intrinsèques et extrinsèques (cf section 1.2 et figure I.3 du

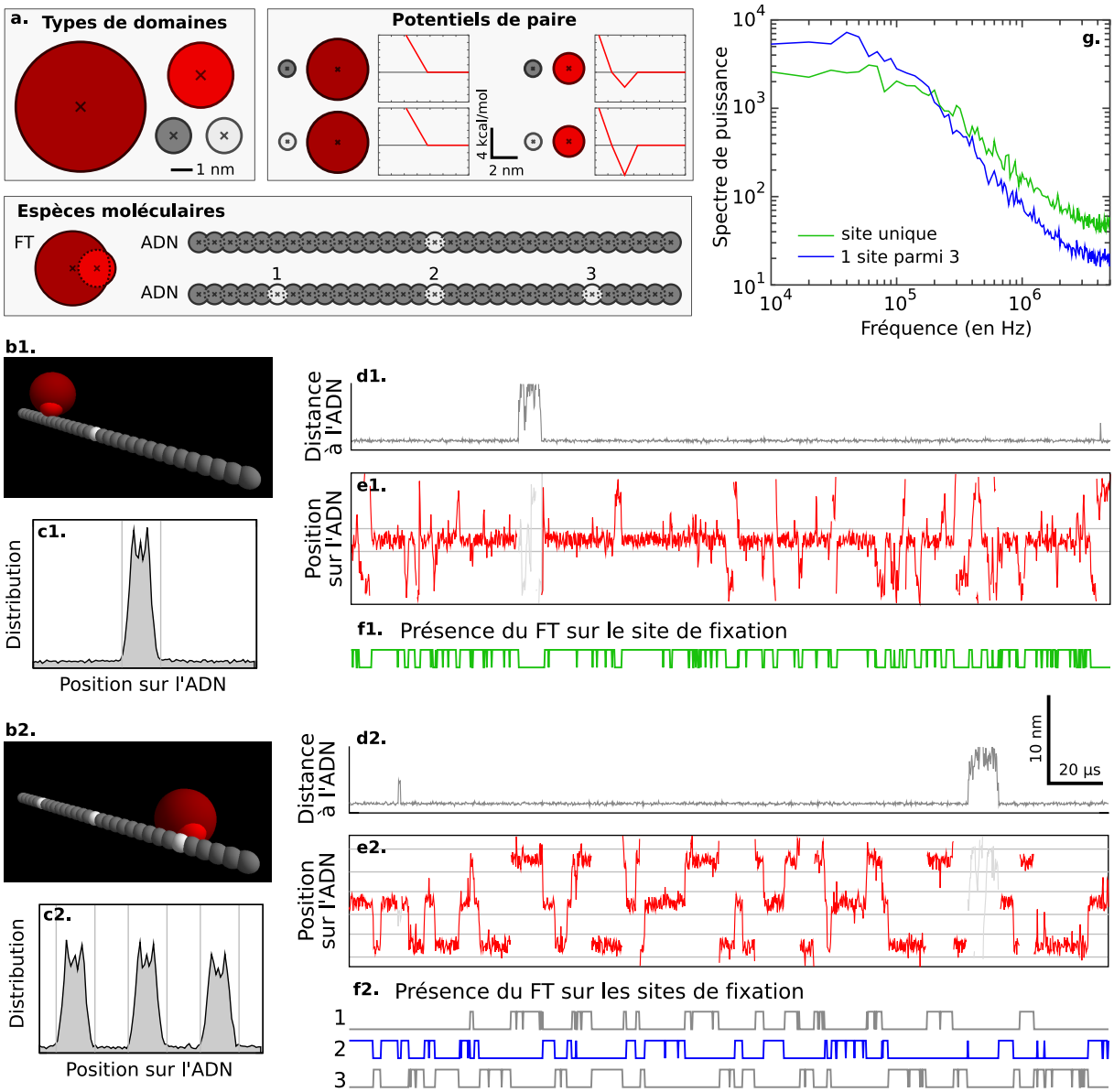


Fig. II.7: Influence du profil d'affinité d'un FT sur sa dynamique. (a) Sur le même principe que la figure II.5, nous définissons ici un système minimaliste représentant un FT (composé d'un domaine principal et d'un domaine de liaison à l'ADN) et une séquence d'ADN pouvant comporter 1 ou 3 sites de fixation (gris clair). (b1-2) Nous comparons alors, dans ces deux cas de figure, la dynamique d'association/dissociation d'un unique FT sur les différents sites de fixation. (c1-2) La distribution stationnaire de la position du FT sur l'ADN reflète directement le profil d'affinité de la séquence. (d1-2) De plus, comme en témoigne l'évolution temporelle de la distance du FT à l'ADN, celui-ci reste ici essentiellement lié et se dissocie occasionnellement pour se réassocier rapidement par la suite. (e1-2) On peut alors observer la diffusion du FT le long de la séquence et sa capture par l'unique site de fixation dans le premier cas (f1) et par les trois sites dans le second (f2). Les lignes grises sur (e1-2) sont les mêmes que sur (c1-2) et délimitent les bassins d'attraction des sites de fixation. Les discontinuités des courbes rouges (e1-2) proviennent du fait que l'espace est torique. (g) Les caractéristiques dynamiques de la présence du FT sur un site donné peuvent être comparées par leurs spectres de puissance (cf Box 1, page 22), obtenus ici à partir de simulations d'une durée de 2 ms. On s'aperçoit alors que, dans le second cas de figure, ces caractéristiques spectrales peuvent être influencée par le simple phénomène de séquestration du FT sur les deux autres sites.

chapitre I). En effet, cette classification binaire suppose qu'une source de stochasticité influence deux copies d'un certain système (ici le gène) soit de façon identique, soit de façon totalement indépendante. Dans le cas que nous considérons, c'est pourtant un même phénomène qui peut aussi bien corréler les niveaux d'expression des deux gènes que les décorréler et ce, non pas de façon binaire, mais avec différents degrés en fonction des rapports entre les différents phénomènes spatio-temporels impliqués (diffusion d'un FT et dégradation des transcrits).

Ainsi, la spatialité de la régulation au sein du noyau serait à même de structurer les propriétés stochastiques conjointes de l'expression de plusieurs gènes. Un aspect important de ce mécanisme est en particulier la probabilité qu'a le FT de se réassocier à un site qu'il vient de quitter et sa probabilité de capture par un autre site plus ou moins proche. Dans ce cadre, il est très probable que la structure de l'espace dans lequel diffuse le facteur joue un rôle majeur. Ainsi, Tkačik et Bialek (2009) ont montré que la diffusion d'un FT selon un mécanisme « 3D+1D » (combinant une diffusion dans l'espace et le long de l'ADN ; cf chapitre I, section 2.7) provoquait des corrélations temporelles importantes dans le retour du FT sur son site de fixation. On sait de plus, par des observations expérimentales, que l'espace accessible à un FT peut avoir une structure fractale (Misteli, 2005; Bancaud *et al.*, 2009; Rouquette *et al.*, 2009). Cette caractéristique confère au mouvement de diffusion du FT des propriétés dites anormales ayant un impact sur la dynamique d'exploration de l'espace et la capture du FT par un site de fixation (Saxton, 2001; Berry, 2002; Guigas et Weiss, 2008). En effet, plus la diffusion est anormale, plus il est probable que le FT soit facilement capturé par un site proche et difficilement capturé par un site lointain (Guigas et Weiss, 2008). Ce mécanisme est donc particulièrement pertinent dans le contexte du phénomène proposé sur la figure II.8.

Le mécanisme représenté sur la figure II.8, ainsi que celui décrit par van Zon *et al.* (2006), constituent cependant des cas particuliers car ils nécessitent que le FT ne soit présent qu'en relativement peu de copies. Mais dans ces deux cas, il semble probable (même si cette affirmation nécessiterait d'être vérifiée) que plus la rencontre du FT avec le(s) site(s) est rare et plus sa réassociation (sur l'un des deux sites proches) est probable, plus le phénomène reste valable pour un nombre important de copies du FT. C'est en particulier le cas lorsque la diffusion du FT présente un fort degré d'anormalité car une particule suivant un tel mouvement explore l'espace beaucoup plus localement que dans le cas d'un mouvement brownien classique, impliquant alors une exploration globale de l'espace très longue mais de nombreuses réassociations une fois proche d'un site de fixation.

Ce phénomène, décrit jusqu'ici comme étant dû à la diffusion des FT, peut être également étendu au cas du déplacement d'autres entités impliquées dans la régulation génique ou au cas du déplacement des deux loci par rapport à certains éléments de l'architecture nucléaire. C'est ainsi ce qu'évoquent Becskei *et al.* (2005) dont les résultats expérimentaux indiquent (tout comme ceux de Raj *et al.* (2006)) qu'une colocalisation spatiale de deux gènes influence la corrélation de leur expression. On peut par exemple envisager que deux gènes proches sur la séquence voient leurs expressions corrélées car ils sont toujours présents dans le même contexte nucléaire (à l'intérieur d'un territoire chromosomique, dans une région inter-chromosomique, ... ; cf chapitre I, section 4.4). Nous reviendrons sur ce point dans la section 3.2.

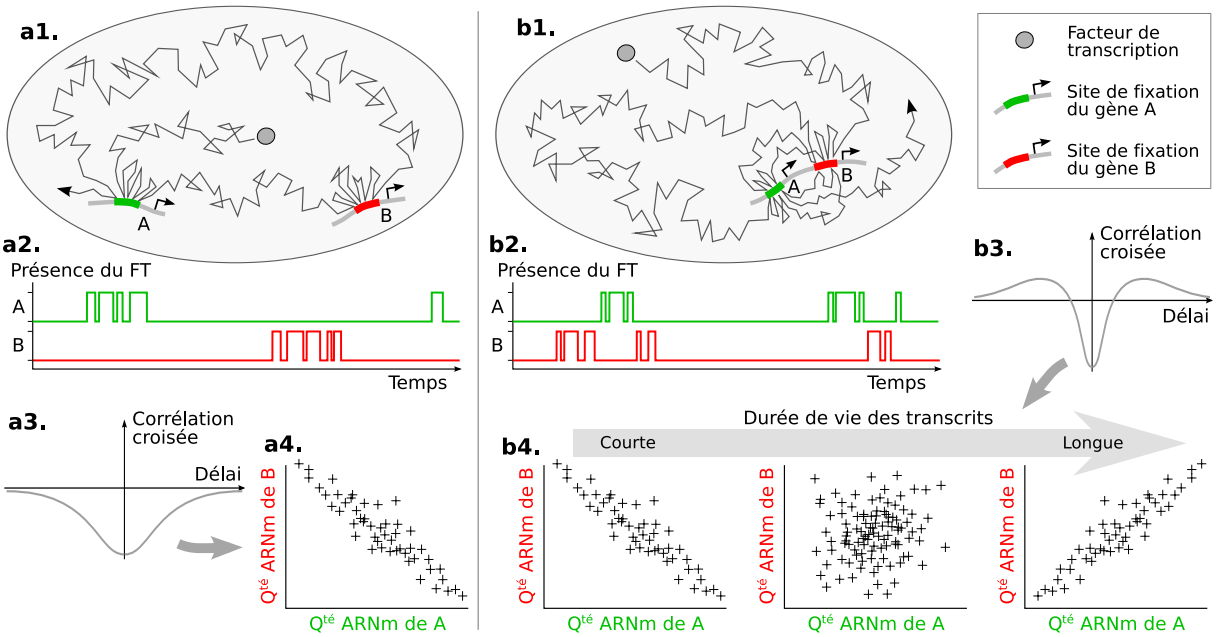


Fig. II.8: (anti-)Corrélation de l'activité de deux gènes par des mécanismes spatiaux. Les idées présentées ici ne sont que des hypothèses et nécessitent d'être testées expérimentalement. Les graphiques ne sont pas le résultat de simulations et ne sont montrés qu'à des fins d'illustration. (a1) Considérons deux gènes A et B placés à des loci éloignés au sein de l'espace nucléaire et régulés par un même FT présent en une unique copie. (a2) Lorsque le FT est présent sur l'un des deux loci, il ne l'est nécessairement pas sur l'autre et son passage de l'un à l'autre requiert un délai pour sa diffusion à travers l'espace nucléaire. (a3) On peut alors s'attendre à ce que la corrélation croisée de l'occupation des deux loci présente une anti-corrélation (dont la durée reflète le temps moyen de diffusion d'un locus à l'autre) se traduisant par (a4) une anti-corrélation entre les deux niveaux d'expression (lorsqu'un gène est exprimé, l'autre ne l'est pas, et *vice versa*). (b1) Dans le cas où les deux loci sont relativement proches l'un de l'autre par rapport à l'espace nucléaire, il est probable que le même phénomène ait lieu sur des durées plus courtes et que s'y ajoute, à des échelles plus longues, un phénomène similaire à celui mis en évidence par van Zon *et al.* (2006) (cf figure I.3, page 33). (b2) En effet, dans cette situation, les deux gènes peuvent être actifs à tour de rôle lorsque le FT se réassocie de nombreuses fois sur les deux sites, puis rester tous deux inactifs pendant une longue période durant laquelle le FT explore d'autres régions de l'espace avant de se rapprocher à nouveau des deux gènes. (b3) Il est alors probable que ces deux phénomènes se traduisent par une anti-corrélation sur des temps courts (reflétant l'alternance du FT sur les deux sites) et une corrélation sur des temps longs (reflétant la diffusion dans l'ensemble de l'espace nucléaire). (b4) C'est alors la durée de vie des transcrits qui détermine la corrélation entre les deux niveaux d'expression : si cette durée de vie est plus courte que la durée d'anti-corrélation, alors les deux niveaux d'expression seront anti-corrélés. Mais si elle est au contraire supérieure à la durée moyenne que le FT passe dans le voisinage des deux gènes, son alternance entre les deux loci n'aura que peu d'effets et les deux gènes apparaîtront comme actifs et inactifs simultanément, provoquant alors une forte corrélation de leurs niveaux d'expression. Ainsi, la corrélation entre les deux niveaux d'expression – une mesure très couramment employée pour séparer des sources intrinsèques et extrinsèques de stochasticité (cf figure I.1, page 20) – serait donc ici régie par une relation complexe entre de nombreux paramètres tels que la distance entre les deux gènes, la taille de l'espace nucléaire, la durée de vie des transcrits, le coefficient de diffusion du FT et la dimension de l'espace accessible au FT (diffusion « 3D+1D », diffusion anormale, ... cf texte).

Structuration de l'activité d'un gène par son environnement local

En ne considérant maintenant qu'un seul des deux gènes sur la figure II.8b, on peut également remarquer que les propriétés de son expression peuvent être influencées par la présence à proximité d'un second site de fixation pour son FT. En effet, comme l'illustre l'occupation de B sur la figure II.8b2, la présence du site A provoque de courtes périodes où le FT est absent car séquestré sur le site A. On peut ainsi s'attendre, dans un cadre général, à ce que l'environnement proche d'un gène soit susceptible de provoquer une certaine structuration de son activité. En effet, comme évoqué précédemment, la dimension de l'espace dans lequel se trouve le gène peut fortement influencer la probabilité de réassociation d'un FT sur son site ainsi que la distribution de durée de recapture lorsque le FT s'est sensiblement éloigné. À un niveau très local, la présence de plusieurs sites de fixation sur une séquence d'ADN peut donc modifier les propriétés temporelles de la présence du FT sur chacun des sites, comme le montre l'expérience préliminaire de la figure II.7. Il paraît alors possible que l'hétérogénéité du profil d'affinité d'un FT le long de l'ADN puisse en partie de structurer l'activité du gène. De plus la présence sur l'ADN de nombreuses autres molécules subissant le même processus et avec lesquelles le FT interagit par exclusion stérique peut également modifier ces propriétés dynamiques.

Il paraît ainsi important de considérer le contexte spatial d'un gène et de prendre en compte, en particulier, l'interaction des molécules le régulant avec des entités tierces (autres sites de fixation, espace accessible, ...). Cependant, les régulateurs ne sont pas les seuls à se déplacer et les loci des gènes eux-mêmes témoignent de mouvements particuliers en relation avec leur expression. En effet, comme nous l'évoquions au chapitre I (section 4), l'expression génique chez les eucaryotes fait intervenir des mécanismes de relocalisation des gènes dans les régions dites inter-chromosomiques et au sein d'usines à transcription. Ces mécanismes sont donc susceptibles de jouer un certain rôle sur les caractéristiques temporelles de l'expression d'un gène et de l'expression conjointe de plusieurs gènes. Il convient alors de les prendre en compte et de les étudier de ce point de vue.

3.2 Régions inter-chromosomiques et usines à transcription

Au sein d'un paysage nucléaire très hétérogène et organisé, l'expression génique chez les eucaryotes fait intervenir un certain nombre de structures et de mécanismes spatiaux à différents niveaux : de la relocalisation des gènes hors des territoires chromosomiques à leur rencontre dans des usines à transcription (phénomène de *gene kissing*) et au sein des régions interchromosomiques où s'intermêlent l'ADN de différents chomosomes avec différentes structures impliquées dans la transcription (cf chapitre I, section 4.4). Ces mécanismes, découverts très récemment, ont probablement des implications majeures sur le processus d'expression génique et nécessitent d'être étudiés du point de vue de la modélisation. En effet, la figure I.5 (page 45), représentant cette organisation à différentes échelles, évoque clairement la possibilité de nombreux mécanismes influençant la dynamique stochastique de la transcription d'un ou de plusieurs gènes.

On peut par exemple s'intéresser à l'effet du déplacement des longues fibres de chromatine s'intermêlant au sein d'une région interchromosomique et pouvant localement se

compacter/décompacter, rendrant alors accessible ou inaccessible différents gènes. Le rapprochement temporaire de différents loci pourrait par exemple provoquer transitoirement des corrélations entre l'expression de différents gènes, sur la base du principe illustré sur la figure II.8. Le mouvement de ces structures est en effet probablement plus lents que ceux de FT, mettant alors en jeu deux échelles de temps distinctes. Il pourrait dans ce cadre être pertinent de s'inspirer des modèles coarse grain de fibre de chromatine (Arya et Schlick, 2006; Lesne, 2006; Schlick et Perišić, 2009) ou d'en développer un, plus simplifié et à une échelle de description plus grossière, ne représentant que les propriétés essentielles que seraient la flexibilité en fonction du degré de compaction/décompaction et les phénomènes de propagation de cette compaction.

On peut également considérer l'effet des usines à transcription, voire les phénomènes de nucléation des usines à transcription, prenant place préférentiellement sur certaines séquences, influençant alors l'expression des gènes avoisinants.

À une échelle plus locale, la dynamique au sein d'une unique usine à transcription semble également pouvoir provoquer certains phénomènes particuliers dans l'expression conjointe de plusieurs gènes. En effet, le faible nombre de polymérases par usine à transcription (estimé à 8 en moyenne (Sutherland et Bickmore, 2009)) provoque certainement des phénomènes de compétition entre les différents gènes partageant une même usine. De plus, une caractéristique particulière à cette échelle est que les mouvements sont en partie influencés par l'activité de transcription. Celle-ci est en effet un moteur moléculaire puissant, capable de déplacer le matériel génétique par rapport à l'usine à transcription (il semblerait en effet que la polymérase reste essentiellement fixe et que c'est l'ADN que se déplace lors de sa transcription). Ainsi, si un gène au promoteur très fort entre dans une usine à transcription, il peut temporairement séquestrer toutes les polymérases et diminuer ainsi la transcription des autres gènes. Mais le simple fait qu'il soit transcrit le relocalise alors hors de l'usine à transcription, empêchant probablement sa transcription de façon continue et laissant alors la possibilité aux autres gènes de s'exprimer à nouveau. Ce type de mécanismes pourrait alors provoquer des phénomènes de bursts et d'anticorrélation entre l'expression de plusieurs gènes.

Il est probable que de nombreux mécanismes de nature similaire puissent avoir lieu au sein des usines à transcription. On peut noter que relativement peu de modèles mésoscopiques intracellulaires (cf section 1) permettraient d'étudier ce type de phénomènes. En effet, il est ici nécessaire de pouvoir représenter les contraintes mécaniques liées à l'effet moteur de la transcription. Remarquons qu'il est possible, avec un modèle intégrant des réactions chimiques et des potentiels d'interaction (comme ce sera, à terme, le cas de 3DSPI v3), de décrire des moteurs moléculaires[27]. Les questions relatives à l'organisation spatiale de la transcription au sein d'un espace nucléaire organisé sont très récentes et encore peu explorées du point de vue de la modélisation.

[27]Considérons une macromolécule (représentant par exemple une polymérase) sur une chaine de particules (représentant par exemple un morceau d'ADN). Si l'interaction de la macromolécule avec un site de forte affinité provoque une réaction rendant alors cette affinité moins bonne que celle du site voisin, la macromolécule va alors se déplacer dessus. Si ce même mécanisme a lieu avec ce nouveau site, la macromolécule se déplacera alors à nouveau sur le site voisin, et ainsi de suite. La variation d'énergie potentielle à chaque réaction constitue l'étape énergie-dépendante du processus et peut être quantifiée très simplement.

3.3 Corps nucléaires : transitions de phase et non-linéarités

L'organisation nucléaire fait intervenir de nombreuses structures, appelées corps nucléaires, correspondant à des régions de l'espace de composition particulière, dépourvues de membrane et dont les composants sont en échange permanent avec l'extérieur (cf chapitre I, section 4). Ces structures paraissent alors très semblables aux différentes phases d'un système multiphasique constitué d'espèces chimiques partiellement miscibles. En effet, les phénomènes de nucléation (apparition d'un corps nucléaire) et de coalescence (fusion de deux corps nucléaires), tels qu'il sont observés au sein des noyaux, sont assez caractéristiques de ce type de systèmes.

On peut imaginer différentes implications pour ces phénomènes de transition de phase dans le cadre d'une expression génique stochastique. Par exemple, l'existence d'un corps nucléaire riche en une espèce biochimique donnée pourrait permettre d'assurer une concentration constante de cette protéine dans le reste du noyau. En effet, dans un système composé de molécules A et de molécules B partiellement miscibles, lorsque l'augmentation de la concentration de l'espèce B provoque la nucléation de gouttelettes riche en B, un équilibre se créé dans l'échange de matière entre les deux phases. Si l'on accroît encore la quantité de B, la nouvelle phase riche en B augmente alors de taille, mais les concentrations de A et de B dans chacune des deux phases restent inchangées. Si la molécule B est une protéine dont l'expression présente des fluctuations stochastiques, la présence d'un tel corps nucléaire permet de maintenir constante la concentration libre de cette protéine dans le reste du noyau.

La valeur seuil de la concentration de molécules B à laquelle se fait cette transition peut être très différente d'un cas à l'autre. En effet, sur la figure II.6, même s'il s'agit là d'un cas légèrement différent car les protéines l'espèce A ne sont pas décrites explicitement[28], la concentration libre de la protéine à l'équilibre dépend fortement des propriétés d'interaction de l'espèce B avec elle-même. Ainsi, dans un cas où la concentration de B reste supérieure à cette valeur seuil de transition de phase, la présence d'un corps nucléaire aura pour effet d'atténuer les fluctuations. Mais dans le cas inverse, où les fluctuations de B font franchir cette valeur de transition, on peut s'attendre à assister à l'apparition et à la disparition stochastique du corps nucléaire. La non-linéarité provoquée par cet effet de seuil pourrait alors avoir pour effet d'amplifier le caractère stochastique des phénomènes liés à l'activité de B. En effet, des fluctuations, même faibles, au bord de cette valeur critique provoquent ponctuellement de fortes surconcentrations locales de B.

Ces mécanismes élémentaires et bien connus de transition de phases dans des systèmes multiphasiques pourraient avoir de nombreuses implications au sein du noyau et mériteraient ainsi davantage d'attention. Comme l'illustre la figure II.6, de telles questions pourraient en particulier être abordées avec notre modèle.

[28] Ce cas s'apparente donc plus à un système liquide-vapeur qu'à un système liquide-liquide.

4 Conclusion

Nous avons proposé un modèle, ainsi que le logiciel de simulation associé, nommé 3DSPI v3, pour l'étude des phénomènes d'organisation spatiale au sein du noyau cellulaire en relation avec le processus d'expression génique. Contrairement à de nombreux modèles à cette échelle, cet outil vise à permettre une description à la fois physique et biochimique des interactions entre molécules, des caractéristiques vraisemblablement essentielles pour l'études de nombreux mécanismes au sein du noyau. Ce modèle est encore relativement préliminaire et certains développements nécessitent d'être poursuivis. Comme nous l'avons vu, l'approche comporte cependant certaines limitations. Il est en effet clair que la prise en compte des interactions physiques précises entre molécules ne permet pas d'atteindre des durées de simulation pertinentes pour décrire, même simplement, le processus global d'expression génique. Malgré ce constat, la démarche de conception du modèle nous a permis de développer un certain nombre d'hypothèses relatives à l'influence du contexte spatial d'un gène sur son expression.

À l'issue de ce chapitre, bien que notre modèle permette de prendre en compte des aspects essentiels du contexte spatial de l'expression génique, il apparaît clairement que certaines questions restent difficilement abordables. En particulier, comme nous l'avons vu au chapitre I, la dynamique des complexes impliqués dans l'activation et la régulation transcriptionnelle chez les eucaryotes fait intervenir des mécanismes complexes d'influence combinatoire entre molécules (coopération/compétition, modification conformationnelle et covalente de la chromatine, ...). Un tel système ne pourrait être représenté dans notre modèle spatial que de façon très simplifiée, nécessitant d'importantes simplifications ayant probablement un impact majeur sur le comportement du système. De plus, les échelles temporelles impliquées dans la dynamique d'un promoteur – de la seconde à l'heure – sont clairement inaccessibles par un tel modèle de simulation. L'étude de ces mécanismes nécessite alors d'emprunter une approche sensiblement différente, faisant en particulier appel à des formalismes autorisant une étude analytique. Nous proposons, dans le chapitre suivant, un tel modèle permettant de représenter spécifiquement les mécanismes d'interaction combinatoire entre molécules sur un promoteur et leur influence sur l'expression génique.

Chapitre III

Dynamique stochastique de la régulation transcriptionnelle

Nous présentons, dans ce chapitre, nos travaux relatifs à la dynamique stochastique du promoteur et à son influence sur la stochasticité de l'expression génique. Nous commençons par décrire, dans la section 1.4, les travaux de modélisation existants, adoptant chacun différents points de vue sur la question. Cela nous amène alors à définir les grandes lignes d'un modèle que nous présentons dans la section 2. Nous en effectuons alors, dans la section 3, les dérivations générales permettant en particulier la déduction de divers indicateurs (aussi bien théoriques que reproduisant des mesures expérimentales). Nous exploitons alors ces dérivations, dans la section 4, afin de mettre en évidence différentes propriétés concernant la génération, par un promoteur, d'une stochasticité structurée de l'expression du gène, mais aussi sur sa dynamique multi-échelle en relation avec les propriétés physico-chimiques des interactions entre molécules. Enfin, dans la section 5, nous évoquons différentes perspectives pour l'exploitation future de ce modèle autant dans un cadre théorique que du point de vue de l'analyse de données expérimentales réelles.

1 Modélisations du promoteur

Il existe, dans la littérature actuelle, relativement peu de modèles considérant de façon détaillée et générale la dynamique des événements moléculaires prenant place sur un promoteur, en particulier dans le cadre de la régulation eucaryote. De plus, parmi ces quelques modèles, ceux qui se placent du point de vue de l'expression stochastique des gènes sont encore plus rares.

1.1 Simulation *vs* dérivations analytiques

Notons tout d'abord que, parmi les modèles décrivant la dynamique d'un promoteur, la plupart sont exploités par l'intermédiaire de la simulation et non par la dérivation de prédictions analytiques. Or, alors que, dans un contexte stochastique, une simulation ne correspond qu'à une trajectoire donnée du système étudié et doit être répétée un grand nombre de fois pour caractériser ce dernier de façon significative, les prédictions analytiques décrivent le système de façon beaucoup plus précise et rapide[1]. Par exemple, pour le promoteur à deux états présenté au chapitre I (figure I.2a, page 26) et pour un jeu de paramètres donné, il est possible de réaliser une simulation stochastique de l'évolution du système (figure I.2b). En répétant un grand nombre de fois cette simulation, on peut alors obtenir une estimation approchée de la distribution stationnaire du niveau d'ARN correspondant au jeu de paramètre utilisé. À l'inverse, il est possible de déduire par un calcul analytique l'expression exacte de cette distribution en fonction des différents paramètres (Innocentini et Hornos, 2007), révélant alors l'ensemble des comportements du système pour tous les jeux de paramètres possibles.

Ainsi, les prédictions analytiques autorisent non seulement une exploration paramétrique beaucoup plus rigoureuse, et donc une meilleure généralisation des résultats, mais permettent également une interprétation plus poussée des résultats en révélant la structure du comportement du système tout en mettant en évidence l'importance de certains paramètres ou de certaines relations entre les paramètres. Cependant, plus un modèle est complexe et hétérogène dans sa formulation plus sa résolution analytique est difficile, voire impossible. Si l'on souhaite obtenir des résultats analytiques, cela impose alors parfois un cadre plus rigide que la simulation, interdisant par exemple la relaxation de certaines hypothèses. Il est de ce point de vue particulièrement intéressant de faire appel à des modèles génériques car ils permettent de décrire un grand nombre de situations différentes avec un même formalisme général (Paulsson, 2005a; Warren *et al.*, 2006; Cox *et al.*, 2006; Lestas *et al.*, 2008). Nous souhaitons donc dans nos travaux adopter ce type d'approche afin de pouvoir représenter des promoteurs très différents tout en ayant une description analytique de leur comportement.

1.2 Modèles de l'expression stochastique

Comme nous l'avons vu au chapitre I (section 2.4), la très grande majorité des modèles d'expression stochastique des gènes prenant en compte une dynamique explicite du promoteur ne considèrent celui-ci que comme un système à deux états. Si l'on excepte ce modèle classique, on peut citer les travaux de Blake *et al.* (2003, 2006), de Murphy *et al.* (2007), de Sánchez et Kondev (2008) et de Boeger *et al.* (2008) décrivant l'activité stochastique de promoteur plus complexes.

Les travaux de Blake *et al.* (2003, 2006) et de Murphy *et al.* (2007) considèrent l'interac-

[1] Il peut être fait appel, dans de nombreux cas, à des méthodes de résolution numérique (par exemple, la résolution d'un systèmes d'équations linéaires). Nous considérons ici ces méthodes comme relevant de la prédiction analytique car elles constituent une automatisation d'un calcul analytique et, même si elles peuvent parfois représenter un temps de calcul significatif, elles donnent un résultat exact.

tion de deux ou trois molécules sur le promoteur pour interpréter des données d'expression stochastique. Mais ces études, ne considérant que des cas particuliers et non un modèle générique et n'exploitant leur modèle que par l'intermédiaire de la simulation, ne permettent pas de tirer des conclusions générales.

À l'inverse, les travaux de Sánchez et Kondev (2008) représentent de façon générique l'interaction d'un nombre arbitraire de molécules sur un promoteur. Les auteurs considèrent l'association/dissociation de chaque molécule, influencée par les autres molécules présentes sur le promoteur, et décrivent les conséquences de cette dynamique sur l'expression stochastique du gène. Ils caractérisent alors le comportement du système en dérivant les différents moments de la distribution stationnaire du niveau d'expression par la résolution numérique de systèmes d'équations linéraires (un par moment de la distribution). Cependant, comme nous l'avons vu (cf Box 1, page 22), l'ensemble des moments ne permet pas une interprétation aussi explicite que la distribution elle-même puisqu'il ne révèle pas d'informations telles que la taille, la position et la forme des différents modes. De plus, ces indicateurs ne décrivent que les aspects stationnaires du système et non sa dynamique qui peut pourtant en constituer un aspect central. Ce modèle reste cependant très expressif puisqu'il consiste en une description générique du promoteur, rendant donc potentiellement compte d'un très grand nombre de situations différentes. Il peut ainsi constituer une base pour de nombreuses investigations. Le modèle que nous proposons dans ce chapitre partage par ailleurs, dans sa formulation, plusieurs caractéristiques communes avec le modèle de Sánchez et Kondev (2008). Cependant, les auteurs restent dans le cadre d'un système procaryote et n'interprètent pas leur modèle ni en termes dynamiques, ni d'un point de vue énergétique, deux aspects qui, comme nous allons l'illustrer tout au long de ce chapitre, occupent un place centrale dans la compréhension du comportement du promoteur.

Les travaux de Boeger *et al.* (2008) décrivent l'influence de la dynamique conformationnelle de la chromatine du promoteur d'un gène sur la stochasticité de son expression. Même si les auteurs ne considèrent qu'un système particulier et relativement réduit (6 états) et que l'expression stochastique n'est décrite qu'en termes de moyenne et de variance, il s'agit là de l'une des rares études considérant l'effet d'une représentation aussi détaillée de la chromatine sur l'expression stochastique.

Il existe dans la littérature d'autres modèles décrivant de façon précise le promoteur d'un gène, mais ne prenant pas le point de vue de l'expression stochastique. On trouve d'une part les modèles thermodynamiques que nous décrivons à la section 1.3. Ces modèles sont utilisés depuis plusieurs années pour décrire la régulation génique (essentiellement chez les procaryotes) mais le formalisme employé impose un certain nombre de limitations. Il existe d'autre part quelques modèles décrivant la dynamique d'un promoteur eucaryote mais n'abordant pas les questions relatives à l'expression même du gène. Ces modèles, très peu nombreux et relativement différents les uns des autres seront décrits à la section 1.4. Nous verrons, au long de ce chapitre, comment tous ces différents modèles peuvent être réinterprétés dans le cadre de celui que nous proposerons à la section 2.

1.3 Modèles de la régulation (modèles thermodynamiques)

La régulation génique a longtemps été étudiée en décrivant l'affinité des molécules avec le promoteur et les comportements de coopération et de compétition entre molécules par des énergies libres d'interaction (Ackers *et al.*, 1982; Scholl et Nixon, 1996; Lehman *et al.*, 1998; Bintu *et al.*, 2005a; Vilar et Saiz, 2005; Saiz et Vilar, 2006a). La combinaison de ces énergies d'interaction permet d'associer à chaque configuration possible de molécules sur le promoteur, une énergie du système. En considérant alors le système à l'équilibre thermodynamique[2], il est possible de déduire analytiquement (par l'application d'un facteur de Boltzmann) la probabilité d'occurrence de chaque configuration – et donc le niveau d'expression moyen – en fonction des concentrations des différentes espèces moléculaires. De plus, il est possible de décrire par un coût énergétique différents changements conformationnels de l'ADN tels que le repliement sous la forme d'un boucle et donc les prendre en compte explicitement dans la régulation du gène (Saiz et Vilar, 2006b). L'énergie associée à ce repliement, qui dépend de la longueur et du type de boucle formée, est par ailleurs caractérisée de façon relativement précise aussi bien expérimentalement que théoriquement (Saiz *et al.*, 2005).

Un tel formalisme ne permet cependant pas de décrire le comportement dynamique du promoteur et donc le caractère stochastique de l'expression du gène qui en résulte. En effet, il ne comporte aucune information cinétique puisque les différences d'énergie entre deux états ne permettent que de connaître la constante d'équilibre de la transition (c'est-à-dire le rapport des constantes cinétiques dans les deux sens). De fait, appliqué aux deux systèmes présentés sur les figures I.2e1 et I.2e2 (page 26) où les constantes d'équilibre sont identiques, mais où les taux de transition entre états du promoteur diffèrent fortement, ce formalisme décrirait dans les deux cas une équiprobabilité des deux états du promoteur et donc une même expression moyenne. Pourtant, c'est bien ici la dynamique de l'alternance entre ces deux états équiprobables qui est à l'origine des différences que l'on observe dans l'expression stochastique du gène. Certains travaux ont alors pris en compte les constantes cinétiques explicitement en les choisissant de sorte à respecter les différences d'énergie libre entre états du promoteur (Vilar et Saiz, 2005; Saiz et Vilar, 2006a). Mais ces travaux n'exploitent alors le modèle résultant que par la simulation.

Par ailleurs, l'hypothèse d'un équilibre thermodynamique contraint implicitement cette approche à ne représenter que des systèmes fermés, c'est-à-dire des systèmes ne comportant aucune réaction consommant de l'énergie extérieure. Ainsi, lorsqu'il s'agit d'appliquer ce cadre théorique à la régulation eucaryote, il est impossible de considérer le caractère énergie-dépendant des réactions de remodelage chromatinien et de modification covalente des histones. Malgré cette limitation, ce formalisme est couramment employé dans ce cadre, en associant des énergies à différents états conformationnels de la chromatine (en particulier à la position des nucléosomes le long de la séquence d'ADN) et à leur interaction avec différentes molécules (Segal et Widom, 2009; Jiang et Pugh, 2009; Raveh-Sadka *et al.*, 2009). L'usage d'un formalisme thermodynamique pour la régulation eucaryote est alors souvent justifié en avancant que les remodelage chromatiniens ne servent qu'à fluidi-

[2]Il s'agit plus précisément du principe d'*équilibre détaillé* selon lequel toutes les réactions ont lieu autant dans un sens que dans l'autre.

fier les réarrangements conformationnels en facilitant le passage de barrières énergétiques importantes et que l'hypothèse d'équilibre thermodynamique reste donc valide (Segal et Widom, 2009). Nous verrons pourtant dans ce chapitre que cette justification n'est en réalité pas valide pour décrire la dynamique des promoteurs eucaryotes. En particulier le comportement clairement périodique montré par les cinétiques des ChIP et illustré sur la figure I.9 (page 57) est caractéristique d'un système hors d'équilibre. Il est donc capital de pouvoir prendre en compte explicitement la consommation d'énergie extérieure que représentent ces réactions.

1.4 Modèles de la dynamique du promoteur

Les modèles décrivant spécifiquement les propriétés dynamiques des promoteurs eucaryotes tels que mises en évidence par les expériences de cinétique de ChIP (figure I.9, page 57) sont plutôt rares. Lemaire *et al.* (2006) ont proposé un modèle considérant que l'état du promoteur suit une séquence cyclique de transitions, reproduisant alors le comportement périodique observé expérimentalement. Ce modèle, bien que relativement restreint puisque ne décrivant qu'une unique séquence de transitions, permet de montrer que la longueur du cycle, ainsi que le caractère plus ou moins réversible et plus ou moins homogène des transitions, joue un rôle important dans la dynamique. Ces propriétés appelent alors à davantage d'attention, en particulier dans le cadre d'un modèle plus complet.

Benecke (2006) montre, sur un modèle d'automate stochastique représentant les modifications de la chromatine en fonction de leurs influences combinatoires, que la dynamique résultante peut en particulier produire des comportements périodiques. Ce modèle a l'avantage de considérer des modifications chromatiniennes non restreintes à une unique séquence de transitions, rendant alors compte de la complexité potentielle du code des histones (voire d'un code chromatine considérant également les phénomènes de remodelage (Lesne, 2006)), mais ne fait cependant pas intervenir explicitement l'association/dissociation des molécules sur le promoteur.

Dans les travaux récents de Degenhardt *et al.* (2009), les auteurs implémentent le concept du *transcriptional ratchet* (cf section 5.5, page 59) afin de reproduire le caractère périodique des cinétiques de ChIP. Ils considèrent explicitement l'association/dissociation de molécules sur le promoteur et une séquence cyclique de 6 modifications de la chromatine supposées totalement irréversibles. Ces travaux sont exclusivement basés sur de la simulation, à l'exception d'une propriété mise en évidence en considérant un modèle plus simple (identique à celui de Lemaire *et al.* (2006)) : les auteurs montrent que les oscillations observées sur une cinétique de ChIP s'amortissent au cours du temps, et ce d'autant plus que le cycle est court – une propriété par ailleurs déjà énoncée par Lemaire *et al.* (2006). Cette approche montre cependant l'intérêt d'utiliser un modèle plus complet dans lequel l'ordre d'assemblage des molécules peut être plus ou moins aléatoire (même si les modifications de la chromatine suivent une séquence préétablie). Les auteurs montrent en particulier que le modèle reproduit mieux les caractéristiques des données expérimentales lorsque l'assemblage des molécules suit un processus qualifié de *preferentially random*, intermédiaire entre un assemblage purement séquentiel et un assemblage totalement désordonné.

Ainsi, si les modèles rendant compte de la dynamique périodique lente des promoteurs eucaryotes sont peu nombreux, ceux mettant en relation cette dynamique avec le caractère très mobile des molécules, tel qu'il est montré par les expériences de FRAP, le sont encore plus. Dans les travaux de Degenhardt *et al.* (2009), les auteurs choisissent les taux de dissociation des molécules de façon à être cohérents avec ces données, mais aucun modèle ne permet à ce jour de dériver d'un même formalisme des données de ChIP et des données de FRAP. Ces dernières sont classiquement analysées par des modèles différents et relativement simples où l'on considère que la molécule d'intérêt peut être associée selon différents *états d'association*, ayant chacun un temps de résidence donné et provoquant alors différentes composantes exponentielles additives dans la récupération de la fluorescence (cf Box 2, page 52). Ainsi le rapport entre les dynamiques temporelles rapides et lentes reste relativement peu exploré du point de vue de la modélisation.

***

Dans la section suivante, nous présentons un modèle destiné à répondre à un tel besoin et dont le but est d'aborder des questions concernant aussi bien la génération, par la dynamique du promoteur, d'une stochasticité de l'expression génique structurée que concernant la dynamique multi-échelle du promoteur. Il est dans ce cadre capital de pouvoir dériver, d'un même formalisme, à la fois les indicateurs employés pour caractériser l'expression stochastique des gènes – tels que la variance normalisée, la distribution stationnaire et le spectre de puissance d'un niveau d'expression (cf Box 1, page 22) – et les mesures expérimentales utilisées dans l'étude de la dynamique du promoteur – tels que des données de FRAP, de FRET et de cinétique de ChIP (cf Box 2, page 52). Notre objectif est, de plus, de proposer une modélisation générique permettant de décrire des systèmes très différents et de tirer des conclusions générales. Le modèle doit aussi être en mesure de représenter à la fois la combinatoire d'association/dissociation des molécules et des modifications conformationnelles et covalentes. Il doit autoriser la représentation de systèmes ouverts en adoptant une description énergétique afin de quantifier l'apport énergétique extérieur des réactions de remodelage et de modification d'histones. Enfin, il doit privilégier une exploitation d'un point de vue essentiellement analytique, faisant éventuellement intervenir des étapes de résolution numérique.

2 Modèle biophysique

Nous proposons dans cette section un modèle d'expression génique stochastique centrée sur une représentation détaillée du promoteur. Nous gardons toutefois explicites, bien que très simplifiées, les étapes de transcription et de traduction afin de décrire l'influence de la dynamique stochastique du promoteur sur l'expression du gène.

2.1 Formulation générale

Le modèle que nous proposons décrit la transition du promoteur entre différents états s au sein d'un ensemble de N états, noté $\mathcal{S}$. Un état du promoteur est déterminé par

- **la combinaison des molécules sur le promoteur :** molécules présentes sur un site de fixation ou au sein d'un complexe multiprotéique fixé sur le promoteur,
- **la position ou l'état conformationnel de certains complexes :** position des nucléosomes le long de l'ADN, contact entre nucléosomes, repliement de l'ADN, ...
- **les modifications post-traductionnelles de certaines molécules :** acétylation, méthylation, phosphorylation, ubiquitilation, ... de certains résidus de queue d'histone ou méthylation de certains sites CpG du promoteur.

(cf figure III.1 pour plus de détails).

Nous désignons de façon générale par *facteur de transcription* (FT) toute molécule pouvant faire partie de la structure multiprotéique de régulation (incluant les cofacteurs, histones, ligands, enzymes, ...) et par *facteur chromatinien* (FC) toute modification covalente ou conformationnelle. Le terme *facteur* pourra désigner à la fois des FT et des FC. Par

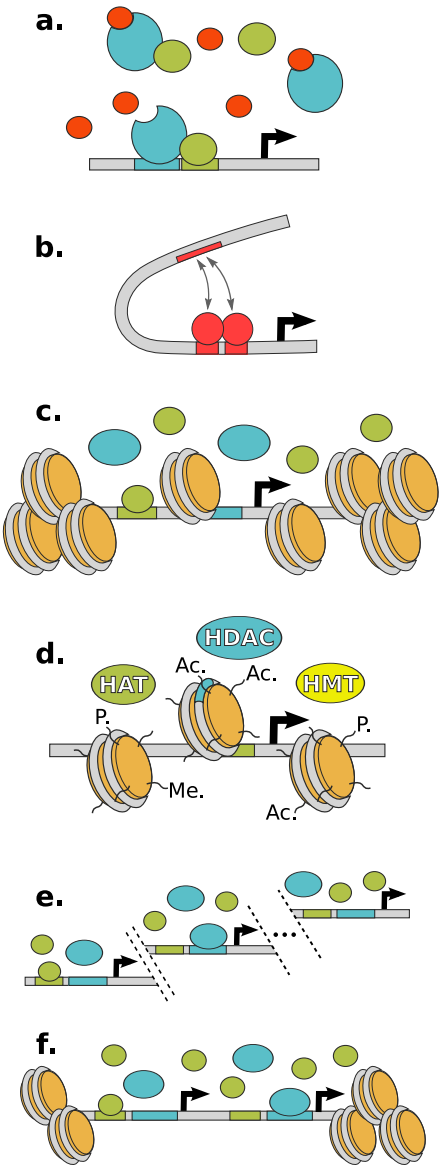


Fig. III.1: Pouvoir descriptif du modèle. De nombreux aspects de l'état du promoteur d'un gène et de sa dynamique peuvent être décrits par le modèle présenté dans ce chapitre. (a) Il permet de représenter l'association/dissociation de molécules sur leur site de fixation ou au sein d'une structure multiprotéique présente sur le promoteur, ainsi que toutes sortes de phénomènes de coopération/compétition et d'influence mutuelle combinatoire entre molécules. De plus, ces dernières peuvent s'associer et se dissocier du promoteur aussi bien séparément que sous la forme de complexes. Le modèle permet également de décrire divers aspects de l'état conformationnel de la structure tels que (b) le repliement de l'ADN sous la forme d'une boucle, (c) la position des nucléosomes le long de la séquence, l'enroulement/déroulement de l'ADN autour des nucléosomes, différentes conformations possibles d'un complexe associé au promoteur, ... (d) L'état du promoteur peut également prendre en compte les nombreuses modifications covalentes des histones (et de l'ADN). Une caractéristique majeure du modèle est la possibilité de représenter les influences réciproques et combinatoires entre les modifications conformationnelles et covalentes et la présence de différents FT, intégrant en particulier explicitement l'activité enzymatique ou de remodelage de certains complexes et la capacité de la chromatine à influencer leur association, ces deux mécanismes étant essentiels à la notion de code des histones (cf chapitre 1, section 5.1). Enfin, on peut également représenter différentes constructions telles que (e) un nombre arbitraires de copies indépendantes du gène ou (f) deux copies du gène partageant un même environnement chromatinien. Cette seconde construction a en particulier été utilisée expérimentalement pour montrer que la dynamique locale du promoteur peut constituer une source majeure de stochasticité dans l'expression d'un gène (Becskei *et al.*, 2005; Raj *et al.*, 2006).

abus de langage, on pourra alors parler d'association/dissociation ou de présence/absence d'un facteur même s'il s'agit d'un FC.

Nous considérons chaque facteur comme ayant un statut binaire, décrivant si la molécule en question fait partie ou non de la structure ou si la modification covalente est présente ou non. Mais les états conformationnels ne sont en réalité pas nécessairement binaires (un nucléosome peut occuper un continuum de positions le long de la séquence). Ils peuvent alors être représentés en théorie par une combinaison de plusieurs facteurs, et ce avec un niveau de détail arbitrairement fin. En employant 5 « facteurs », on peut par exemple représenter la position d'un nucléosome le long de la séquence selon 32 emplacements différentes (puisque $2^5 = 32$). Chacun de ces 5 facteurs ne représente alors pas une propriété donnée mais c'est leur combinaison qui décrit la position du nucléosome. Même si nous n'emploierons pas ce niveau de détail par la suite, il est important de mentionner que le formalisme que nous présentons ici permet une telle représentation. En outre, l'interaction entre un nucléosome et la séquence d'ADN décrit souvent un profil énergétique avec plusieurs bassins séparés par de hautes barrières énergétiques, rendant alors relativement pertinente l'approximation de leur position par un faible nombre d'états discrets.

L'ensemble des n facteurs (incluant les FT et les FC) est noté $\mathcal{F}$ et permet 2^n combinaisons. L'ensemble $\mathcal{S}$ des états du promoteur compte donc au maximum 2^n éléments : $\mathcal{S} \subseteq 2^{\mathcal{F}}$ (où $2^{\mathcal{F}}$ est l'ensemble des parties de $\mathcal{F}$). Par exemple, pour un promoteur ayant $n = 3$ facteurs $\mathcal{F} = \{A, B, C\}$ (figure III.2a), les combinaisons possibles sont $2^{\mathcal{F}} = \{\varnothing, A, B, C, AB, AC, BC, ABC\}$ (figure III.2b). Mais, si l'on prend en compte le fait que certains états ne peuvent avoir lieu, l'ensemble $\mathcal{S}$ des états du promoteur n'est qu'un sous ensemble de $2^{\mathcal{F}}$. Ainsi, si, sur ce promoteur, A et B sont deux molécules qui s'excluent mutuellement parce qu'elles partagent le même site de fixation, ou si A représente une position d'un nucléosome empêchant la fixation de B, on a alors $\mathcal{S} = \{\varnothing, A, B, C, AC, BC\} \subseteq 2^{\mathcal{F}}$. Cette simplification de l'espace des états est utile d'un point de vue computationnel, mais dans le cas général on peut considérer que $\mathcal{S} = 2^{\mathcal{F}}$, certains états étant éventuellement inatteignables.

Nous décrivons la dynamique du promoteur au sein de cet ensemble d'états, ainsi que les processus de transcription et de traduction par le système de réactions biochimiques suivant :

$$\mathcal{G}_s \xrightarrow{M_{s',s}} \mathcal{G}_{s'}, \qquad \forall (s, s'), s \neq s' \tag{III.1}$$

$$\mathcal{G}_s \xrightarrow{\rho_s} \mathcal{G}_s + \mathcal{R}, \qquad \forall s \tag{III.2}$$

$$\mathcal{R} \xrightarrow{\gamma} \varnothing \tag{III.3}$$

$$\mathcal{R} \xrightarrow{\tilde{\rho}} \mathcal{R} + \mathcal{P} \tag{III.4}$$

$$\mathcal{P} \xrightarrow{\tilde{\gamma}} \varnothing \tag{III.5}$$

où $\mathcal{R}$ et $\mathcal{P}$ sont les molécules d'ARN et de protéine et $\mathcal{G}_s$ correspond au gène avec son promoteur dans l'état $s \in \mathcal{S}$ (par exemple, $\mathcal{G}_{AC}$ représente le gène et son promoteur lorsque seules les molécules A et C sont présentes sur leur site de fixation). Les différentes réactions de ce système correspondent respectivement à un changement d'état du promoteur (association/dissociation de FT, modification de FC), un événement de transcription, la

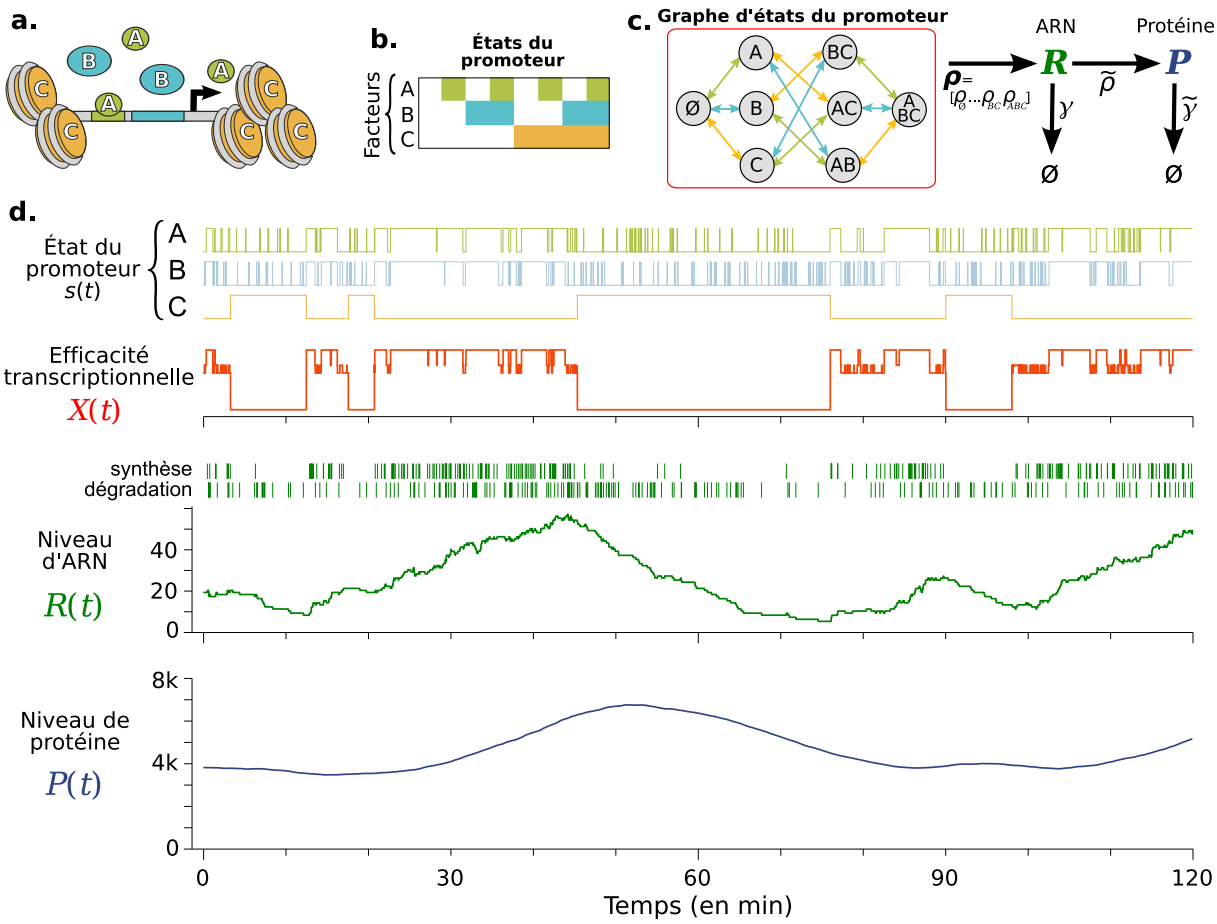


Fig. III.2: Instanciation et simulation d'un système très simple. Cette figure illustre le modèle présenté dans ce chapitre par un exemple-jouet dans lequel l'état du promoteur est déterminé par 3 facteurs (2 FT et un FC). (a) Deux molécules A et B s'associent de façon coopérative au promoteur et ont une meilleure affinité avec leur site de fixation lorsque la chromatine (représentée ici de manière simpliste par un unique FC noté C) est ouverte. (b) Pour ces 3 facteurs (lignes), l'énumération des 8 combinaisons possibles décrit l'ensemble des états du promoteur (colonnes). (c) Le modèle correspond à un système de réactions biochimiques stochastiques représenté schématiquement ici. La réaction (III.1) définit le graphe d'état du promoteur (encadré en rouge), graphe orienté et pondéré, dont la matrice d'adjacence est notée $\mathbf{M}$ et décrivant à lui seul toute la dynamique du promoteur. La réaction (III.2) définit que la probabilité de synthèse des transcrits est déterminée par l'état courant du promoteur selon un taux ρ_s propre à chaque état. De façon similaire aux modèles classiques d'expression génique (figure I.2a, page 26) la traduction est proportionnelle au niveau d'ARN et les protéines et les transcrits subissent une dégradation standard de premier ordre. (d) La simulation de ce système par un algorithme de Gillespie (1977) permet de visualiser un exemple de réalisation de sa dynamique stochastique. On peut observer (de haut en bas) les événements d'association/dissociation ou de modification des 3 facteurs (décrivant collectivement une marche aléatoire sur le graphe d'état du promoteur), l'efficacité transcriptionnelle du promoteur (capacité à initier des événements de transcription), les événements stochastiques de synthèse et de dégradation des transcrits, l'évolution du niveau d'ARN en résultant et l'évolution du niveau de protéines (par le même type de processus stochastique de synthèse/dégradation que pour l'ARN, mais non montré ici).

dégradation d'un transcrit, un événement de traduction et la dégradation d'une protéine. Ce modèle, schématisé sur la figure III.2c, est similaire à celui présenté au chapitre 1 (section 2.1) en ce que concerne les réactions (III.3) à (III.5) mais est beaucoup plus détaillé sur la dynamique du promoteur (III.1) et sur son influence sur les événements de transcription (III.2). La réaction (III.1) décrit l'ensemble des transitions possibles de l'état du promoteur et peut alors être représentée sous la forme d'un graphe dans lequel chaque nœud est un état $s \in \mathcal{S}$ et chaque arc, pondéré et orienté, est un taux de transition[3] $M_{s',s}$ (figure III.2c). De plus, contrairement à la représentation binaire classique de l'activité du promoteur employée par la plupart des modèles d'expression stochastique (cf chapitre 1, section 2.4), la réaction (III.2) associe spécifiquement à chaque état du promoteur une efficacité transcriptionnelle donnée, décrite par un taux de transcription ρ_s.

Tous les taux de transition $M_{s',s}$ entre les différents états du promoteur peuvent être représentés par une matrice, notée $\mathbf{M}$, correspondant simplement à la matrice d'adjacence du graphe d'état. Elle permet de représenter tous les phénomènes de coopération/compétition et d'influences cinétiques mutuelles entre molécules (Browning et Busby, 2004; Dodd *et al.*, 2004; Adams et Workman, 1995; Agresti *et al.*, 2005; Mellor, 2006; Li *et al.*, 2007). Classiquement une molécule A s'associe sur son site de fixation avec un taux $k_{\text{on}}[A]$ (où $[A]$ dénote la concentration de A) et se dissocie à un taux k_{off}. Dans notre modèle, ces deux taux dépendent de la combinaison de tous les autres facteurs (FT et FC). Par exemple, $M_{\varnothing,A} > M_{B,AB}$ et $M_{A,\varnothing} = M_{AB,B}$ signifie que B coopère avec A en le stabilisant sur son site. En revanche, si $M_{\varnothing,A} > M_{B,AB}$ et $M_{\varnothing,A}/M_{A,\varnothing} = M_{B,AB}/M_{AB,B}$, alors B ne fait qu'augmenter le turnover de A sans modifier son équilibre. Enfin, $M_{AC,C} = M_{ABC,BC}$ et $M_{C,AC} = M_{BC,ABC}$ signifie que toute influence de B sur A disparaît lorsque C est présent.

Un aspect essentiel du modèle est que l'on peut également représenter les interactions mutuelles entre les associations/dissociations de molécules ou de complexes et les modifications conformationnelles et covalentes de la structure. On peut par exemple choisir les valeurs de $\mathbf{M}$ de façon à représenter le fait que, dans certains états conformationnels, le site de fixation d'un FT donné est peu accessible, rendant alors plus difficile son association, ou que certaines combinatoires de modifications covalentes des queues d'histones peuvent favoriser la fixation de certaines FT ou encore influencer les propriétés d'interaction des nucléosomes avec l'ADN ... Enfin, il est possible de décrire explicitement l'activité enzymatique ou de remodelage de certaines molécules ou de certains complexes sur la chromatine (figure III.1d). Par exemple, les valeurs de $\mathbf{M}$ peuvent être telles qu'une modification covalente donnée d'une queue d'histone ne puisse être réalisée que lorsque qu'un complexe prédéfini est présent sur le promoteur, ou bien telles qu'un nucléosome ne puisse être déplacé ou expulsé que lorsque qu'un complexe de remodelage donné est présent.

La figure III.2 présente un exemple de simulation dans le cas d'un système très simple à 3 facteurs (2 FT et un FC, ce dernier représentant ici de façon très simpliste l'état ouvert ou fermé de la chromatine). Les trois premières courbes décrivent respectivement les événements d'association/dissociation des deux FT et d'ouverture/fermeture de la chromatine, correspondant à une marche aléatoire – biaiséé par les taux de transition de la matrice $\mathbf{M}$ – sur le graphe d'état. Dans cet exemple, la chromatine est sensiblement plus lente que les

[3]Remarquons que le taux de la transition de l'état s vers l'état s' est $M_{s',s}$ (et non $M_{s,s'}$).

FT et influence leurs temps de résidence (cf légende). La courbe suivante décrit l'évolution de l'efficacité transcriptionnelle du promoteur au cours du temps, c'est-à-dire la valeur du taux de transcription ρ_s associé à l'état courant. La dynamique du promoteur influence alors les événements de synthèse du processus de synthèse/dégradation du niveau d'ARN, lui-même influençant le processus de synthèse/dégradation des protéines.

L'essentiel de l'expressivité de ce modèle correspond à la réaction (III.1). Matérialisée par le graphe d'état du promoteur et paramétrée uniquement par la matrice $\mathbf{M}$, elle décrit de façon très simple la dynamique de promoteurs potentiellement très complexes. Nous nous concentrons, dans le reste de la section 2 ainsi que dans le début de la section 3, plus particulièrement sur cet objet.

2.2 Structure et contenu de la matrice de transition

La matrice $\mathbf{M}$, définissant tous les taux de transition dans le graphe d'état du promoteur, est contrainte à différents niveaux. D'une part, la façon dont ont lieu les modifications de l'état du promoteur lui impose un certain nombre de contraintes structurelles et, d'autre part, ses valeurs, intégrant implicitement la concentration des FT, dépendent en partie de ces dernières. De plus, comme nous le décrirons dans la section 2.4, la reformulation du modèle d'un point de vue énergétique fait également ressortir certaines contraintes sur les valeurs de $\mathbf{M}$ liées au caractère énergie-dépendant ou non des différentes réactions.

Contraintes structurelles

Considérons temporairement que tous les facteurs $f \in \mathcal{F}$ sont des FT. Le taux de transition du promoteur d'un état s vers un état s', noté $M_{s',s}$, correspond à une association si $s' \supset s$ et à une dissociation si $s' \subset s$ (rappelons ici qu'un état est décrit par l'ensemble des facteurs présents sur le promoteur). Par exemple, $\mathcal{G}_A \rightarrow \mathcal{G}_{ABC}$ décrit l'association du complexe BC sur un promoteur où seule la molécule A est présente. Si s n'est ni un sous-ensemble ni un sur-ensemble de s', la réaction $\mathcal{G}_s \rightarrow \mathcal{G}_{s'}$ ne peut avoir lieu en une seule étape et $M_{s',s} = 0$. Ainsi, la transition de $\mathcal{G}_{AB}$ à $\mathcal{G}_{AC}$ s'effectue nécessairement au moins en deux étapes : dissociation de B, puis association de C ($\mathcal{G}_{AB} \rightarrow \mathcal{G}_A \rightarrow \mathcal{G}_{AC}$) ou l'inverse ($\mathcal{G}_{AB} \rightarrow \mathcal{G}_{ABC} \rightarrow \mathcal{G}_{AC}$). Même si l'état $\mathcal{G}_{ABC}$ peut éventuellement être très instable (auquel cas $\mathcal{G}_{ABC} \rightarrow \mathcal{G}_{AC}$ se fait très rapidement), C ne peut pas remplacer B en une seule étape cinétique. Comme l'illustre la figure III.3a, cette propriété réduit grandement l'espace des paramètres en impliquant qu'un grand nombre d'éléments de $\mathbf{M}$ sont nuls[4].

Bien que plusieurs FT puissent être modifiés ensemble lors d'une même transition (correspondant à l'association ou à la dissociation d'un complexe), dans de nombreux cas, la transition d'un FC se fait seule. De fait, même si une modification post-traductionnelle ou

[4]On remarque que les valeurs de $\mathbf{M}$ qui restent libres (figure III.3a) forment une structure fractale similaire à celle de l'automate de Sierpiński ou au motif de divisibilité par 2 du triangle de Pascal (Peitgen *et al.*, 2004). Cette structure ayant une dimension de Hausdorff de $\log 3/\log 2 \simeq 1.585$, le nombre de paramètres libres de $\mathbf{M}$ n'augmente pas comme N^2 mais comme $N^{1.585}$.

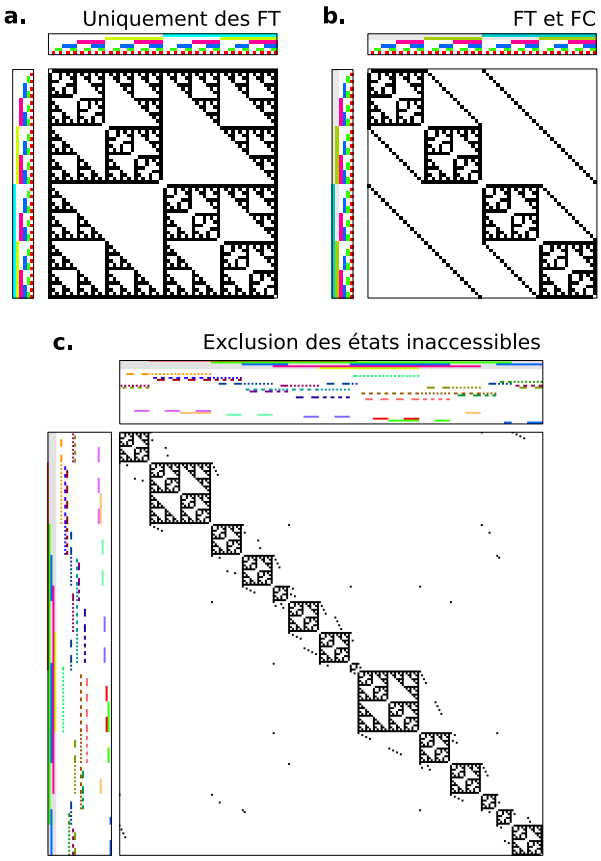


Fig. III.3: Contraintes structurelles sur la matrice $\mathbf{M}$. La matrice $\mathbf{M}$ décrit les taux de transition entre chaque paires d'états. Cependant, le fait que certaines transitions ne peuvent s'effectuer (en une seule étape élémentaire) contraint de très nombreux éléments à être nuls (ici en blanc). (a) Lorsque tous les facteurs sont des FT, toute transition $\mathcal{G}_s \rightarrow \mathcal{G}_{s'}$ qui n'est ni une association ($s \subset s'$) ni une dissociation ($s \supset s'$) d'un FT ou d'un complexe ne peut avoir lieu[4]. (b) La plupart des modifications de FC constituent une étape élémentaire et ne peuvent donc avoir lieu en même temps qu'une association ou qu'une dissociation de FT, impliquant alors davantage de valeurs nulles. (c) Lorsque les états inaccessibles sont éliminés (comme c'est le cas ici pour l'exemple que nous présenterons à la figure III.4), les éléments non-contraints sont encore moins nombreux et peuvent ne représenter qu'une faible fraction des valeurs de $\mathbf{M}$.

un remodelage peut produire un état très instable pour un FT donné qui se dissocie alors très rapidement, le phénomène se produit en (au moins) deux étapes. Cette contrainte supplémentaire sur l'espace des paramètres rend nuls davantage de valeurs de $\mathbf{M}$ (figure III.3b). Les exceptions à cette règle concernent par exemple les cas où l'on considère explicitement les histones comme pouvant s'associer et se dissocier (des FT au sens défini précédemment). Dans ce cas, les modifications covalentes de la queue d'une histone sont perdues lors de sa dissociation, de même que la position d'un nucléosome le long de l'ADN lorsque celui-ci est expulsé. Il est cependant tout à fait possible de prendre en compte ces phénomènes, mais nous n'irons pas ici dans un tel niveau de détail. Notons également que si l'on emploie une combinaison de plusieurs « facteurs » pour décrire de façon précise la position d'un nucléosome, les contraintes sur $\mathbf{M}$, sont évidemment également différentes.

Comme évoqué précédemment, certains états du promoteur peuvent, dans de nombreux cas, ne pas représenter une configuration physiquement réalisable. L'élimination de ces états simplifie encore significativement l'espace paramétrique, comme l'illustre la figure III.3c sur un exemple que nous décrirons dans la section 2.3. Notons finalement que des critères structurels de contacts physiques entre molécules peuvent également permettre de simplifier l'espace paramétrique[5].

[5]Par exemple, si B et C ne forment pas de contacts au sein de la structure mais que D touche à la fois B et C, alors $\mathcal{G}_{ABCD} \rightarrow \mathcal{G}_{AD}$ ne peut se faire en une seul étape alors que $\mathcal{G}_{ABCD} \rightarrow \mathcal{G}_{A}$ le peut.

Dépendance à la concentration des FT

La dynamique d'un système donné peut être étudiée à partir de la simple définition de $\mathbf{M}$. Cependant, cette matrice dépend des concentrations des différents FT, de sorte qu'il est possible, en détaillant davantage sa structure, d'étudier comment le comportement d'un système change en fonction des concentrations des FT. En effet, les éléments de $\mathbf{M}$ sont des constantes cinétiques d'ordre 0 dont celles qui représentent des associations (cas où $s' \supset s$) résultent de la multiplication d'une constante d'ordre 1 et de la concentration du complexe qui s'associe (qu'il s'agisse d'un FT libre ou d'un complexe de plusieurs FT). On peut alors définir une matrice de référence $\mathbf{M}^0$ permettant de déduire $\mathbf{M}$ à partir de l'ensemble des concentrations des FT et des complexes. On remarque que, dans le cas d'une association, $s' - s$ correspond au complexe qui s'associe et vaut $\varnothing$ pour une dissociation[6]. En prenant par convention que $[\varnothing] = 1$ et que $[f] = 1$ si f est un FC (la concentration de f n'ayant pas de signification physique dans ce cas), on peut écrire

$$M_{s',s} = [s' - s]M^0_{s',s} \quad \forall s, s' \tag{III.6}$$

De plus, la concentration des FT libres et des différents complexes résulte de leurs interactions hors du promoteur. La modification de la quantité totale d'un FT influence donc non seulement sa concentration libre mais également celle des complexes (qu'il en fasse ou non partie). Ceci peut être décrit de façon classique en définissant les concentrations totales des FT (notées $[f^*], \forall f \in \mathcal{F}$) et un système de réactions (noté $\mathbf{\Theta}$) dont l'état stationnaire donne toutes les concentrations des FT libres et des complexes[7].

2.3 Exemple d'instanciation : un promoteur eucaryote

Afin d'illustrer différentes notions abordées jusqu'ici, nous introduisons un exemple d'instanciation du modèle représentant de façon simplifiée un promoteur eucaryote (figure III.4) en nous appuyant en particulier sur les caractéristiques évoqués au chapitre 1, section 5. Cet exemple sera réutilisé par la suite, mais rappelons ici qu'il ne s'agit que d'une instanciation particulière d'un modèle générique et que toutes les dérivations des sections 2 et 3 s'appliquent au cas général.

Dans cet exemple, nous considérons 26 facteurs : 5 FC et 21 FT (figure III.4a1). Parmi les 21 FT, 14 réalisent explicitement les modifications des FC. Comme décrit sur la figure III.4a2, chaque modification d'un FC peut être réalisée par un FT ou une combinaison de FT particuliers (représentant un complexe de remodelage ou une enzyme de modification d'histone), certains pouvant éventuellement être fonctionnellement redondants. Nous définissons également que chaque état de la chromatine (combinaison de FC) détermine le spectre de FT pouvant s'associer au promoteur (figure III.4a3). Ainsi, l'ensemble des états

[6] De façon générale, $s' - s$ (resp. $s - s'$) représente le FT ou le complexe qui s'associe (resp. se dissocie) lorsque $\mathcal{G}_s \to \mathcal{G}_{s'}$ est une association (resp. dissociation). $s' \ominus s = (s' - s) \cup (s - s')$ (où $\ominus$ est l'opérateur de *différence symétrique*) correspond dans tous les cas au complexe qui s'associe ou se dissocie.

[7] Par exemple, si $[A^*] = 6$, $[B^*] = 4$ et $[C^*] = 6$ et que $\mathbf{\Theta} = \{A + B \overset{K_1=3}{\rightleftarrows} AB,\ A + C \overset{K_2=.5}{\rightleftarrows} AC\}$ alors on a $[A] = 1$, $[B] = 3$, $[C] = 2$, $[AB] = 1$ et $[AC] = 4$.

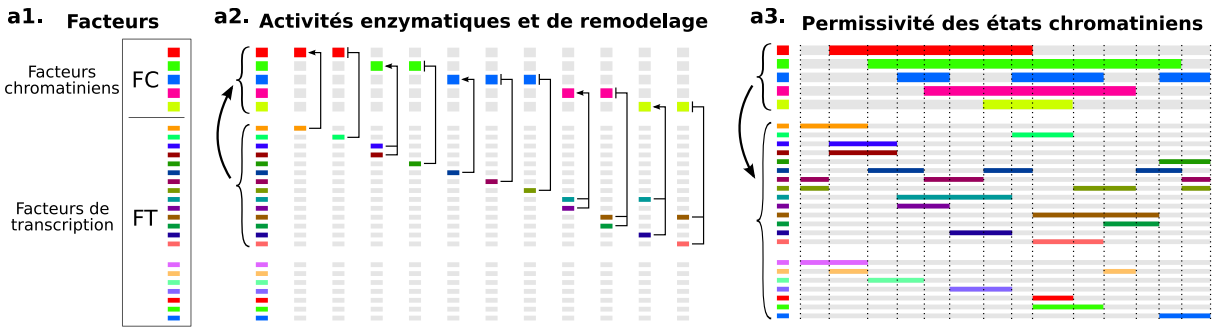
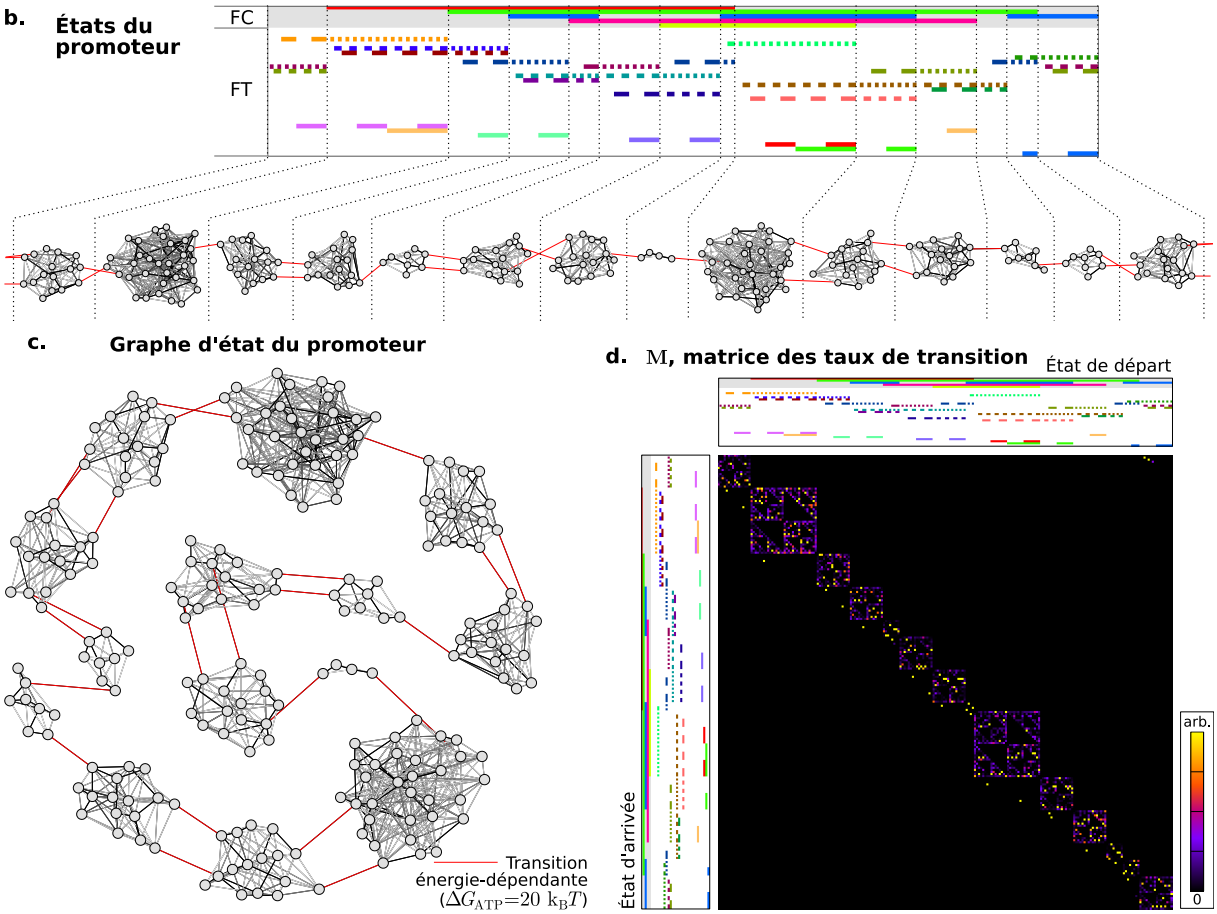


Fig. III.4: Représentation d'un promoteur eucaryote. Cet exemple d'instanciation du modèle permet d'illustrer plusieurs notions centrales et montre comment il possible de représenter explicitement différentes caractéristiques essentielles des promoteurs eucaryotes (cf chapitre 1, section 5). (a1) Dans cet exemple, chacun des 26 facteurs (5 FC et 21 FT) est représenté par une couleur donnée. (a2) Chaque modification des 5 FC est due à l'action de FT/complexes spécifiques ayant des propriétés de remodelage ou de modification des histones. Chaque FC ne peut ainsi être basculé d'un état à l'autre qu'en la présence de ces FT/complexes particuliers. On peut de plus représenter explicitement et quantitativement l'apport énergétique extérieur au système occasionné par l'hydrolyse des coenzymes lors de chacune de ces réactions (cf section 2.4). (a3) On définit également comment l'état de la chromatine détermine le spectre de molécules pouvant s'associer ou non au promoteur (décrit simplement ici de façon binaire, ce qui n'est pas nécessairement le cas). Par exemple, la troisième colonne de ce tableau indique que lorsque seuls les FC ■ et ■ sont « actifs », seuls les FT ■, ■, ■ et ■ peuvent s'associer au promoteur. Ces interactions réciproques FC/FT (a2 et a3), définissant un *code des histones*, ont été choisies de façon relativement simple : les modifications de la chromatine ne peuvent ici suivre qu'une séquence cyclique prédéterminée (la même que celle des colonnes de a3). (b) L'ensemble des états possibles du promoteur peut être obtenu en énumérant, pour chaque état chromatinien, toutes les combinaisons des FT pouvant s'associer au promoteur (similairement à la figure III.2b). Chaque colonne représente une unique combinaison de facteurs. (c) Le *graphe d'état du promoteur* est un graphe orienté et pondéré dans lequel chaque nœud correspond à un état du promoteur et chaque arc à une transition élémentaire (réaction (III.1)). (d) La matrice d'adjacence de ce graphe, notée $\mathbf{M}$, contient les taux de transition entre chaque paire d'états et comporte un certain nombre de valeurs nulles (cf figure III.3c). La dynamique stochastique du promoteur décrit alors simplement une marche aléatoire sur ce graphe. Pour le promoteur défini ici, le graphe d'état et la matrice $\mathbf{M}$ reflètent clairement la structure du système : chaque module fortement connecté

du promoteur peut être obtenu en énumérant, pour chaque combinaison de FC, toutes les combinaisons possibles des FT pouvant s'associer (figure III.4b). L'activité enzymatique et de remodelage des FT et la permissivité sélective de la chromatine ont été choisies de sorte à ce que les modifications de la chromatine décrivent un cycle. Par exemple, à partir du premier état chromatinien de la figure III.4a3, la seule modification possible de la chromatine (d'après la figure III.4a2) conduit au deuxième état chromatinien de la figure III.4a3, lui-même ne pouvant conduire qu'au troisième, etc.

Le graphe des états du promoteur et des transitions entre états résultant de cette définition est représenté sur la figure III.4c. Il s'agit du même objet que celui représenté, dans le cas d'un système beaucoup plus simple, sur la figure III.2c (encadré en rouge) et la dynamique du promoteur correspond également à une marche aléatoire sur ce graphe



(Fig. III.4, cont. : cf page précédente)
correspond à un état chromatinien. Les transitions intra-module représentent des associations/dissociations de FT/complexes et sont définies comme étant énergie-indépendantes. Les transitions inter-modules sont les modifications de la chromatine (décrites par a1) et sont les seules à être énergie-dépendantes. Bien que la structure du graphe d'état soit déterminée par (a1-3), sa pondération (échelle de gris des arcs) ne l'est pas. Ainsi, le système peut encore être sous-instancié en choisissant les valeurs libres de $\mathbf{M}$ de façon à représenter des mécanismes d'interaction entre molécules très différents (cf section 4.4).

orienté et pondéré. Le graphe d'état du promoteur (figure III.4c) ainsi que sa matrice de transition $\mathbf{M}$ (figure III.4d) présentent certaines caractéristiques reflétant la structure du système. On observe en effet un certain nombre de modules très connectés correspondant chacun à un état chromatinien et décrivant les associations/dissociations de FT et de complexes sur la chromatine. Les transitions entre ces modules (représentées en rouge sur le graphe) sont des événement de modifications de la chromatine (tels que décrits par la figure III.4a2). Ces transitions nécessitent la formation d'un complexe spécifique pour avoir lieu[8] et sont donc moins nombreuses que les associations/dissociations de FT, provoquant alors le caractère très modulaire du graphe. À une échelle supérieure, les transitions entre les différents modules décrivent un cycle de modification de la chromatine

[8]Notons que, lorsque le complexe de modification de la chromatine est formé, la présence ou non d'une autre molécule n'en modifie pas l'activité. Cela résulte alors en deux transititions d'un module à un autre.

qui dans cet exemple correspond à une unique séquence mais pourrait tout à fait être instancié différemment, en rajoutant par exemple différents chemins alternatifs. En effet, les figures III.4a2 et III.4a3, décrivant les interactions mutuelles entre la chromatine et les complexes réalisant ses modifications, représentent directement le code des histones. La structure du graphe d'état résultant de ce code peut être très variée et peut constituer un objet d'étude en soi.

Cependant, bien que le graphe de transition ait, dans cet exemple, une structure prédéfinie, sa pondération – à savoir, le choix des taux de transition définissant l'affinité des différentes molécules, les phénomènes de coopération/compétition et d'influence mutuelle combinatoire, l'efficacité des réactions enzymatiques, ... – n'est pas encore fixée. On peut alors étudier, pour un tel système, comment la nature des interactions entre molécules influence la dynamique du promoteur à différents niveau. C'est en particulier ce que nous ferons dans la section 4.4 où nous décrirons comment différents modes d'interaction des FT entre eux et avec la chromatine sont à même de reproduire ou non des observations expérimentales de diverses natures. Dans l'exemple de la figure III.4, les taux de transition ont simplement été obtenus par un algorithme ne faisant que renforcer l'activité périodique spontanée de ce système (cf annexe B.2).

2.4 Reformulation énergétique

La régulation transcriptionnelle a longtemps été approchée par des modèles thermodynamiques (Ackers *et al.*, 1982; Bintu *et al.*, 2005a, cf section1.3). De par leur définition, ceux-ci décrivent nécessairement le comportement *moyen* et *à l'équilibre* de systèmes *fermés* (c'est-à-dire ne consommant pas d'énergie extérieure). Nous montrons ici comment l'approche que nous employons constitue une généralisation de ces modèles et permet d'étudier les aspects *stochastiques* et *dynamiques* (c'est-à-dire hors d'équilibre) de systèmes potentiellement *ouverts*, comme c'est le cas des promoteurs eucaryotes où de nombreux remodelages chromatiniens sont effectués par des complexes ATP-dépendants et où les modifications covalentes des histones impliquent l'hydrolyse d'une co-enzyme.

La formulation cinétique de notre système rend possible sa description par une équation maîtresse (cf §3). Ce formalisme a été largement étudié pour représenter des systèmes biochimiques (Schnakenberg, 1976) (pour une utilisation plus récente, voir (van Kampen, 2007)) et en particulier la relation entre consommation énergétique, cycles et oscillations. Nous appliquons ici ces notions dans le contexte de la régulation transcriptionnelle et mettons en évidence leurs implications d'un point de vue biologique.

Modèle thermodynamique classique

L'approche classique empruntée par les modèles thermodynamiques consiste à définir une différence d'énergie libre de Gibbs ΔG pour l'association de chaque FT à l'ADN et pour l'interaction entre plusieurs FT (pour chaque paire, triplet, ...) (Saiz et Vilar, 2006a). Cela revient à définir une énergie G_s pour chaque état du promoteur (figure III.5a) et correspond, dans notre cas, à un vecteur à N éléments, noté $\mathbf{G}$. Cependant, $\mathbf{G}$ ne permet

de décrire que les *constantes d'équilibre* du système, c'est-à-dire le rapport entre les deux *constantes cinétiques* – aller et retour – pour chaque transition (Kaufman, 2002) :

$$G_{s'} - G_s = -\mathrm{k_B} T \log \frac{M_{s',s}}{M_{s,s'}} \tag{III.7}$$

où $\mathrm{k_B}$ est la constante de Boltzmann et T la température. L'équation (III.7) signifie que, la réaction $s \rightleftarrows s'$ est orientée vers l'état dont l'énergie libre est la plus basse (figures III.5b1 et III.5b2) et ce d'autant plus que la différence d'énergie est importante. Si les deux énergies sont égales, alors les deux taux de transition sont égaux (figure III.5b3). Ainsi, sur un profil énergétique tel que ceux représentés sur la figure III.5, le système a spontanément tendance à aller vers les états de plus basse énergie.

Par construction, cette définition contraint la matrice $\mathbf{M}$ à être dépourvue de cycles orientés[9]. En effet, pour n'importe quel chemin cyclique dans le graphe de transition, c'est-à-dire pour n'importe quelle séquence de m états notés $s_0, s_1, \ldots, s_{m-2}, s_{m-1}, s_m$ (avec $s_m = s_0$), la différence d'énergie libre le long du chemin est nécessairement nulle $\sum_{p=0}^{m-1} G_{s_{p+1}} - G_{s_p} = 0$. En employant la relation (III.7), cette propriété se reformule en $\prod_{p=0}^{m-1} M_{s_{p+1},s_p} = \prod_{p=0}^{m-1} M_{s_p,s_{p+1}}$, ce qui signifie que le produit des constantes cinétiques le long d'un cycle est le même dans les deux sens.

Introduction de la cinétique et représentation de systèmes ouverts

Pour obtenir les constantes cinétiques $M_{s',s}$ (et non uniquement les rapports $M_{s',s}/M_{s,s'}$), il est nécessaire de définir les énergies d'activation entre paires d'états. Nous notons $E_{s',s}$ l'énergie de la barrière d'activation entre les états s et s' (figure III.5), définissant la matrice $\mathbf{E}$ de dimensions $N \times N$ (considérée symétrique pour l'instant : $E_{s',s} = E_{s,s'}$) dont les éléments non-contraints sont les mêmes que ceux de $\mathbf{M}$ (figure III.3). Les constantes cinétiques sont alors obtenues par l'équation d'Arrhenius[10] :

$$E_{s',s} - G_s = -\mathrm{k_B} T \log M_{s',s} \tag{III.8}$$

On note en effet que lorsque $\mathbf{E}$ est symétrique, l'équation (III.8) implique l'équation (III.7).

L'équation (III.8) signifie que plus la barrière énergétique à franchir pour une transition donnée est élevée plus cette dernière est rare. Ainsi, des barrières d'activation différentes entre deux états modifient les temps de transition mais laissent inchangé l'équilibre de la réaction (figure III.5b3 et III.5b4).

Lors de la réaction $s \rightarrow s'$, pour passer la barrière d'activation entre les deux états, le système dans l'état s reçoit une quantité d'énergie $E_{s',s} - G_s$ provenant des fluctuations thermiques du milieu. Une fois cette barrière passée, le système libère immédiatement une quantité d'énergie thermique $E_{s,s'} - G_{s'}$ en se plaçant dans l'état s'. Dans le cas d'un système fermé ($E_{s',s} = E_{s,s'}$), le bilan énergétique correspond simplement à la différence

[9] Chemin fermé le long duquel le produit des taux de transition n'est pas le même dans les deux sens.

[10] Par simplicité, le facteur pré-exponentiel dans l'équation d'Arrhenius (multiplicatif sur $M_{s',s}$) est implicitement considéré comme une constante additive dans la définition de $E_{s',s}$.

d'énergie libre entre les deux états $G_{s'} - G_s$. Ainsi, lorsque le système sera revenu dans son état d'origine s (soit directement, soit par un autre chemin), il aura rendu autant d'énergie qu'il en aura reçu. À l'inverse, si l'on considère que le système est ouvert et reçoit de l'énergie chimique (par exemple, par l'hydrolyse d'une molécule d'ATP) au cours de la réaction $s \to s'$, alors on a $E_{s',s} < E_{s,s'}$, la différence étant simplement la quantité d'énergie chimique reçue (figure III.5c). L'énergie thermique à recevoir pour franchir la barrière d'activation dans les sens $s \to s'$ (soit $E_{s',s} - G_s$) est alors moins importante et la réaction dans ce sens se fait plus facilement. Lorsque le système sera retourné dans son état d'origine s, il aura alors rendu plus d'énergie thermique qu'il n'en aura reçu, mais il aura consommé de l'énergie chimique. Ainsi, tous les états atteints après cette réaction apparaissent comme ayant une énergie libre plus basse qu'avant cette réaction. Cela résulte en un profil énergétique apparent dans lequel un même état aura une énergie libre apparente d'autant plus basse que de l'énergie extérieure aura été consommée. Si l'on considère par exemple une séquence cyclique de transitions contenant au moins une réaction énergie-dependante (figure III.5c2), le profil apparent consiste en une descente énergétique sans fin dans laquelle il est toujours possible d'atteindre des états de plus basse énergie. Le système ne parvient alors jamais à l'équilibre. Il s'agit là d'un exemple illustrant, sur un système simple, comment un apport énergétique extérieur provoque un cycle orienté. Dans le cas général, dès que la matrice $\mathbf{E}$ n'est plus symétrique, la matrice $\mathbf{M}$ peut contenir des cycles orientés :

$$\sum_{p=0}^{m-1} E_{s_{p+1},s_p} - E_{s_p,s_{p+1}} = -\mathrm{k_B} T \log \prod_{p=0}^{m-1} \frac{M_{s_{p+1},s_p}}{M_{s_p,s_{p+1}}} \tag{III.9}$$

Ainsi, dans n'importe quel graphe d'état du promoteur, un cycle donné n'est orienté que s'il inclut des réactions qui consomment de l'énergie extérieure et il l'est d'autant plus que cette consommation énergétique cumulée est importante.

Un intérêt majeur de la description du promoteur d'un point de vue énergétique (en obtenant $\mathbf{M}$ à partir de la définition de $\mathbf{G}$ et de $\mathbf{E}$) est que l'on peut définir explicitement pour chaque transition s'il y a ou non consommation d'énergie extérieure et quelle quantité d'énergie est éventuellement consommée. Dans l'exemple de la figure III.4, les réactions de modification de la chromatine (décrites par la figure III.4a2 et représentées en rouge sur la figure III.4c) sont les seules réactions énergie-dépendantes et consomment toutes une énergie[11] ΔG_{ATP}. Toutes les autres réactions, c'est-à-dire les associations/dissociations de FT et des complexes sur le promoteur (représentées dans le graphe d'état par des arcs en niveaux de gris), sont définies comme n'impliquant pas d'apport énergétique extérieur. Ces propriétés imposent des contraintes sur les valeurs de $\mathbf{M}$ qui ne pourraient pas être prises en compte si l'on définissait cette matrice directement et qui sont susceptibles d'influencer le comportement de la structure, aussi bien localement que globalement.

Le fait de représenter le promoteur comme un système ouvert a un certain nombre d'implications sur la façon dont il peut être étudié. En effet, un système énergie-indépendant (et donc dépourvu de cycles orientés) respecte le principe d'*équilibre détaillé* : à l'état stationnaire, les concentrations des espèces chimiques sont telles que chaque réaction a lieu autant dans un sens que dans l'autre (principe de *micro-réversibilité*) et on est alors à

[11] On pourrait définir de façon plus réaliste une énergie différente pour chacune.

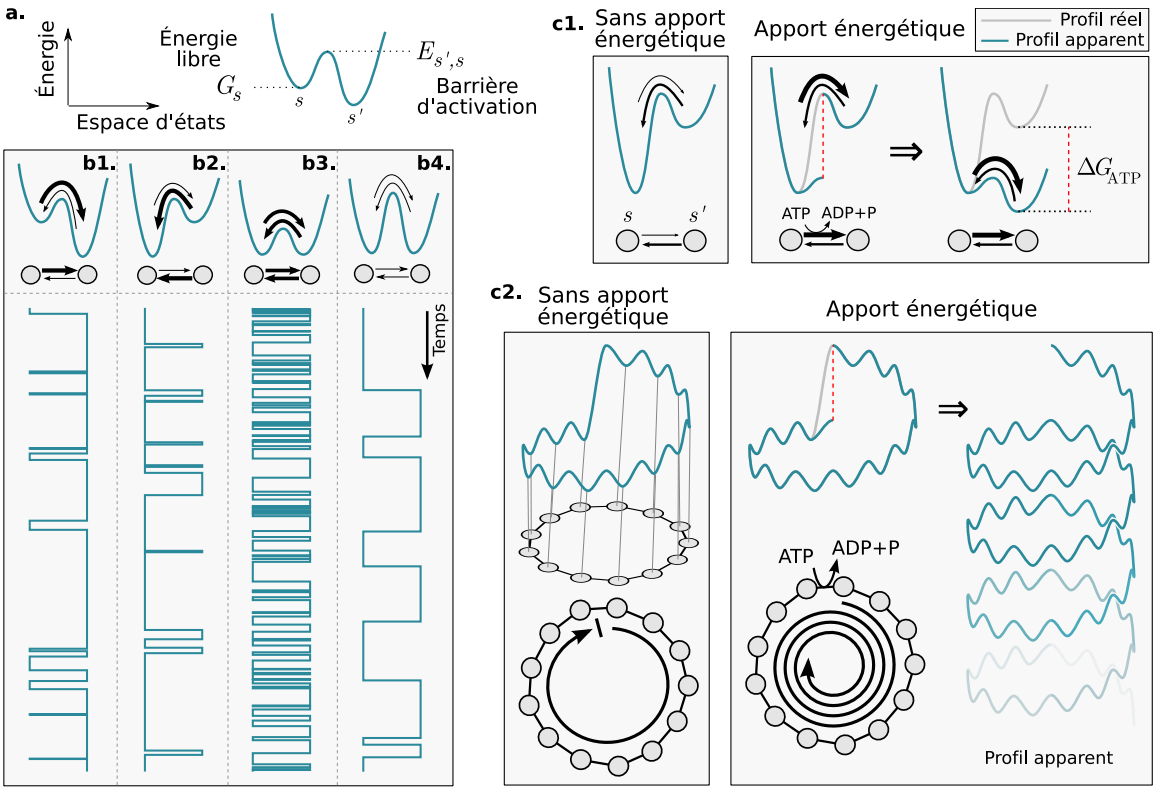


Fig. III.5: Interprétation énergétique des transitions d'état du promoteur. (a) L'ensemble des paramètres cinétiques du promoteur, à savoir les taux de transitions entre paires d'états décrits par la matrice $\mathbf{M}$, peut être défini en associant à chaque état s une énergie libre G_s et à chaque transition $s \to s'$ une énergie de barrière d'activation $E_{s',s}$, définissant alors un vecteur $\mathbf{G}$ à N éléments et une matrice $\mathbf{E}$ de dimensions $N \times N$. On peut représenter schématiquement ces paramètres par un profil énergétique dans un espace abstrait d'états du promoteur (chaque puits d'énergie correspondant à un état). La métaphore du profil énergétique est utile d'un point de vue intuitif car l'évolution stochastique du système qu'il décrit est qualitativement similaire au déplacement, sur ce profil, d'une particule soumise à la fois à une agitation spontanée (d'autant plus importante que la température du milieu est élevée) et à une attraction vers le bas due à la tendance du système à minimiser son énergie. Le passage d'un état à un autre – c'est-à-dire la sortie d'un puits – est d'autant plus facile (et donc rapide) que la barrière à passer $E_{s',s} - G_s$ est peu importante. (b) Ainsi, tel qu'illustré ici sur des systèmes fermés très simples à deux états, le promoteur tend globalement à aller vers – et à passer plus de temps dans – les états de plus basse énergie (b1,b2). Il passe de fait autant de temps dans deux états de même énergie (b3,b4). La hauteur des barrières d'activation détermine quant à elle la fréquence des transitions : plus la barrière entre deux états est basse plus les transitions sont fréquentes (b3) et inversement (b4). (c) Dans le cadre de systèmes ouverts, l'apport énergétique extérieur peut modifier significativement le profil énergétique du système et donc son comportement. (c1) Pour un système simple à deux états, si la transition $s \to s'$ vers l'état s' le moins favorable est accompagnée par exemple de l'hydrolyse d'une coenzyme (ATP, acetyl-CoA, NAD$^+$, ...), la barrière à franchir apparaît alors d'autant moins haute que l'apport énergétique extérieur ΔG est élevé. Cela provoque un *profil énergétique apparent* dans lequel tous les états atteints après une transition énergie-dépendante apparaissent plus bas que dans le profil réel. (c2) Ainsi, dans un système où l'on peut retourner dans l'état s par d'autres transitions, la réaction énergie-dépendante provoque un profil apparent dans lequel il est toujours possible d'atteindre des états de plus basse énergie, correspondant d'un point de vue cinétique à un cycle orienté dans le graphe d'état.

l'équilibre thermodynamique (Schnakenberg, 1976; van Kampen, 2007). Cette propriété, également connue comme la réversibilité d'un processus de Markov ou comme l'absence de circulation au sein du système, n'est plus valable dès lorsque le système est ouvert et consomme de l'énergie extérieure. Cela proscrit l'utilisation de nombreux outils utilisés par les approches thermodynamiques (tels que l'application d'un facteur de Boltzmann par exemple) et impose alors d'étudier le promoteur comme un système hors d'équilibre.

Dépendance à la concentration des FT

Les énergies associées aux états du promoteur $\mathbf{G}$ et aux barrières d'activation entre états $\mathbf{E}$ dépendent de la concentration des FT et des complexes (Kaufman, 2002). Elles peuvent être déduites pour des concentrations arbitraires à partir de la définition de $\mathbf{G}^0$ et $\mathbf{E}^0$ (valeurs de $\mathbf{G}$ et $\mathbf{E}$ en conditions standards[12]) :

$$G_s = G_s^0 + \mathrm{k_B}T \sum_{f \notin s} \log[f] \tag{III.10}$$

$$E_{s',s} = E_{s',s}^0 - \mathrm{k_B}T \log[s'-s] + \mathrm{k_B}T \sum_{f \notin s} \log[f] \tag{III.11}$$

où, comme précédemment, $[s'-s]$ est la concentration de $s'-s$ (c'est-à-dire le complexe qui s'associe si $\mathcal{G}_s \to \mathcal{G}_{s'}$ est une association et $\varnothing$ sinon) et par convention $[\varnothing] = 1$. En effet, la traduction de ces équations en formulation cinétique (en appliquant l'équation (III.8)) aboutit bien à $M_{s',s} = [s'-s]M_{s',s}^0$, la relation de concentration-dépendance (III.6).

L'interprétation intuitive de l'équation (III.10) est que l'énergie libre G_s associée aux états pour lesquels un FT donné est absent du promoteur augmentent avec la concentration de ce facteur, rendant donc ces états moins stables. L'équation (III.11) signifie que, pour un système de réactions hors promoteur $\mathbf{\Theta}$ donné, les énergies d'activation entre deux états (dans les deux sens, $E_{s',s}$ et $E_{s,s'}$) varient comme l'état le plus associé[13]. Ainsi la barrière énergétique à franchir pour une dissociation est indépendante des concentrations et, pour une association, diminue avec la concentration du complexe qui s'associe.

De plus, on remarque que des réactions énergie-dépendantes ayant lieu hors du promoteur peuvent transmettre leur énergie au promoteur. En effet, si le système $\mathbf{\Theta}$ contient des réactions qui maintiennent activement (c'est-à-dire avec consommation d'énergie) les FT et complexes dans des concentrations différentes de celles définies[14] par $\mathbf{E}^0$, alors cette énergie est transmise au promoteur. C'est par exemple le cas des modifications des queues

[12]Les conditions standards sont celles où toutes les espèces chimiques ont une concentration unitaire.

[13]Si $\mathcal{G}_s \to \mathcal{G}_{s'}$ est une dissociation, le terme $-\mathrm{k_B}T\log[s'-s]$ est nul et l'énergie thermique $E_{s',s} - G_s$ à recevoir pour que la réaction ait lieu vaut toujours $E_{s',s}^0 - G_s^0$ et est donc indépendante des changements de concentration. Si $\mathcal{G}_s \to \mathcal{G}_{s'}$ est une association, les variations du terme $\mathrm{k_B}T\sum_{f\notin s}\log[f]$ sont contre-balancées par le terme $-\mathrm{k_B}T\log[s'-s]$. En effet, le système de réaction $\mathbf{\Theta}$ étant à l'état stationnaire, la concentration du complexe qui s'associe $[s'-s]$ est proportionnelle au produit des concentrations libres des FT qui le composent $\prod_{f\in(s'-s)}[f]$ (par exemple $[AB] = K_1[A][B]$).

[14]L'énergie chimique reçue au cours de la réaction $\mathcal{G}_s \to \mathcal{G}_{s'}$ vaut $E_{s,s'} - E_{s',s} = E_{s,s'}^0 - E_{s',s}^0 + \mathrm{k_B}T\log\frac{[s'-s]}{\prod_{f\in(s'-s)}[f]}$ et ne peut être nulle que si les constantes de $\mathbf{\Theta}$ sont définies en fonction de $\mathbf{E}^0$.

d'histones : une acétyltransférase utilise un donneur qui s'associe avec un groupe acétyl et se dissocie sans, mais dont la forme acétylée est activement synthétisée hors du promoteur.

2.5 Un cadre simplifié : association/dissociation de FT seuls

Il est possible de simplifier significativement la définition du modèle en supposant que les FT ne peuvent s'associer et se dissocier que seuls et non sous forme de complexes, ce qui permet également d'ignorer les interactions des FT hors du promoteur. Dans ce cadre, l'espace paramétrique est alors fortement réduit : tous les facteurs se comportent comme des FC dans la figure III.3b et le système de réactions hors promoteur est nécessairement vide : $\mathbf{\Theta} = \{\varnothing\}$.

Ces suppositions rendent en particulier beaucoup plus simple la description de l'effet de concentration des FT. Tous les FT étant libres, leurs concentrations $[f]$ sont directement des paramètres du système et sont indépendantes les unes des autres. Tous les équations présentées précédemment restent valides mais ne sont définies que pour $|s \ominus s'| = 1$ (et non $s' \supset s$ ou $s' \subset s$), de sorte que $[s' - s]$ est simplement la concentration $[f]$ du FT qui s'associe ou se dissocie (et non celle d'un complexe). L'équation (III.11) peut alors être simplifiée en $E_{s',s} = E^0_{s',s} + \mathrm{k_B}T \sum_{f \notin s \cup s'} \log[f]$.

Ignorant la possibilité pour des molécules de s'associer ou de se dissocier ensemble sur le promoteur, les travaux de Sánchez et Kondev (2008) utilise implicitement cette version simplifiée. Notons par ailleurs que les auteurs n'emploient qu'une définition cinétique et n'interprètent pas leur modèle en termes énergétiques (tel que décrit dans la section 2.4). Cependant, bien que ce cadre simplifié puisse être utile dans certaines cas, il rend le modèle beaucoup moins expressif, empêchant en particulier d'étudier le mode d'assemblage des complexes : sur le promoteur, hors du promoteur ou tout intermédiaire entre ces deux situations. Ce point est pourtant un aspect essentiel de la dynamique du promoteur, comme le montrent des études récentes (Gorski *et al.*, 2008; Darzacq et Singer, 2008; Degenhardt *et al.*, 2009; Hager *et al.*, 2009) et comme nous l'étudierons à la section 4.4.

*
* *

Le modèle que nous avons défini dans cette section 2 constitue un cadre générique permettant de représenter des promoteurs très différents : aussi bien des systèmes simples et idéaux que des systèmes arbitrairement complexes. De plus, de nombreuses caractéristiques essentielles des promoteurs peuvent être décrites de façon précise et quantitative. Ce modèle constitue donc un cadre théorique pouvant servir à l'étude de nombreux aspects de la dynamique stochastique du promoteur et de son influence sur l'expression génique. Nous présentons dans la section suivante les dérivations que l'on peut faire d'un point de vue général quant au comportement dynamique de ce système, en reproduisant en particulier à la fois des mesures employées pour étudier la dynamique du promoteur et pour étudier la stochasticité de l'expression génique.

3 Dérivations générales

Bien qu'il puisse représenter des systèmes très différents, le modèle de promoteur décrit par la réaction (III.1) a une définition particulièrement simple et homogène. Il s'agit d'un ensemble de réactions unimoléculaires de même nature et la complexité de ce qui peut être représenté ne provient que de la paramétrisation par la matrice $\mathbf{M}$. L'intérêt d'une telle définition est que l'on peut, avant même d'instancier le modèle (et donc la matrice $\mathbf{M}$), dériver de manière générique un grand nombre d'indicateurs et en comprendre alors la structure en fonction des paramètres. Cela permet également de déduire des propriétés générales, par exemple les conditions nécessaires sur $\mathbf{M}$ pour obtenir un certain comportement.

Nous commençons ici par décrire la dynamique d'un tel système en montrant en particulier qu'elle peut être décomposée en composantes élémentaires simples, appelées *modes*. Nous en déduisons alors différents indicateurs à la fois théoriques et reproduisant des mesures expérimentales (et dont l'ordre de présentation suit par simplicité la logique mathématique des dérivations). Nous nous intéressons ensuite au comportement de systèmes idéalisés (en particulier du cycle homogène, orienté et isolé de transitions entre états) afin de mettre à jour certaines relations au sein du système. Nous montrons enfin comment la dynamique stochastique du promoteur se propage aux niveaux d'ARN et de protéines et rivalise alors avec les autres sources de stochasticité gène-intrinsèques.

3.1 Dynamique du promoteur

Les changements d'état du promoteur sont décrits par l'ensemble des réactions (III.1) et correspondent à un processus de Markov à états discrets et à temps continu. Jusqu'ici, nous n'avons pas défini la valeur des éléments diagonaux de la matrice $\mathbf{M}$ (en effet, la réaction (III.1) n'est définie que pour $s \neq s'$). Si l'on fixe ces éléments diagonaux de façon à ce que la somme sur chaque colonne soit nulle $\sum_{s'} M_{s',s} = 0, \forall s$, on peut alors décrire très simplement l'évolution de l'état du promoteur par l'équation maîtresse :

$$\frac{\mathrm{d}\phi(t)}{\mathrm{d}t} = \mathbf{M}\phi(t) \tag{III.12}$$

où $\phi(t)$ est le vecteur dont chacun des N éléments décrit la probabilité que le promoteur soit, à l'intant t, dans chacun des N états. Cette équation différentielle signifie que la probabilité que le promoteur soit dans un état donné augmente d'autant plus qu'il est probable d'arriver dans cet état (élément non-diagonaux de $\mathbf{M}$ dans le produit $\mathbf{M}\phi(t)$) et diminue d'autant plus qu'il est probable de le quitter (élément diagonaux de $\mathbf{M}$).

Ce formalisme mathématique est couramment employé pour représenter la dynamique stochastique d'une entité unique (Schnakenberg, 1976; van Kampen, 2007) et est identique, d'un point de vue mathématique, au formalisme décrivant l'évolution déterministe par des lois de vitesse d'un système de réactions de premier ordre (chaque éléments de $\phi(t)$ représentant alors la concentration d'une espèce chimique et $M_{s',s}$ le taux auquel l'espèce s est transformée en l'espèce s'), par exemple dans le cas des réseaux métaboliques (Hearon, 1953; Summers et Scott, 1988).

Dans sa formulation la plus simple, l'intégration de l'équation maîtresse (III.12) donne l'équation de Chapman-Kolmogorov suivante :

$$\phi(t+\tau) = e^{\tau\mathbf{M}}\phi(t) \tag{III.13}$$

Elle permet de déterminer comment une distribution de probabilité connue à un instant t aura évolué après un délai τ. La nouvelle distribution $\phi(t+\tau)$ s'obtient par le produit matriciel du vecteur $\phi(t)$ avec l'exponentielle matricielle $e^{\tau\mathbf{M}}$. Ainsi, la matrice $e^{\tau\mathbf{M}}$ en fonction de τ est centrale dans notre étude car elle décrit entièrement la dynamique du promoteur. L'élément sur la s-ième ligne et la s'-ième colonne de $e^{\tau\mathbf{M}}$ est $\mathrm{P}(s', t+\tau|s, t)$, la probabilité d'être passé de $\mathcal{G}_s$ à $\mathcal{G}_{s'}$ après un délai τ. Chaque colonne représente, pour un état de départ donné, la répartition du système sur les différents états. Pour $\tau = 0$, l'exponentielle matricielle $e^{\tau\mathbf{M}}$ est simplement la matrice identité, indiquant qu'après un délai nul, le promoteur n'a pas changé d'état. Lorsque τ tend vers l'infini, toutes colonnes tendent vers le même vecteur, que nous notons Λ_0, décrivant la distribution stationnaire du système[15] (distribution d'une population de promoteurs identiques et indépendants les uns des autres).

La figure III.6 montre l'évolution de $e^{\tau\mathbf{M}}$ pour l'exemple de promoteur eucaryote introduit précédemment (figure III.4) et illustre en particulier les différents niveaux d'observation qu'elle permet. Il est cependant important de garder à l'esprit qu'il s'agit ici d'un exemple particulier et que l'exponentielle matricielle peut témoigner de comportements très différents en fonction du système auquel elle s'applique.

De nombreux indicateurs peuvent être déduits à partir de l'exponentielle matricielle $e^{\tau\mathbf{M}}$ de $\mathbf{M}$ ou d'une matrice analogue (par exemple une sous-matrice). En effet, toute condition sur l'état du promoteur peut être représentée par un vecteur dont le s-ième élément vaut 1 si l'état s vérifie la condition en question et 0 sinon. Ainsi, pour une distribution ϕ_0 donnée, la probabilité qu'une condition c soit vérifiée est simplement $c^{\mathrm{T}}\,\phi_0$ (où ^{T} désigne la transposition) et la probabilité qu'elle le soit après un délai τ est $c^{\mathrm{T}}\,e^{\tau\mathbf{M}}\,\phi_0$. De plus, la probabilité que c soit vérifiée un délai τ après qu'une autre condition c' l'ait été est $c^{\mathrm{T}}\,e^{\tau\mathbf{M}}\,\mathbf{D}_{c'}\Lambda_0$ (où $\mathbf{D}_{c'}$ est la représentation du vecteur c' sous la forme d'une matrice diagonale et Λ_0, la distribution stationnaire du système). Dans le cas général, l'évolution de la probabilité qu'une condition soit satisfaite après une perturbation ou sachant qu'une autre condition l'a été peut se mettre sous la forme

$$v_{\mathrm{post}}{}^{\mathrm{T}}\,e^{\tau\mathbf{M}}\,v_{\mathrm{pre}} \tag{III.14}$$

Cela revient donc à calculer une projection de l'exponentielle matricielle, une somme partielle de ses valeurs qui dépend de l'indicateur considéré.

[15] Sous une conditions d'ergodicité que nous clarifierons dans la section 3.2 mais qui est valable pour tout promoteur physiquement réaliste.

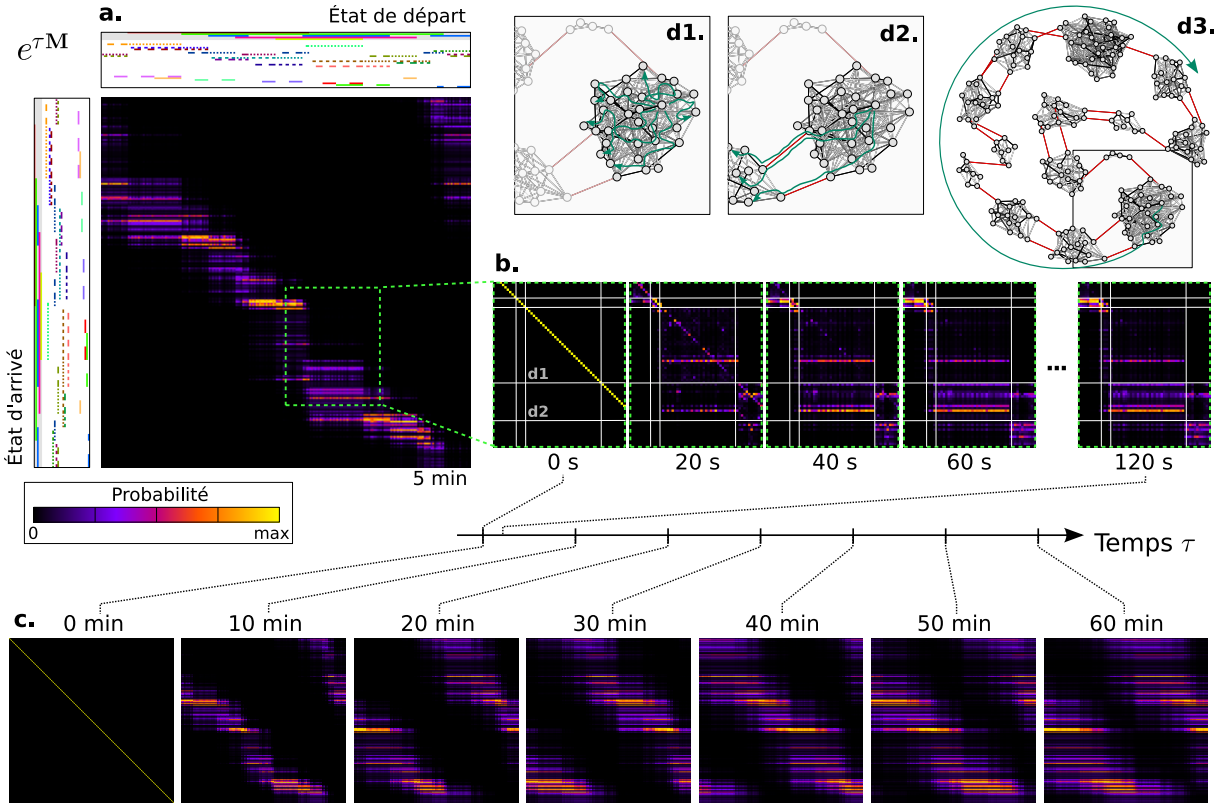


Fig. III.6: Exponentielle matricielle $e^{\tau \mathbf{M}}$ pour l'exemple du promoteur eucaryote. Cette figure montre l'évolution des valeurs de $e^{\tau \mathbf{M}}$ en fonction du temps τ pour l'exemple introduit à la figure III.4. Le terme $e^{\tau \mathbf{M}}$ correspond à l'*exponentielle matricielle* du produit du réel τ par la matrice $\mathbf{M}$. Il s'agit simplement d'une matrice de mêmes dimensions que $\mathbf{M}$ et dont les valeurs dépendent de τ. La figure (a) représente cette matrice pour τ=5 min. Comme pour $\mathbf{M}$, les différentes lignes et colonnes de $e^{\tau \mathbf{M}}$ correspondent chacune à un état du promoteur. Cependant, alors que l'élément $M_{s',s}$ de la matrice $\mathbf{M}$ représente le taux de transition $s \rightarrow s'$ (cf figure III.4d), le même élément dans la matrice $e^{\tau \mathbf{M}}$ décrit la probabilité que, partant de s à un instant donné, le système se retrouve dans l'états s' au bout d'un temps τ, soit $\mathrm{P}(s', t+\tau | s, t)$. Ainsi, pour un état de départ s donné, la s-ième colonne de $e^{\tau \mathbf{M}}$ décrit simplement la distribution du système parmi tous les états, en fonction du délai τ. En effet, la somme sur chaque colonne vaut toujours 1 et, pour τ=0, $e^{\tau \mathbf{M}}$ est la matrice identité. La figure (b) représente l'évolution de l'exponentielle matricielle pour des temps courts. Seule une partie de $e^{\tau \mathbf{M}}$ est représentée ici (cf cadre pointillé sur (a)) afin de montrer l'évolution des valeurs dans différentes régions définies par les états chromatiniens (rappelons que l'ordre de représentation des états est choisi ici tel qu'un même état chromatinien constitue un bloc contigu). Cette échelle d'observation de $e^{\tau \mathbf{M}}$ décrit de façon précise la dynamique locale du promoteur : l'association/dissociation des FT sur la chromatine au sein d'un même état chromatinien (région *d1* dans (b) et schéma (d1)) et la transition vers un autre état chromatinien (région *d2* dans (b) et schéma (d2)). Dans l'exemple présenté ici, il est peu probable d'avoir quitté l'état chromatinien au bout de 20 s, mais fort probable d'avoir atteint le suivant après 120 s. (c) L'exponentielle matricielle $e^{\tau \mathbf{M}}$, observée sur des temps longs, décrit la dynamique lente du promoteur, si celle-ci existe. Dans cet exemple, la définition du système (cf figure III.4) provoque un comportement périodique lent de modification de la chromatine (d3). On observe ici de façon qualitative, qu'en partant d'un état donné, le système quitte rapidement son état chromatinien et a peu de chance d'y revenir avant 40 min (valeurs élevées proches de la diagonale). De plus, la largeur du nuage de valeurs témoigne de l'imprécision dans ce délai, indiquant la probabilité pour un promoteur d'être en avance ou en retard. Ainsi, l'exponentielle matricielle décrit entièrement la dynamique du promoteur et peut alors servir de base pour dériver de nombreux indicateurs.

3.2 Décomposition en modes propres

Le calcul d'une exponentielle matricielle est un problème connu pour poser un certain nombre de difficultés computationnelles (Moler et Van Loan, 2003). Un grand nombre de méthodes ont été proposées, chacune étant plus ou moins adaptée à différentes situations. Une méthode donnant de relativement bons résultats dans le cas général et employée par défaut dans la plupart des logiciels de calcul numérique[16] est l'utilisation d'un approximant de Padé[17]. Cependant, la plupart des méthodes ne permettent d'évaluer $e^{\tau\mathbf{M}}$ que pour un τ donné et doivent donc être répétées si l'on souhaite connaître la dépendance à τ (c'est par exemple le cas de l'approximant de Padé). À l'inverse, d'autres techniques permettent d'exprimer $e^{\tau\mathbf{M}}$ comme une fonction de τ. Une méthode particulièrement intéressante dans notre cas se base sur la décomposition de la matrice $\mathbf{M}$ en valeurs propres et vecteurs propres, une opération très classique en algèbre linéaire. Elle permet d'exprimer $e^{\tau\mathbf{M}}$, non seulement comme une fonction de τ, mais surtout comme une somme de composantes élémentaires simples, appelées *modes*. Appliquée à notre modèle, cette méthode de décomposition permet alors de comprendre la structure de la dynamique du promoteur et des différents indicateurs.

L'ensemble des valeurs propres de la matrice $\mathbf{M}$, notées chacune λ_i, est appelé le *spectre* de $\mathbf{M}$. Rappelons qu'une valeur propre est un nombre complexe $\lambda_i \in \mathbb{C}$ et peut donc être vue comme un point sur un plan 2D (le plan complexe) dont les coordonnées sont respectivement la partie réelle $\mathrm{Re}(\lambda_i)$ et la partie imaginaire $\mathrm{Im}(\lambda_i)$ de λ_i. Ainsi, le spectre de $\mathbf{M}$ forme un ensemble de points sur le plan complexe dont la disposition dépend du contenu de cette matrice et donc du système représenté. La figure III.7a1 décrit le spectre de $\mathbf{M}$ pour l'exemple du promoteur eucaryote évoqué précédemment.

Il est à noter que, en théorie des graphes, la matrice $-\mathbf{M}$ est appélée le *laplacien* du graphe[18] et que la *théorie spectrale des graphes* (Cvetković *et al.*, 1979; Chung, 1997) consiste en l'étude de la structure des graphes par le spectre de leur laplacien. Bien que cette discipline s'intéresse essentiellement à des graphes non-pondérés et en général non-orientés, de nombreux résultats peuvent être exploités dans le cadre de nos travaux.

Nous notons $\boldsymbol{\lambda} = [\lambda_0, \lambda_1, \ldots, \lambda_{N-1}]$ le vecteur contenant les N valeurs propres λ_i de $\mathbf{M}$ (il y a autant de valeurs propres que d'états du promoteur), $\boldsymbol{\Lambda} = [\Lambda_0|\Lambda_1|\ldots|\Lambda_{N-1}]$ la matrice $N \times N$ dont chaque colonne Λ_i est un vecteur propre à droite (ainsi, $\Lambda_{s,i}$ est le s-ième élément de Λ_i) et $\boldsymbol{\Lambda}^{-1}$, la matrice inverse de $\boldsymbol{\Lambda}$, dont chaque ligne Λ_i^{-1} est un vecteur propre à gauche. Dans un tel système (Schnakenberg, 1976), toutes les valeurs propres ont une partie réelle négative, c'est pourquoi il est plus simple de considérer $-\lambda_i$, les valeurs propres de $-\mathbf{M}$. De plus, celles qui sont non-réelles ($\mathrm{Im}(\lambda_i) \neq 0$) viennent nécessairement par paires avec leur conjuguée $\lambda_i = \overline{\lambda_{i'}}$. Autrement dit, le spectre de $-\mathbf{M}$ forme toujours un ensemble symétrique par rapport à l'axe des réels sur le demi-plan complexes $\mathrm{Re}(-\lambda_i) \geq 0$ (figure III.7a1).

[16] Tels que Matlab ou SciPy par exemple.

[17] Analogue à l'approximation par un développement limité de Taylor, mais par une fonction rationnelle et non un polynôme.

[18] La matrice d'adjacence soustraite à la matrice diagonale des degrés (sortants, dans le cas orienté).

À partir de cette décomposition, on a $\mathbf{M} = \mathbf{\Lambda}\mathbf{D}_{\boldsymbol{\lambda}}\mathbf{\Lambda}^{-1}$ (où $\mathbf{D}_{\boldsymbol{\lambda}}$ est la représentation du vecteur $\boldsymbol{\lambda}$ sous forme d'une matrice diagonale). L'exponentielle matricielle peut alors se réécrire

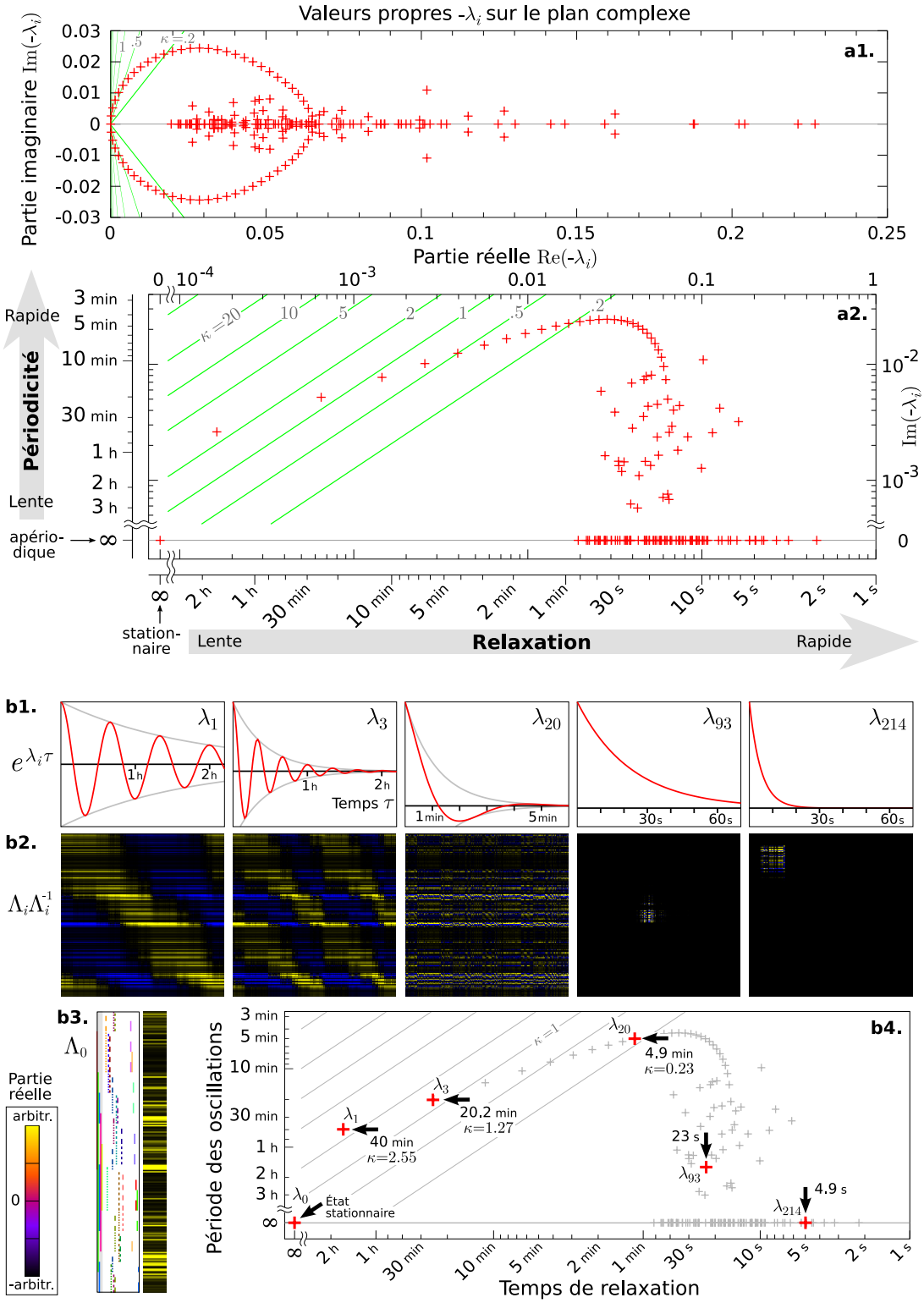
$$e^{\tau\mathbf{M}} = \mathbf{\Lambda} e^{\tau\mathbf{D}_{\boldsymbol{\lambda}}} \mathbf{\Lambda}^{-1} \tag{III.15}$$

$$= \sum_i e^{\lambda_i \tau}\, \Lambda_i \Lambda_i^{-1} \tag{III.16}$$

L'exponentielle matricielle $e^{\tau\mathbf{M}}$ peut donc s'exprimer comme une somme dont chacun des N termes est le produit de $e^{\lambda_i\tau}$ (une fonction de τ ; figure III.7b1) par $\Lambda_i\Lambda_i^{-1}$ (une matrice indépendante de τ ; figure III.7b2). Ainsi, tous les indicateurs obtenus par projection de $e^{\tau\mathbf{M}}$ (cf équation (III.14)) ne seront que des sommes des mêmes fonctions $e^{\lambda_i\tau}$ pondérées différemment par la projection des matrices $\Lambda_i\Lambda_i^{-1}$.

Le spectre de la matrice $\mathbf{M}$ détermine la « bibliothèque » de fonctions que l'on va retrouver dans ces différents indicateurs puisque la position de chaque point sur le plan complexe (ou de chaque paire de points symétriques par rapport à l'axe des réels) détermine la

Fig. III.7: Décomposition modale pour l'exemple de promoteur eucaryote. Cette figure illustre le principe de décomposition en modes propres de l'exponentielle matricielle sur l'exemple présenté précédemment (figure III.6). (a1) Les valeurs propres $-\lambda_i$ de la matrice $-\mathbf{M}$ (figure III.4d) forment, sur le plan complexe, un ensemble de N points symétrique par rapport à l'axe des réels et contenu sur le demi-plan $\mathrm{Re}(-\lambda_i) \geq 0$. Chacun de ces points (ou paires de points symétriques) correspond à une composante élémentaire simple de l'activité du promoteur, appelée *mode*. Comme illustré en détails par la figure III.8, la position de chaque point sur le plan détermine les caractéristiques temporelles du mode (temps de relaxation, caractère périodique ou non, période d'oscillation, cohérence, ...). (a2) La représentation en échelle logarithmique des deux axes du plan et leur interprétation en termes temporels met en évidence la dynamique du système à des échelles très variées. On observe en particulier de nombreuses relaxations apériodiques rapides de l'ordre de quelques secondes, ainsi que des périodicités lentes et fortement cohérentes de plusieurs dizaines de minutes (les droites passant par l'origine sur (a1) correspondent aux droites parallèles de pente 1 sur (a2) et indiquent la cohérence κ des différents modes ; cf figure III.8). (b) L'observation des fonctions $e^{\lambda_i\tau}$ (b1) et des matrices $\Lambda_i\Lambda_i^{-1}$ (b2) associées à certaines valeurs propres (b4) permet de comprendre, sur cet exemple de promoteur eucaryote, la contribution des différents modes dans la décomposition de l'exponentielle matricielle $e^{\tau\mathbf{M}}$ (équation (III.16)). La distribution stationnaire (vers laquelle le système tend lorsque $\tau \to \infty$) est simplement donnée par Λ_0 (b3), le vecteur propre associé à la valeur propre nulle $\lambda_0 = 0$. Certaines valeurs propres forment, sur le plan complexe, un ensemble proche d'une ellipse (provenant, comme nous le verrons par la suite, du caractère cyclique du graphe de transition), dans lequel se trouvent les modes les plus lents et les plus cohérents. Les matrices $\Lambda_i\Lambda_i^{-1}$ qui leur sont associées montrent que ces modes influencent l'ensemble des valeurs de l'exponentielle matricielle $e^{\tau\mathbf{M}}$ et décrivent donc un mécanisme global dans la dynamique du promoteur. À l'inverse, les valeurs propres qui ne font pas partie de cet ensemble mais se trouvent sur (ou proche de) l'axe des réels décrivent des dynamiques (quasi-)apériodiques rapides (bien que λ_{93} ne soit pas réelle, $\kappa(\lambda_{93})$ est tellement faible que $e^{\lambda_{93}\tau}$ est indicernable d'une exponentielle réelle). La répartition des valeurs au sein des matrices $\Lambda_i\Lambda_i^{-1}$ montre que ces modes décrivent des phénomènes beaucoup plus locaux. Par exemple, les modes associés aux valeurs propres λ_{214} et λ_{93} influencent respectivement une région de $e^{\tau\mathbf{M}}$ du type de *d1* et de *d2* sur la figure III.6b, décrivant donc respectivement les transitions internes à un état chromatinien (association/dissocaition de FT) et les transitions d'un état chromatinien à un autre (modification conformationnelle ou covalente). De nombreux indicateurs peuvent être exprimés



(Fig. III.7, cont. : cf page précédente)
sous la forme de la projection d'une exponentielle matricielle (sous la forme de l'équation (III.14)) et correspondent donc à la somme des fonctions $e^{\lambda_i \tau}$ pondérées par la projection des matrices $\Lambda_i \Lambda_i^{-1}$. Ainsi, pour un indicateur donné, c'est la relation entre cette projection et la répartition des valeurs au sein des matrices $\Lambda_i \Lambda_i^{-1}$ qui détermine la contribution de chaque mode.

forme de la fonction $e^{\lambda_i \tau}$ qui lui est associée (figure III.8). Cette fonction[19] de τ est une sinusoïde dont l'amplitude décroît exponentiellement vers 0. Sa fréquence d'oscillation est proportionnelle à la partie imaginaire de $-\lambda_i$ et vaut $|\mathrm{Im}(-\lambda_i)/2\pi|$, sa vitesse de relaxation (c'est-à-dire la vitesse de décroissance exponentielle de l'enveloppe) est déterminée par la partie réelle $\mathrm{Re}(-\lambda_i)$ de la valeur propre $-\lambda_i$ considérée (figure III.8a). Le temps de relaxation (inverse de la vitesse de relaxation) décrit la mémoire du système relativement au mode considéré. Il correspond au temps que deux promoteurs, partant dans un même état à un instant donné, vont mettre pour être significativement désynchronisés (qu'il s'agisse d'un mode périodique ou non). Il est important de noter que cela ne signifie pas que les deux promoteurs ont une dynamique parfaitement identique pendant cette durée, mais bien que leurs comportements divergent l'un de l'autre petit à petit et deviennent significativement différents au-delà de ce temps de relaxation.

Ainsi, sur le spectre de $\mathbf{M}$, les axes peuvent être interprétés quantitativement en termes dynamiques : pour chaque valeur propre, l'ordonnée décrit la période d'oscillation du mode et l'abscisse, son temps de relaxation. Pour l'exemple du promoteur eucaryote (figure III.7a2), le spectre de $\mathbf{M}$ met alors en évidence une dynamique comportant des composantes à des échelles temporelles très différentes : de nombreuses relaxations apériodiques rapides (de l'ordre de quelques secondes ou dizaines de seconde) et des oscillations lentes (avec une période pouvant atteindre 40 min). Si l'on observe les fonctions $e^{\lambda_i \tau}$ et les matrices $\Lambda_i \Lambda_i^{-1}$ associées à certaines valeurs propres (figure III.7b), on remarque que les modes périodiques lents décrivent des phénomènes très globaux impliquant tous les états du promoteur (λ_1, λ_3 et λ_{20}), alors que les modes rapides et (quasi-)apériodiques décrivent des phénomènes beaucoup plus localisés, tels que l'association/dissociation de FT sans modification de la chromatine (λ_{214}) ou, au contraire, conduisant à une modification de la chromatine (λ_{93}). Ce sont ces différentes dynamiques que l'on observe sur l'exponentielle matricielle $e^{\tau \mathbf{M}}$ pour des délais respectivement longs et courts (figures III.6c et III.6b).

Comme l'illustre la figure III.8, on peut caractériser les comportements périodiques par un *facteur de cohérence*, noté κ, et défini comme le rapport entre le temps de relaxation et la période des oscillations

$$\kappa(\lambda_i) = \left| \frac{\mathrm{Im}(\lambda_i)}{2\pi \mathrm{Re}(\lambda_i)} \right| \tag{III.17}$$

Il s'agit simplement du nombre d'oscillations du mode considéré pendant son temps de relaxation, décrivant alors à quel point le processus périodique sous-jacent qu'il représente est régulier. On notera par la suite

$$\kappa = \max_i \big[\kappa(\lambda_i)\big] \tag{III.18}$$

la cohérence maximale parmi tous les modes (dans la figure III.7, elle vaut par exemple $\kappa = \kappa(\lambda_1) \simeq 2.55$). Cette mesure, sans dimension et indépendante de la fréquence et de l'amplitude des oscillations, caractérise à quel point ces dernières s'amortissent plus

[19] Notons que l'on ne considère ici que la partie réelle de $e^{\lambda_i \tau}$ car, dans le cas où $\lambda_i \in \mathbb{R}$, sa partie imaginaire est nulle et, pour $\lambda_i \notin \mathbb{R}$, les deux fonctions $e^{\lambda_i \tau}$ et les deux matrices $\Lambda_i \Lambda_i^{-1}$ associées aux deux valeurs propres conjugées sont également conjuguées (leurs parties imaginaires s'annulent donc toujours dans l'équation (III.16)).

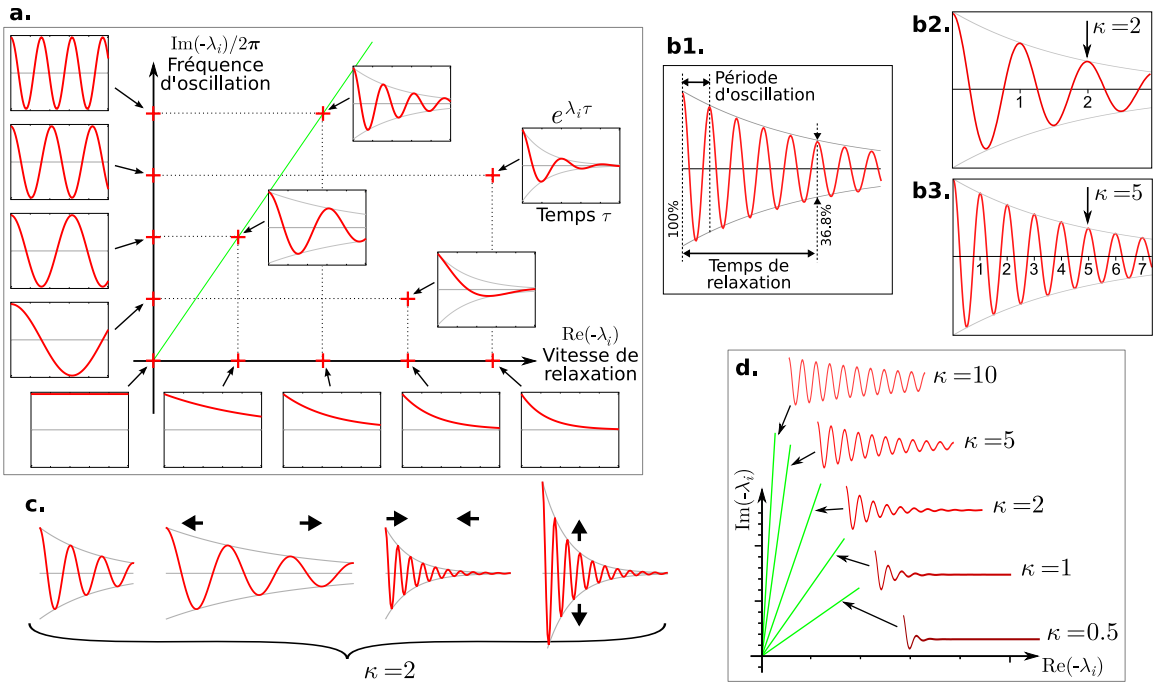


Fig. III.8: Allure des fonctions $e^{\lambda_i \tau}$ et facteur de cohérence κ. (a) Cette figure illustre comment la position d'une valeur propre $-\lambda_i$ sur le plan complexe détermine la forme de la fonction $e^{\lambda_i \tau}$ qui lui est associée[19]. Une valeur propre placée sur l'axe des ordonnées ou sur l'axe des abscisses correspond respectivement à une sinusoïde et à une décroissance exponentielle vers 0. Dans ces deux cas, la position sur l'axe détermine respectivement la fréquence des oscillations ou la vitesse de la décroissance exponentielle (appelée vitesse de relaxation). Elles valent respectivement $|\mathrm{Im}(-\lambda_i)/2\pi|$ et $\mathrm{Re}(-\lambda_i)$ et sont donc d'autant plus lentes que $-\lambda_i$ est proche de l'origine. Pour toute autre valeur de $-\lambda_i$, la fonction $e^{\lambda_i \tau}$ est simplement le produit de la sinusoïde et de la décroissance exponentielle associées respectivement à la partie imaginaire et à la partie réelle de $-\lambda_i$, décrivant alors des oscillations amorties. (b1) On peut caractériser l'allure d'une fonction $e^{\lambda_i \tau}$ par un *facteur de cohérence*, noté $\kappa(\lambda_i)$ et défini comme le rapport entre le temps de relaxation (correspondant graphiquement au délai pour que l'enveloppe exponentielle ait réduit d'un facteur $e^{-1} \simeq 0.368$) et la période d'oscillation. (b2,b3) Ce facteur décrit comment les oscillations perdurent comparativement à leur période et, comme nous le verrons plus en détail par la suite, il caractérise à quel point le processus périodique stochastique sous-jacent est *cohérent*, c'est-à-dire de période peu variable. (c) Ce facteur est insensible à des modifications d'échelles de la courbe et une même valeur peut donc correspondre à des situations très différentes. (d) L'équation (III.17) indique que la cohérence d'un mode n'est déterminée que par la pente de la droite reliant sa valeur propre à l'origine du plan complexe (à un facteur 2π près). Ainsi, deux fonctions $e^{\lambda_i \tau}$ associées à des valeurs propres placées sur une même droite passant par l'origine ont alors la même cohérence (ligne verte sur (a)). Le mode le plus cohérent dans un spectre est alors simplement celui associé à la valeur propre pour laquelle une telle droite a la plus grande pente. (NB : échelles différentes pour les deux axes sur (d))

ou moins rapidement[20]. Comme nous le verrons tout au long de ce chapitre, le facteur

[20]Bien qu'il n'y ait pas une équivalence directe entre le *facteur de cohérence* κ décrit ici et le *facteur d'amortissement* ζ employé pour qualifier un oscillateur harmonique amorti (OHA) (Tipler et Mosca, 1997), on peut remarquer une relation forte entre ces deux quantités. Notons tout d'abord que l'équation

de cohérence est contraint à différents niveaux par les paramètres du système et peut être mesuré sur des données réelles. Cela permet alors de faire un lien entre théorie et expérimentation avec, en particulier, certaines déductions quantitatives sur le système biologique observé.

La décomposition en modes propres permet également de déterminer très simplement la distribution stationnaire du système, c'est-à-dire la distribution de probabilité de trouver le promoteur dans chacun des différents états au sein d'une population désynchronisée ou après avoir attendu un temps très long. Par construction, la matrice $\mathbf{M}$ a toujours au moins une valeur propre nulle, notée $\lambda_0 = 0$, et son vecteur propre associé Λ_0 (considéré normalisé $\sum_s \Lambda_{s,0} = 1$) décrit cette distribution stationnaire (figure III.7b3). Lorsqu'il existe un sous-ensemble d'états accessibles depuis n'importe quel état, alors le système est ergodique : il n'y a qu'une valeur propre nulle et toutes les colonnes de $e^{\tau\mathbf{M}}$ tendent vers le même vecteur Λ_0. Dans le cas contraire, physiquement peu réaliste[21], on a alors autant de valeurs propres nulles qu'on a de tels sous-ensembles et les vecteurs propres associés décrivent les différents états stationnaires.

3.3 Dérivation d'indicateurs : ChIP, FRAP, corrélations croisées, spectres de puissance et distributions de durées

La description générale de la dynamique du promoteur que représente l'exponentielle matricielle peut servir de base pour dériver de nombreux indicateurs, aussi bien théoriques que reproduisant des mesures expérimentales. De plus, la décomposition en modes propres se retrouve dans chacun de ces indicateurs, décrivant alors leur structure et permettant d'y reconnaître certaines caractéristiques (telle que la cohérence des oscillations) et de comprendre l'influence de certains paramètres.

Cinétiques de ChIP

Une expérience de cinétique de ChIP consiste à observer, au cours du temps et sur une population de promoteurs indépendants et préalablement synchronisés, la proportion de ceux qui sont occupés par une certaine molécule ou qui présentent une certaine modification de la chromatine (cf Box 2, page 52). Il s'agit donc de suivre au cours du temps τ la probabilité de présence d'un certain facteur f, en partant à l'instant $\tau = 0$ d'une distribution Λ_0^α où tous les promoteurs sont réprimés par la présence dans le milieu d'α-amanitine (Shang *et al.*, 2000; Métivier *et al.*, 2003) ou, dans d'autres travaux, par l'absence de

différentielle décrivant un OHA (dont la solution est uniquement la somme de deux fonctions $e^{\lambda_i\tau}$) a une forme différente de (III.12). D'autre part, le facteur d'amortissement ζ ne permet pas de qualifier directement une fonction $e^{\lambda_i\tau}$ mais se rapporte à l'oscillateur lui même. En effet, ζ correspond à un rapport entre les paramètres de l'OHA dont la valeur détermine si les deux λ_i sont réels ($\zeta \geq 1$, régime sur-amorti) ou non-réels et conjugées ($0 \leq \zeta < 1$, régime sous-amorti). Les facteurs de cohérence des deux fonctions $e^{\lambda_i\tau}$ sont alors nuls dans le cas sur-amorti et valent tous les deux $\kappa = \sqrt{1-\zeta^2}/2\pi\zeta$ dans le cas sous-amorti.

[21] Il s'agit d'un système où l'on peut trouver au moins deux sous-ensembles d'états $\mathcal{S}_1$ et $\mathcal{S}_2$ tels qu'il soit impossible de passer de $\mathcal{S}_1$ à $\mathcal{S}_2$ ou de $\mathcal{S}_2$ à $\mathcal{S}_1$, même en un temps infini.

cuivre (Karpova *et al.*, 2008). On peut considérer explicitement, dans notre modèle, cette molécule servant à synchroniser la population comme pouvant interagir avec le promoteur (c'est-à-dire comme un FT au sens de notre modèle). La distribution initiale Λ_0^α peut alors

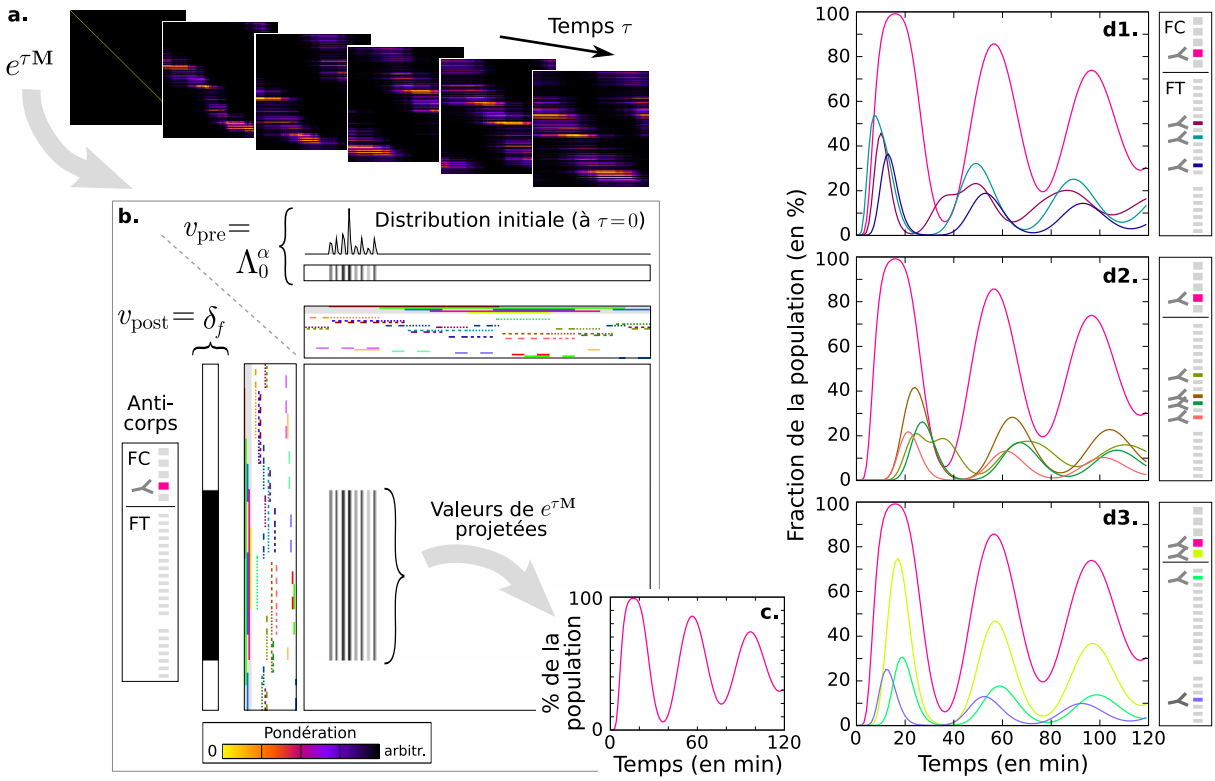


Fig. III.9: Prédiction des cinétiques de ChIP. (a) Les différentes cinétiques de ChIP que l'on peut obtenir sur un promoteur sont toutes des projections de l'exponentielle matricielle $e^{\tau \mathbf{M}}$, identique à celle présentée à la figure III.6. (b) Ces projections consistent à faire une somme partielle des valeurs de $e^{\tau \mathbf{M}}$ dont les pondérations sont déterminées par deux vecteurs v_{pre} et v_{post} : chaque colonne est pondérée par un élément de v_{pre} et chaque ligne, par un élément de v_{post} (équation III.14). Dans le cas d'une cinétique de ChIP, ceux-ci correspondent respectivement à la distribution initiale du système Λ_0^α et à un vecteur représentant l'efficacité de l'anticorps sur les différents états du promoteur. Dans cet exemple, la distribution initiale Λ_0^α est obtenue comme l'état stationnaire en l'absence d'un certain FT (le FT ▬, participant à la modification du FC ■, et dont l'absence bloque le système dans un état chromatinien donné). L'anticorps utilisé ici est $\boldsymbol{\delta}_{\blacksquare}$, un cas idéal considérant une spécificité et une efficacité parfaites. (c) La somme pondérée des valeurs de $e^{\tau \mathbf{M}}$ correspond à l'évolution au cours du temps τ de la fraction de la population de cellules dont le promoteur va immuno-précipiter avec l'anticorps utilisé, décrivant donc la cinétique de ChIP. (d1-3) L'utilisation de différents anticorps sur ce même promoteur permet d'obtenir des cinétiques décrivant la présence de différents FT ou la modification des différents FC (présentées ici en fonction de leur instant d'apparition dans le cycle). Chaque cinétique est la somme des fonctions $e^{\lambda_i \tau}$ définies par le spectre de la matrice $\mathbf{M}$ (figures III.7 et III.8). Pour une cinétique donnée, la contribution de chaque mode est déterminée par la projection de sa matrice $\Lambda_i \Lambda_i^{-1}$ comme décrit par (b). On remarque ici clairement dans toutes les cinétiques (d1-3) l'importance du mode périodique le plus lent du système (associé à la valeur propre λ_1 ; figure III.7b4) dont la période d'oscillation est de 40 min.

être obtenue comme étant la distribution stationnaire du système lorsque ce FT est en concentration forte. On peut alors observer, à partir de cette distribution, l'évolution du système décrit par la matrice $\mathbf{M}$ pour laquelle l'α-amanitine est en concentration nulle. La distribution après un délai τ est donc le vecteur $e^{\tau\mathbf{M}}\Lambda_0^\alpha$ et la probabilité qu'un facteur f soit présent s'obtient par le produit scalaire avec $\boldsymbol{\delta}_f$, le vecteur à N éléments dont le s-ième élément vaut 1 si $f \in s$ et 0 sinon.

$$\mathrm{ChIP}_f(\tau) = {\boldsymbol{\delta}_f}^{\mathrm{T}} e^{\tau\mathbf{M}}\Lambda_0^\alpha \tag{III.19}$$

$$= \sum_i e^{\tau\lambda_i}\left[{\boldsymbol{\delta}_f}^{\mathrm{T}}\Lambda_i\Lambda_i^{-1}\Lambda_0^\alpha\right] \tag{III.20}$$

Ainsi, les profils révélés par les différents anticorps lors d'une cinétique de ChIP correspondent à différentes projections de la même exponentielle matricielle (figure III.9) et leur décomposition modale montre qu'il s'agit d'une somme des mêmes fonctions $e^{\tau\lambda_i}$, mais pondérées différemment. On retrouve en particulier, dans ces cinétiques, la périodicité lente de 40 min décrite par le spectre de $\mathbf{M}$ (figure III.7b3).

Les équations (III.19) et (III.20) peuvent être généralisées pour représenter des cinétiques réalisées avec des anticorps plus réalistes que ceux décrits idéalement par $\boldsymbol{\delta}_f$. En effet, au lieu d'être binaires, les éléments du vecteur représentant l'anticorps peuvent décrire de façon beaucoup plus fine la probabilité, pour chaque état, que l'anticorps précipite sur le promoteur[22]. Cela permet en particulier de rendre compte des phénomènes d'aspécificité (le fait que l'anticorps puisse immuno-précipiter sur un promoteur où la molécule d'intérêt n'est pas présente), d'efficacité partielle et/ou dépendant de plusieurs facteurs, ... Les cinétiques de ChIP obtenues décrivent alors le rapport entre la quantité d'ADN immuno-précipité et la quantité de référence (obtenue sans anticorps), de façon similaire à la technique expérimentale (cf Box 2, page 52).

Corrélations croisées et spectres de puissance

Bien qu'ils ne correspondent pas directement à la reproduction de mesures expérimentales telles que la prédiction de données de ChIP, les corrélations croisées (dont l'autocorrélation est un cas particulier) et les spectres de puissance sont des indicateurs essentiels pour quantifier des dynamiques complexes. Il s'agit de mesures relativement communes en théorie du signal permettant de décrire les relations temporelles entre signaux (cf Box 1, page 22). L'utilisation de ces indicateurs pour caractériser notre modèle illustre typiquement que celui-ci fournit un cadre commun pour l'étude de la dynamique du promoteur et pour l'étude de la stochasticité de l'expression génique. En effet, d'une part, ces indicateurs partagent d'importants points communs avec les cinétiques de ChIP (en particulier dans leur structure modale) et, d'autre part, ce sont eux qui nous permettront par la suite de déterminer comment la stochasticité du promoteur se propage au niveau d'expression du gène.

De nombreux signaux peuvent être définis à partir de la succession stochastique d'états du promoteur, en associant simplement à chaque état une valeur définie par un vecteur.

[22] D'un point de vue pratique, il est possible d'inférer ces vecteurs à partir de données de Re-ChIP.

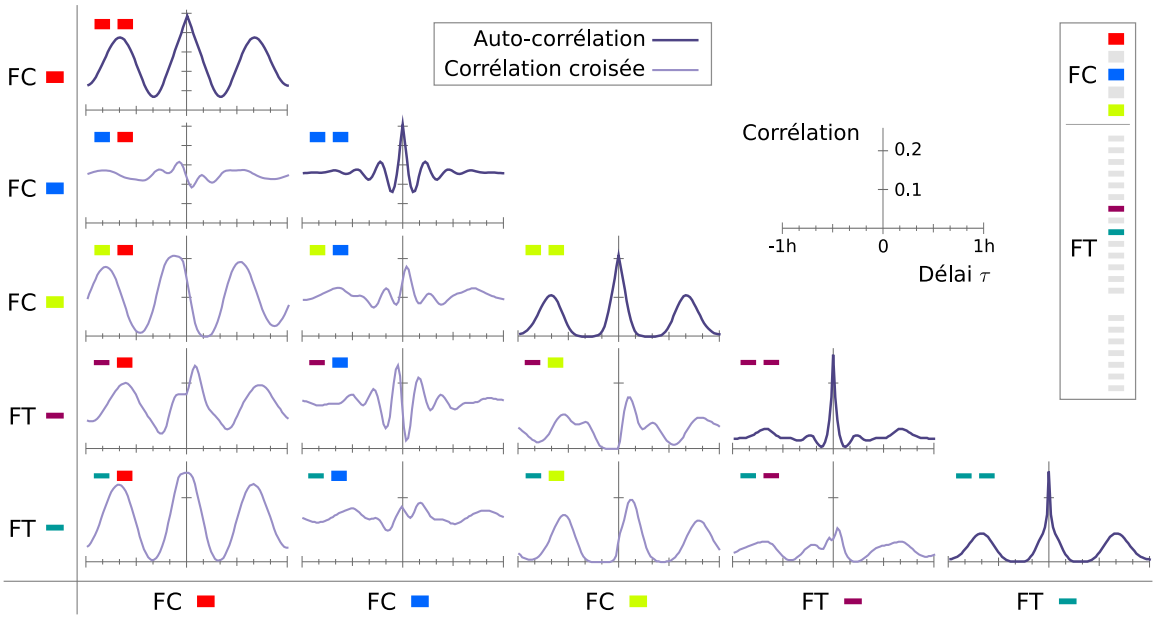


Fig. III.10: Corrélations entre la présence des facteurs. La répartition dans le temps de la présence d'un FT sur le promoteur ou de la modification d'un FC, relativement à la présence/modification d'un autre facteur peut être décrite par leur corrélation croisée. Il s'agit simplement d'une mesure de la similarité entre les deux signaux décrivant leur présence au cours du temps, comparés avec différents délais. Cette figure illustre toutes les corrélations croisées et auto-corrélations (corrélation d'un facteur avec lui même) possibles entre 5 facteurs (3 FC et 2 FT) sur l'exemple de promoteur eucaryote décrit à la figure III.4. L'auto-corrélation du FC ■ montre par exemple que, 10 min après avoir observé la présence de ce facteur, il est fort probable qu'il soit absent, et ce, pendant les 20 min suivantes. De plus, la corrélation croisée ▬ ■ indique que, lorsque le FT ▬ est présent, le FC ■ a peu de chance d'avoir été présent les 15 min précédentes mais le sera probablement au cours des 15 min suivantes. De la même façon, on peut observer des profils plus complexes (tels que ▬ ■ et ▬ ▬) témoignant de niveaux de corrélation distincts sur différents intervalles de temps. Par ailleurs, l'auto-corrélation des différents facteurs met en évidence des périodicités lentes dans leur probabilité de présence sur le promoteur. On note que les oscillations du FC ■, sensiblement plus rapides que celles de deux autres, proviennent du fait que ce FC est modifié 6 fois au cours du cycle et non 2 (cf figure III.4a3). Les corrélations croisées montrent alors que les facteurs peuvent osciller de façon synchrone (▬ ■) ou avec un certain décalage de phase (■ ■). Enfin, les auto-corrélations des deux FT reflètent la dynamique rapide d'association/dissociation de ces molécules, contrainte sur des temps longs par la dynamique périodique de la chromatine. En effet, on observe (en particulier sur ▬ ▬) un régime rapide (un pic fin autour de τ=0) en plus du régime lent et périodique, indiquant que lorsque la molécule est associée, il est fort probable qu'elle se dissocie rapidement, mais que d'autres pourront se réassocier typiquement dans les 10 min suivantes, puis de nouveaux après 20 min et pendant 20 min, ... (NB : pour des raisons de clarté, les échelles des ordonnées sont différentes d'une ligne à l'autre).

Par exemple, si ce vecteur est $\boldsymbol{\delta}_f$ le signal obtenu décrit la présence/absence du facteur f au cours du temps (comme sur les trois premières courbes de la figure III.2d). Si ce vecteur est $\boldsymbol{\rho}$, le signal en question est simplement l'efficacité transcriptionnelle du promoteur au cours du temps, soit le processus $X(t)$ (quatrième courbe de la figure III.2d). On peut également définir un vecteur dont chaque élément représente par exemple la

distance physique entre deux nucléotides donnés du promoteur (dépendant en particulier de l'enroulement/déroulement de l'ADN autour d'un nucléosome et/ou de la position de ce dernier le long de la séquence), voire la puissance d'émission du donneur ou de l'accepteur si ces deux nucléotides sont marqués avec des tags fluorescents comme lors d'une expérience de FRET (cf Box 2, page 52). Le signal obtenu dans ce dernier cas est alors le signal d'émission du donneur au cours du temps, tel qu'analysé dans certaines expériences de FRET (Li *et al.*, 2005; Poirier *et al.*, 2009).

La corrélation croisée entre deux signaux $a(t)$ et $b(t)$ est définie comme $\big\langle a(t)\, b(t+\tau)\big\rangle_t$ (où $\langle\ \rangle_t$ dénote la moyenne temporelle) et correspond à une auto-corrélation si $a(t)$ et $b(t)$ sont identiques[23] (cf Box 1, page 22). Pour tout signaux $a(t)$ et $b(t)$ définis à partir de deux vecteurs $\mathbf{a}$ et $\mathbf{b}$, leur corrélation croisée est alors[24]

$$\big\langle a(t)\, b(t+\tau)\big\rangle_t = \mathbf{b}^{\mathrm{T}}\, e^{\tau\mathbf{M}}\, \mathbf{D_a}\, \Lambda_0 \tag{III.21}$$

$$= \sum_i e^{\tau\lambda_i} \left[\mathbf{b}^{\mathrm{T}} \Lambda_i \Lambda_i^{-1} \mathbf{D_a} \Lambda_0\right] \tag{III.22}$$

pour $\tau \geq 0$ (dans le cas $\tau < 0$, il suffit de prendre $-\tau$ et d'inverser $\mathbf{a}$ et $\mathbf{b}$).

Cette expression générale de la corrélation entre signaux arbitraires décrivant la dynamique du promoteur peut alors être déclinée de nombreuses façons :

- Pour $\mathbf{a} = \boldsymbol{\delta}_f$ et $\mathbf{b} = \boldsymbol{\delta}_{f'}$, il s'agit de la corrélation croisée entre la présence de f et celle de f' sur le promoteur. Elle indique simplement comment, au sein de la dynamique stochastique de toute la structure, la présence de f' se répartit dans le temps par rapport à celle de f (figure III.10).
- Le cas particulier $\mathbf{a} = \mathbf{b} = \boldsymbol{\delta}_f$ permet d'obtenir l'auto-corrélation de la présence de f (figures III.10 et III.11a2, b2 et c2).
- En prenant $\mathbf{a} = \mathbf{b}$, représentant la puissance d'émission du donneur dans une expérience de FRET, on prédit alors l'auto-corrélation du signal d'émission du donneur observé sur une structure unique. Cet indicateur est par exemple utilisé expérimentalement pour déterminer les temps caractéristiques dans la dynamique de la chromatine à différentes échelles (Li *et al.*, 2005; Poirier *et al.*, 2009).
- Si $\mathbf{a} = \mathbf{b} = \boldsymbol{\rho}$, le taux de transcription associé aux différents états (figure III.11d1), on obtient l'auto-corrélation du processus $X(t)$, l'efficacité transcriptionnelle du promoteur (figure III.11d2). Comme nous le verrons, cette auto-corrélation sert de base à de nombreuses prédictions quant à la transmission de la stochasticité du promoteur aux niveaux d'ARN et de protéines.
- Enfin, pour un système comportant deux gènes (partageant par exemple un même environnement chromatinien ; cf figure III.1f), si $\mathbf{a} = \boldsymbol{\rho}$ et $\mathbf{b} = \boldsymbol{\rho}'$ représentent les taux de transcription respectifs de ces gènes en fonction de l'état de toute la structure, on obtient alors la corrélation croisée des deux processus d'efficacité transcriptionnelle[25].

[23] La définition de la corrélation croisée et de l'auto-corrélation varie d'une discipline à l'autre. Elle est par exemple parfois considérée comme centrée et réduite $\big\langle (a(t) - \langle a\rangle)\,(b(t+\tau) - \langle b\rangle)\big\rangle_t / \sigma_a \sigma_b$.

[24] $\big\langle a(t)\, b(t+\tau)\big\rangle_t = \sum_{s,s'} a_s\, b_{s'}\, \mathrm{P}(s', t+\tau \wedge s, t)$, ce qui se reformule en $\sum_{s,s'} b_{s'}\, \mathrm{P}(s', t+\tau | s, t)\, a_s\, \mathrm{P}(s,t)$ et donc en (III.21).

[25] Nous n'irons pas plus loin dans le cadre de cette thèse, mais cet indicateur peut servir de point

Ainsi, comme pour les cinétiques de ChIP, toutes les corrélations (croisées ou non) que l'on peut mesurer sur le promoteur constituent des projections différentes de $e^{\tau \mathbf{M}}$ (selon le même principe que la figure III.9b mais avec d'autres vecteurs v_{pre} et v_{post}). Elles sont donc également construites à partir du même ensemble de fonction $e^{\lambda_i \tau}$, décrit par le spectre de la matrice $\mathbf{M}$ (figures III.7 et III.8).

---

de départ pour prédire un certain nombre de mesures sur l'expression conjointe des deux gènes, et en particulier, leur coefficient de corrélation. Cette mesure, souvent déduite de données de cytométrie, est utilisée pour séparer les source de stochasticité d'origine intrinsèque et extrinsèque (figure I.1, page 20), permettant ici de quantifier l'importance de la dynamique locale de la chromatine.

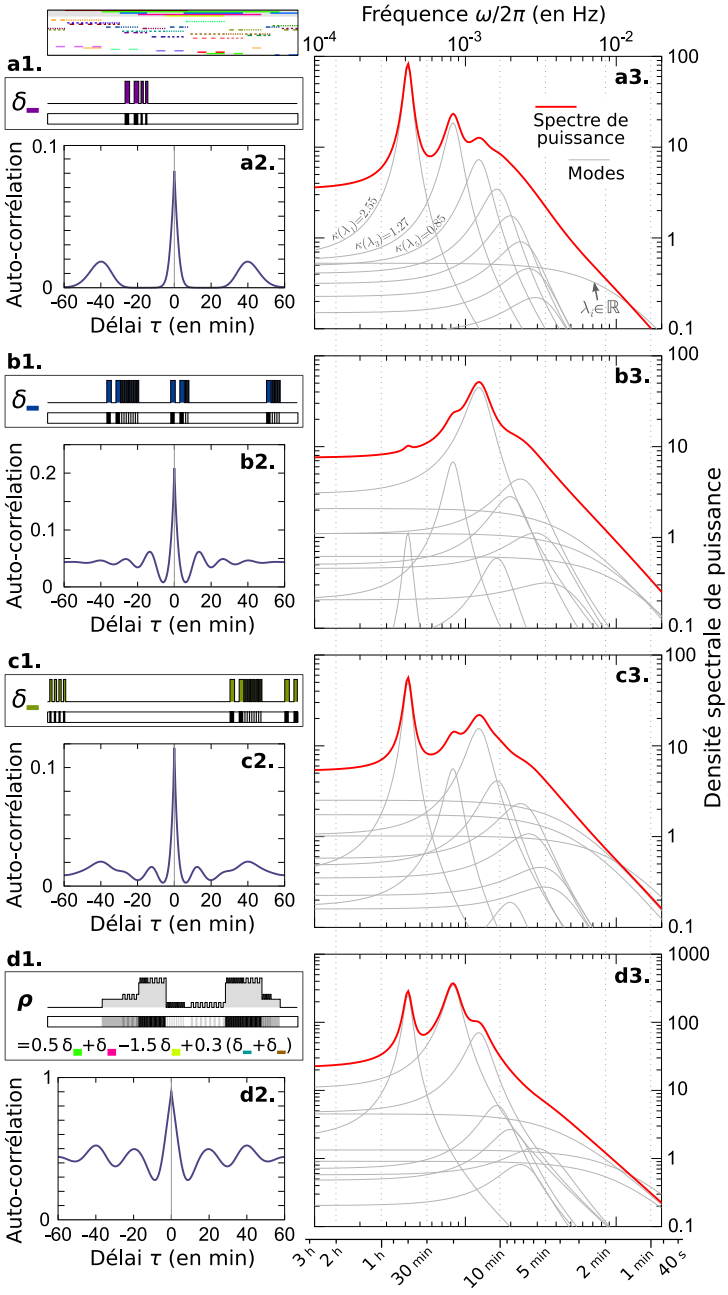


Fig. III.11: Spectres de puissance et auto-corrélations. Le contenu fréquentiel d'un signal peut être décrit par son spectre de puissance, un indicateur obtenu par la transformée de Fourier de l'auto-corrélation et ne constituant donc qu'une représentation différente des même informations. Cette figure montre les spectres de puissance de quatre signaux représentant respectivement la présence de 3 FT (a,b,c) et le processus $X(t)$ d'efficacité transcriptionnelle du promoteur (d). Ces signaux sont chacuns déduits de la succession d'états du promoteur à partir d'un vecteur (a1-d1), et leurs auto-corrélations décrivent des comportements dynamiques différents (a2-d2) se traduisant dans leurs profils spectraux (a3-d3). La décomposition modale permet de comprendre la structure de ces spectres (seuls les modes les plus importants sont présentés ici). Chaque pic observé dans un spectre correspond à un mode, c'est-à-dire à une valeur propre de $\mathbf{M}$ (cf figure III.7a) et tous les spectres dérivés du même promoteur ne sont que des pondérations différentes des même modes. Dans le cas (a3) (où ces pondérations sont très régulières car les valeurs dans le vecteur (a1) sont très localisées), on reconnaît clairement les modes périodique lents et cohérents, oscillant à des périodes de 40 min, 20.2 min, 13.6 min, ..., et pour lesquels la finesse du pic décrit la cohérence du mode.

Le spectre de puissance permet de décrire le contenu fréquentiel d'un signal, c'est-à-dire comment les fluctuations se répartissent sur les différentes fréquences. Il correspond simplement à la transformée de Fourier de l'auto-corrélation (théorème de Wiener-Khinchin). Pour un signal $a(t)$, son expression est[26]

$$S_a(\omega) = \sum_i \frac{-2\lambda_i}{{\lambda_i}^2 + \omega^2} \left[\mathbf{a}^{\mathrm{T}} \Lambda_i \Lambda_i^{-1} \mathbf{D_a} \Lambda_0\right] \tag{III.23}$$

où ω, appelé fréquence angulaire, correspond à la fréquence à un coefficient 2π près. L'auto-corrélation et le spectre de puissance constituent deux représentations différentes de la même information (respectivement dans le domaine temporel et dans le domaine fréquentiel). Ce dernier reflète donc également la décomposition en modes propres décrite par le spectre de la matrice $\mathbf{M}$. En effet, de la même façon que les auto-corrélations sont des sommes pondérées des mêmes fonctions $e^{\lambda_i \tau}$, les spectres de puissance sont des sommes, identiquement pondérées, des même fonctions $\frac{-2\lambda_i}{{\lambda_i}^2+\omega^2}$. La forme de chacune de ces composantes est également déterminée par la position sur le plan complexe de la valeur propre λ_i qui lui est associée. En représentation logarithmique, ces composantes décrivent toujours un plateau dans la limite des basses fréquences et présentent une décroissance linéaire en $1/\omega^2$ dans la limite des hautes fréquences. Lorsque la valeur propre est réelle ($\lambda_i \in \mathbb{R}$ sur la figure III.11a3), la région intermédiaire ne comporte qu'un coude décroissant à la fréquence $\omega = \mathrm{Re}(-\lambda_i)$, décrivant simplement la vitesse de relaxation de la fluctuation apériodique correspondante. Une valeur propre non réelle décrit une fluctuation périodique stochastique dans le signal et produit quant à elle un pic à la fréquence $\omega = |\mathrm{Im}(-\lambda_i)|$. Ce pic est d'autant plus fin que la cohérence $\kappa(\lambda_i)$ du mode est élevée. Par exemple, les modes périodiques les plus lents et cohérents de la figure III.7a sont représentés sur la figure III.11a3.

Par la suite, on notera β_i les coefficients pour le cas particulier du processus d'efficacité transcriptionnelle $X(t)$

$${\beta_i}^2 = \boldsymbol{\rho}^{\mathrm{T}} \Lambda_i \Lambda_i^{-1} \mathbf{D}_{\boldsymbol{\rho}} \Lambda_0 \tag{III.24}$$

L'auto-corrélation et le spectre de puissance de $X(t)$ s'écrivent alors respectivement

$$\left\langle X(t)\, X(t+\tau) \right\rangle_t = \sum_i {\beta_i}^2\, e^{|\tau|\lambda_i} \tag{III.25}$$

$$S_X(\omega) = \sum_i {\beta_i}^2 \frac{-2\lambda_i}{{\lambda_i}^2 + \omega^2} \tag{III.26}$$

et le taux de transcription moyen correspond simplement à $\langle X \rangle = \beta_0$.

Distributions de durée : résidence d'un FT, durée d'un état chromatinien, ...

Il est possible de déterminer la distribution de la durée $\tau_{\mathcal{U}}$ que le système passe dans un certain sous-ensemble $\mathcal{U} \subset \mathcal{S}$ d'états. Si $\mathcal{U}$ représente tous les états où un facteur donné est présent (resp. absent), il s'agit alors de la distribution du temps de résidence (resp.

[26] La transformée de Fourier (en fréquence angulaire) de $e^{-\alpha|t|}, \alpha > 0$ est la fonction lorentzienne $\frac{2\alpha}{\alpha^2+\omega^2}$.

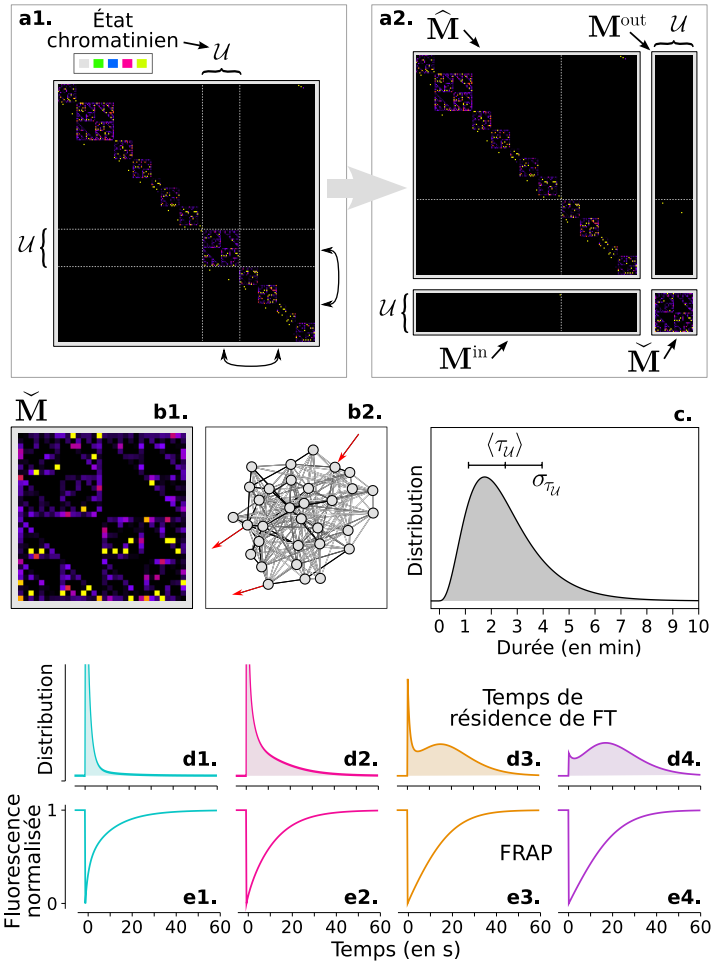
d'absence) du facteur. Si $\mathcal{U}$ correspond à tous les états ayant une même configuration de FC, on obtient la distribution de la durée que le système passe dans un certain état chromatinien indépendamment des associations/dissociations des FT.

On peut réorganiser la matrice $\mathbf{M}$ de la façon suivante[27]

$$\mathbf{M} = \begin{array}{c} \\ s'\notin\mathcal{U} \\ s'\in\mathcal{U} \end{array} \begin{array}{c} \begin{array}{cc} s\notin\mathcal{U} & s\in\mathcal{U} \end{array} \\ \left[\begin{array}{c|c} \text{Dynamique hors de } \mathcal{U} & \text{Sorties de } \mathcal{U} \\ \hline \text{Entrées dans } \mathcal{U} & \text{Dynamique dans } \mathcal{U} \end{array}\right] \end{array} = \begin{bmatrix} \hat{\mathbf{M}} & \mathbf{M}^{\text{out}} \\ \mathbf{M}^{\text{in}} & \check{\mathbf{M}} \end{bmatrix} \tag{III.27}$$

[27] Les lignes et les colonnes de $\mathbf{M}$ peuvent être arbitrairement permutées entre elles tant que l'ordre des états reste le même sur les lignes et sur les colonnes.

Fig. III.12: Distributions de durée et courbes de FRAP. Pour un ensemble $\mathcal{U}$ d'états, correspondant ici à un même état de la chromatine, la matrice $\mathbf{M}$ (a1) peut être réorganisée[27] de façon à faire apparaître quatre sous-matrices (a2), représentant chacune différents aspects de la dynamique relativement à $\mathcal{U}$ (cf équation III.27). La sous-matrice $\check{\mathbf{M}}$ (b1) décrit alors la dynamique au sein de cet état chromatinien (b2) et permet en particulier de dériver la distribution de la durée entre l'instant d'entrée et l'instant de sortie de ce sous ensemble d'états (c) et l'ensemble de ses moments. Cette méthode permet aussi bien d'étudier la durée des états chromatiniens que le temps de résidence ou d'absence des facteurs. (d) On observe diverses distributions du temps de résidence d'un FT pour des systèmes faisant intervenir différents mécanismes d'interaction entre molécules (ces systèmes, similaires dans leur structure au promoteur eucaryote présenté jusqu'ici, seront décrits en détails dans la section 4.4). (d) Dans ces différents cas, les courbes de FRAP présentent de légères différences dont nous montrons qu'elles comportent – en théorie – l'information nécessaire pour en déduire la distribution complète des temps de résidence de la molécule. (d1,e1 : figure III.29b ; d2,e2 : figure III.30b1 ; d3-4,e3-4 : figure III.30b2).



comme l'illustre la figure III.12a, dans le cas où $\mathcal{U}$ représente un état de la chromatine. On note alors $\check{\lambda}_i$ les valeurs propres de la matrice $\check{\mathbf{M}}$, et $\check{\Lambda}_i$ et $\check{\Lambda}_i^{-1}$ ses vecteurs propres à droite et à gauche. Les vecteurs $\Lambda_0^{\mathcal{U}}$ (resp. $\Lambda_0^{\overline{\mathcal{U}}}$) représentent les éléments de Λ_0 correspondant aux états de $\mathcal{U}$ (resp. hors de $\mathcal{U}$). On remarque que, Λ_0 étant solution stationnaire de l'équation maîtresse (III.12) $\mathbf{M}\Lambda_0 = \mathbf{0}$, la décomposition (III.27) implique alors que[28]

$$\begin{cases} \hat{\mathbf{M}}\,\Lambda_0^{\overline{\mathcal{U}}} & = -\mathbf{M}^{\text{out}}\,\Lambda_0^{\mathcal{U}} \\ \mathbf{M}^{\text{in}}\,\Lambda_0^{\overline{\mathcal{U}}} & = -\check{\mathbf{M}}\,\Lambda_0^{\mathcal{U}} \end{cases} \tag{III.28}$$

On note $\mathbf{1} = [1, 1, \ldots, 1]$ le vecteur ligne permet de faire la somme des éléments d'un vecteur, par exemple $\mathbf{1}\,\Lambda_0 = 1$. La probabilité d'être dans $\mathcal{U}$ est donc simplement $\mathbf{1}\,\Lambda_0^{\mathcal{U}}$ et les taux moyens d'entrée et de sortie de $\mathcal{U}$ (qui sont égaux à l'état stationnaire) valent

$$J = \mathbf{1}\,\mathbf{M}^{\text{in}}\,\Lambda_0^{\overline{\mathcal{U}}} = \mathbf{1}\,\mathbf{M}^{\text{out}}\,\Lambda_0^{\mathcal{U}} = -\mathbf{1}\,\check{\mathbf{M}}\,\Lambda_0^{\mathcal{U}} = -\mathbf{1}\,\hat{\mathbf{M}}\,\Lambda_0^{\overline{\mathcal{U}}} \tag{III.29}$$

(les deux dernières expressions sont déduites de (III.28)). Le temps moyen entre deux entrées consécutives est alors $1/J$ et la durée moyenne passée dans $\mathcal{U}$ est $\langle \tau_{\mathcal{U}} \rangle = \mathbf{1}\,\Lambda_0^{\mathcal{U}}/J$.

La somme sur chaque colonne de $\check{\mathbf{M}}$ n'est pas nulle car les éléments diagonaux prennent en compte les transitions sortantes décrites dans $\mathbf{M}^{\text{out}}$. Ainsi, pour l'exponentielle matricielle $e^{\tau\check{\mathbf{M}}}$ de la sous-matrice $\check{\mathbf{M}}$, la somme sur chaque colonne s ne vaut 1 que pour $\tau = 0$ et décroît jusqu'à 0 lorsque τ augmente. Cette quantité représente simplement la probabilité, en partant de s, de ne pas avoir quitté $\mathcal{U}$ au bout d'un temps τ. Ainsi, l'expression

$$1 - \mathbf{1}\; e^{\tau\check{\mathbf{M}}}\; \phi_0 \tag{III.30}$$

est la *fonction de répartition* décrivant la probabilité d'avoir quitté $\mathcal{U}$ après un temps τ, en partant d'une distribution d'états ϕ_0. La distribution de probabilité de l'état par lequel le système rentre dans $\mathcal{U}$ est $\mathbf{M}^{\text{in}}\Lambda_0^{\overline{\mathcal{U}}}/J$ (et vaut également $-\check{\mathbf{M}}\,\Lambda_0^{\mathcal{U}}/J$ d'après (III.28)). Ainsi, la distribution de probabilité de la durée $\tau_{\mathcal{U}}$ que le système passe dans $\mathcal{U}$ est

$$\mathrm{P}(\tau_{\mathcal{U}} = \tau) = \frac{\mathrm{d}}{\mathrm{d}\tau}\left[1 - \mathbf{1}\; e^{\tau\check{\mathbf{M}}}\; \frac{\mathbf{M}^{\text{in}}\Lambda_0^{\overline{\mathcal{U}}}}{J}\right] \tag{III.31}$$

$$= \mathbf{1}\, e^{\tau\check{\mathbf{M}}}\; \check{\mathbf{M}}^2\, \frac{\Lambda_0^{\mathcal{U}}}{J} \tag{III.32}$$

$$= \sum_i e^{\tau\check{\lambda}_i} \left[\check{\lambda}_i{}^2\, \mathbf{1}\check{\Lambda}_i\, \check{\Lambda}_i^{-1} \frac{\Lambda_0^{\mathcal{U}}}{J}\right] \tag{III.33}$$

On peut également obtenir les différents moments de la distribution par[29]

$$\left\langle \tau_{\mathcal{U}}{}^q \right\rangle = \int_0^\infty \tau^q\, \mathrm{P}(\tau_{\mathcal{U}} = \tau)\, \mathrm{d}\tau \tag{III.34}$$

$$= q!\, \mathbf{1}\big(-\check{\mathbf{M}}\big)^{1-q} \frac{\Lambda_0^{\mathcal{U}}}{J} \tag{III.35}$$

$$= q! \sum_i \big(-\check{\lambda}_i\big)^{1-q}\, \mathbf{1}\check{\Lambda}_i\, \check{\Lambda}_i^{-1} \frac{\Lambda_0^{\mathcal{U}}}{J} \tag{III.36}$$

[28] Cela revient à remarquer, pour chaque état s, que les sorties de s (élément diagonal de $\check{\mathbf{M}}\,\Lambda_0^{\mathcal{U}}$) contrebalancent les entrées dans s provenant à la fois des états hors de $\mathcal{U}$ (le produit $\mathbf{M}^{\text{in}}\,\Lambda_0^{\overline{\mathcal{U}}}$) et des états de $\mathcal{U}$ (éléments non diagonaux de $\check{\mathbf{M}}\,\Lambda_0^{\mathcal{U}}$).

[29] En remarquant que $\int_0^\infty e^{ax}x^b\mathrm{d}x = b!/(-a)^{b+1}, \forall a < 0, b \geq 0$, par intégrations par parties récurrentes.

Ainsi, la distribution de probabilité de la durée passée dans le sous-ensemble $\mathcal{U}$ d'états ne dépend que de la sous-matrice $\check{\mathbf{M}}$ et de la distribution stationnaire du système Λ_0. Si l'on s'intéresse, par exemple, à la durée d'un état chromatinien, cette sous-matrice et le sous-graphe correspondant ne constituent qu'une faible partie du système complet (figures III.12b1-2). De même, pour obtenir la distribution du temps de résidence d'un facteur donné (figure III.12d1-4), l'ensemble $\mathcal{U}$, représentant tous les états où le facteur est présent, ne constitue souvent qu'une faible fraction de l'espace d'états $\mathcal{S}$. Du point de vue de la décomposition modale (équations (III.33) et (III.36)), cela implique en particulier que le nombre de modes (autant qu'il y a d'éléments dans $\mathcal{U}$) est bien inférieur que dans les cas précédents (ChIP, corrélation et spectres de puissance). De plus, les valeurs propres de $\check{\mathbf{M}}$ sont différentes de celles de $\mathbf{M}$ et ne présentent pas la structure en ellipse due à la dynamique globale lente du système complet (figure III.7a). Comme pour le ChIP et les corrélations croisées, les distributions ne sont que des sommes pondérées de fonctions de la forme $e^{\check{\lambda}_i \tau}$ (telles que décrites sur la figure III.8a). Mais les valeurs propres étant toutes réelles ou à cohérence très faible, ces modes sont essentiellement apériodiques.

Courbes de FRAP

Le FRAP permet de mesurer le turnover d'une molécule d'intérêt rendue fluorescente sur un promoteur donné, en observant le rétablissement de la fluorescence d'une petite zone de l'espace nucléaire contenant un grand nombre de copies du promoteur après que celle-ci ait été photoblanchie (cf Box 2, page 52). Pour reproduire cette mesure, nous considérons explicitement comme deux molécules différentes le FT fluorescent $f^\star$ et le FT photoblanchi f°, définissant ainsi un ensemble $\tilde{\mathcal{S}}$ d'états où tous les états de $\mathcal{S}$ pour lesquels le FT d'intérêt est présent ont été dupliqués, et une matrice $\tilde{\mathbf{M}}$ déterminée à partir de $\mathbf{M}$ de façon à ce que $f^\star$ et f° aient le même comportement mais s'excluent mutuellement et que $[f^\circ] = 0$. À partir de la réorganisation de la matrice $\mathbf{M}$ décrite par (III.27) où $\mathcal{U}$ représente les états pour lesquels f est présent

$$\mathbf{M} = \begin{array}{c} \\ f\notin s' \\ f\in s' \end{array} \begin{array}{c} \begin{array}{cc} f\notin s & f\in s \end{array} \\ \left[\begin{array}{cc} \boxed{\text{Dynamique sans } f} & \boxed{\text{Dissociation de } f} \\ \boxed{\text{Association de } f} & \boxed{\text{Dynamique avec } f} \end{array}\right] \end{array} = \begin{bmatrix} \hat{\mathbf{M}} & \mathbf{M}^{\mathrm{d}} \\ \mathbf{M}^{\mathrm{a}} & \check{\mathbf{M}} \end{bmatrix} \tag{III.37}$$

la matrice $\tilde{\mathbf{M}}$ est construite simplement selon

$$\tilde{\mathbf{M}} = \begin{array}{c} \\ f^\star\notin s' \\ f^\circ\notin s' \\ f^\star\in s' \\ f^\circ\notin s' \\ f^\star\notin s' \\ f^\circ\in s' \end{array} \begin{array}{c} \begin{array}{ccc} f^\star\notin s,\, f^\circ\notin s & f^\star\in s,\, f^\circ\notin s & f^\star\notin s,\, f^\circ\in s \end{array} \\ \left[\begin{array}{ccc} \boxed{\text{Dynamique sans } f^\star \text{ ni } f^\circ} & \boxed{\text{Dissociation de } f^\star} & \boxed{\text{Dissociation de } f^\circ} \\ \boxed{\text{Association de } f^\star} & \boxed{\text{Dynamique avec } f^\star} & 0 \\ \boxed{\text{Association de } f^\circ} & 0 & \boxed{\text{Dynamique avec } f^\circ} \end{array}\right] \end{array} = \begin{bmatrix} \hat{\mathbf{M}} & \mathbf{M}^{\mathrm{d}} & \mathbf{M}^{\mathrm{d}} \\ \mathbf{M}^{\mathrm{a}} & \check{\mathbf{M}} & 0 \\ 0 & 0 & \check{\mathbf{M}} \end{bmatrix} \tag{III.38}$$

(le bloc en bas à gauche vaut 0 parce que $[f^\circ] = 0$). On note $\tilde{\lambda}_i$ les valeurs propres de $\tilde{\mathbf{M}}$, et $\tilde{\Lambda}_i$ et $\tilde{\Lambda}_i^{-1}$ les vecteurs propres à droite et à gauche. L'état stationnaire $\tilde{\Lambda}_0$ vaut 0 pour les états où f° est présent et Λ_0 pour les autres. Une courbe de FRAP correspond alors à la probabilité au cours du temps τ que $f^\star$ soit présent sur le promoteur, en partant à $\tau = 0$ de la distribution $\tilde{\Lambda}_0^\circ$ obtenue en photoblanchissant $\tilde{\Lambda}_0$ (c'est-à-dire en transférant les probabilités associées aux états contenant $f^\star$ aux états équivalents contenant f°). Le niveau de fluorescence, renormalisé par sa valeur avant photoblanchiment, s'écrit alors

$$\mathrm{FRAP}_f(\tau) = \boldsymbol{\delta}_{f^\star}{}^{\mathrm{T}}\, e^{\tau\tilde{\mathbf{M}}}\, \frac{\tilde{\Lambda}_0^\circ}{\boldsymbol{\delta}_{f^\star}{}^{\mathrm{T}}\, \tilde{\Lambda}_0} \tag{III.39}$$

On peut montrer (cf annexe A) que l'ensemble des valeurs propres de $\tilde{\mathbf{M}}$ est l'union de celles de $\mathbf{M}$ et de celle de $\check{\mathbf{M}}$. De plus, les vecteurs propres de $\tilde{\mathbf{M}}$ (que l'on peut déterminer à partir de ceux de $\mathbf{M}$ et de $\check{\mathbf{M}}$) sont tels que tous les modes de (III.39) associés aux valeurs propres de $\mathbf{M}$ sont nuls, à l'exception de λ_0. Ainsi, une courbe de FRAP ne dépend que de $\check{\mathbf{M}}$ et ne reflète donc que la dynamique de dissociation d'un FT. On peut alors reformuler l'expression (III.39) plus simplement en

$$\mathrm{FRAP}_f(\tau) = 1 - \mathbf{1} e^{\tau\check{\mathbf{M}}} \frac{\Lambda_0^{\mathcal{U}}}{\mathbf{1}\,\Lambda_0^{\mathcal{U}}} \tag{III.40}$$

$$= 1 - \sum_i e^{\tau\check{\lambda}_i} \left[\mathbf{1}\,\check{\Lambda}_i \check{\Lambda}_i^{-1} \frac{\Lambda_0^{\mathcal{U}}}{\mathbf{1}\,\Lambda_0^{\mathcal{U}}} \right] \tag{III.41}$$

Une courbe de FRAP est de la forme (III.30) et correspond donc à la fonction de répartition décrivant la probabilité que le FT f ait quitté le promoteur au bout d'un temps τ, non pas à partir de son instant d'association (comme c'est le cas dans l'équation (III.31)), mais à partir de l'instant du photoblanchiment (donc pris uniformément sur son temps de présence). Cela a pour conséquence de masquer certains détails de la dynamique de dissociation du FT (figures III.12e1-4).

Cependant, on remarque que cette distribution de probabilité, obtenue en dérivant (III.40), correspond (à une renormalisation près) à la fonction de répartition $\mathrm{P}(\tau_{\mathcal{U}} \geq \tau)$ (entre crochets dans (III.31)). En d'autres termes, en dérivant une courbe de FRAP, on obtient une fonction dont la dérivée est la distribution du temps de résidence du FT (depuis son association). Ainsi, on peut obtenir simplement cette distribution à partir de la dérivée seconde de la courbe de FRAP[30]

$$\mathrm{P}(\tau_{\mathcal{U}} = \tau) = -\frac{1}{a}\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}\mathrm{FRAP}_f(\tau) \tag{III.42}$$

où a est une constante de normalisation. Cette propriété permet de faire ressortir des caractéristiques importantes de $\mathrm{P}(\tau_{\mathcal{U}} = \tau)$ à partir d'une courbe de FRAP. Intuitivement, en tout point d'une courbe de FRAP, c'est la vitesse de changement de pente qui indique à quel point un temps de résidence donné du FT est probable. Par exemple, la partie quasi-linéaire sur l'intervalle 0-10 s sur la figure III.12e4 indique qu'il est peu probable

[30] Cela implique par ailleurs que, bien qu'une courbe de FRAP soit une somme d'exponentielles pondérées positivement et négativement (équation (III.41)), elle reste nécessairement croissante et concave.

que le FT réside sur son site moins de 10 s (figure III.12d4). À l'inverse, sur ce même intervalle, la courbure importante observées sur la figure III.12e1 témoigne de la forte probabilité d'un temps de résidence inférieur à 10 s (figure III.12d1).

Cependant, la mise en application pratique de cette propriété pose certaines difficultés techniques. En effet, la dérivation amplifiant les erreurs de mesure, elle requière un signal très précis et nécessite donc de faire des moyennes sur un nombre d'acquisitions bien supérieur à ce qui se fait dans la plupart des études actuelles.

*
* *

Les différents indicateurs que nous venons de présenter décrivent la dynamique d'un même promoteur de points de vue différents et complémentaires : par des indicateurs aussi bien expérimentaux que théoriques et révélant des dynamiques aussi bien locales que globales. De plus, nous avons mis en évidence leurs relations et leur structure modale commune décrite par le spectre de la matrice $\mathbf{M}$ ou d'une sous-matrice. Nous montrons, dans les deux sections suivantes, comment, d'une part, ce spectre est contraint par la consommation énergétique du système et, d'autre part, comment, dans des cas simples, il peut révéler certaines relations entre des paramètres globaux et des comportement observables expérimentalement.

3.4 Consommation énergétique et périodicités

Contrairement à notre approche cinétique de la régulation génique, les formalismes thermodynamiques purs (Ackers *et al.*, 1982; Bintu *et al.*, 2005a) ne peuvent prédire que la distribution du système à l'équilibre et sont restreints à l'étude de systèmes fermés (c'est-à-dire ne consommant pas d'énergie extérieure). Dans ce cadre, la distribution à l'équilibre Λ_0 peut être obtenue à partir des énergies libres $\mathbf{G}$ des différents états système fermé, correspondant dans notre modèle à une matrice $\mathbf{E}$ symétrique, il s'agit en effet d'une solution stationnaire $\mathbf{M}\, e^{\frac{-\mathbf{G}}{\mathrm{k_B}T}} = \mathbf{0}$ de l'équation maîtresse (III.12) qui vérifie par ailleurs le principe d'équilibre détaillé[31]. On peut alors montrer, dans ce cas, que la matrice $\mathbf{M}$ n'a nécessairement que des valeurs propres réelles[32]. Cela implique que tout système ne consommant pas d'énergie extérieure a nécessairement un comportement apériodique (tous les modes sont de simples exponentielles). Ainsi, un comportement périodique tel qu'observé sur de nombreux promoteurs eucaryotes (Shang *et al.*, 2000; Métivier *et al.*, 2003; Nagaich *et al.*, 2004; Karpova *et al.*, 2008) n'est possible qu'en la présence de réactions énergie-dépendantes.

Il n'y a cependant pas de relation directe entre consommation énergétique et oscillations car d'autres aspects de la structure du graphe d'état entrent en jeu. En effet, un système

[31] $\forall s, \sum_s M_{s',s} e^{\frac{-G_s}{\mathrm{k_B}T}} = \sum_{s\neq s'} \left[e^{-\frac{\hat{E}_{s',s}-G_s}{\mathrm{k_B}T}} e^{-\frac{G_s}{\mathrm{k_B}T}}\right] - \sum_{s''\neq s'} \left[e^{-\frac{\hat{E}_{s'',s'}-G_{s'}}{\mathrm{k_B}T}} e^{-\frac{G_{s'}}{\mathrm{k_B}T}}\right] = \sum_{s\neq s'} \left[e^{-\frac{\hat{E}_{s',s}}{\mathrm{k_B}T}} - e^{-\frac{\hat{E}_{s,s'}}{\mathrm{k_B}T}}\right] = 0$ et $M_{s',s} e^{\frac{-G_s}{\mathrm{k_B}T}} = e^{-\frac{\hat{E}_{s',s}-G_s}{\mathrm{k_B}T}} e^{-\frac{G_s}{\mathrm{k_B}T}} = e^{-\frac{\hat{E}_{s,s'}-G_{s'}}{\mathrm{k_B}T}} e^{-\frac{G_{s'}}{\mathrm{k_B}T}} = M_{s,s'} e^{\frac{-G_{s'}}{\mathrm{k_B}T}}$ donc $M_{s',s}\Lambda_{s,0} = M_{s,s'}\Lambda_{s',0}$.

[32] Par exemple, Mukherjee *et al.* (1994) montrent que l'équilibre détaillé implique que la matrice $(\mathbf{D}_{\Lambda_0})^{-1/2}\mathbf{M}(\mathbf{D}_{\Lambda_0})^{1/2}$, cospectrale à $\mathbf{M}$, est symétrique et n'a donc que des valeurs propres réelles.

ouvert peut très bien n'avoir que des valeurs propres réelles[33]. Il est donc capital, pour comprendre cette relation, d'être capable de quantifier la consommation énergétique d'un système donné. Mais, bien que la présence de cycles orientés atteste d'une consommation énergétique, il n'est pas possible, à partir de $\mathbf{M}$, de déterminer laquelle ou lesquelles des transitions consomme effectivement de l'énergie. En effet, un même système cinétique (décrit par $\mathbf{M}$) peut être obtenu pour différents systèmes énergétiques (décrits par $\mathbf{G}$ et $\mathbf{E}$). De plus, même si l'on connaît la quantité d'énergie extérieure éventuellement consommée par chacune des transitions, la consommation globale dépend de leur fréquence d'occurrence qui dépend elle-même de la dynamique globale du système. Il est pourtant possible de dériver un indicateur synthétique décrivant, pour un système donné, l'apport énergétique extérieur, en tenant compte sa dynamique globale, et ce uniquement à partir de $\mathbf{M}$. Il s'agit du taux de consommation énergétique moyen :

$$\dot{E} = \mathrm{k_B} T \sum_s \Lambda_{s,0} \sum_{s' \neq s} M_{s',s} \log \frac{M_{s',s}}{M_{s,s'}} \tag{III.43}$$

correspondant simplement à l'espérance, à l'état stationnaire, des variations d'énergie libre du système (chaque transition $\mathcal{G}_s \to \mathcal{G}_{s'}$ a lieu à un taux $M_{s',s}\Lambda_{s,0}$ et provoque une variation d'énergie libre $\mathrm{k_B} T \log M_{s',s}/M_{s,s'}$; équation (III.7)).

La consommation énergétique d'un promoteur pouvant être quantifiée de façon précise, il est alors possible d'étudier sa relation avec diverses propriétés dynamiques du système, et en particulier, avec son comportement périodique.

3.5 Systèmes idéalisés

Dans le cas général, la décomposition de la matrice $\mathbf{M}$ en valeurs propres et vecteurs propres doit être faite numériquement. Dans certains cas particuliers, il est cependant possible d'en obtenir une expression analytique. Celle-ci indique alors comment la positions des valeurs propres sur le plan complexe change en fonction des paramètres du système et permet donc de mieux comprendre le comportement de promoteur, même dans le cadre de systèmes non-idéalisés.

(i) Promoteur à deux états : Ce cas très simple ($N = 2$) correspond à la représentation du promoteur employé dans la plupart des modèles d'expression stochastique qui le considèrent explicitement (cf chapitre I, section 2.4). Les valeurs propres et vecteurs propres sont

$$\begin{cases} \lambda_0 = 0 \\ \lambda_1 = -k_{\mathrm{off}} - k_{\mathrm{on}} \end{cases} \quad \text{et} \quad \begin{cases} \Lambda_0 = [k_{\mathrm{off}}, k_{\mathrm{on}}]^{\mathrm{T}}/(k_{\mathrm{off}} + k_{\mathrm{on}}) \\ \Lambda_1 = [1, -1]^{\mathrm{T}} \end{cases} \tag{III.44}$$

Le spectre de $\mathbf{M}$ décrit donc, en plus de l'état stationnaire λ_0, un unique mode λ_1, apériodique et dont le temps de relaxation est $-1/\lambda_1 = 1/(k_{\mathrm{off}} + k_{\mathrm{on}})$.

[33] C'est le cas, par exemple, de la matrice $\mathbf{M} = [[-2, 1, 1, 0]^T, [1, -2, 0, 1]^T, [1, 0, -2, 1]^T, [0, 1, 2, -3]^T]$

(ii) Cycle orienté homogène : Nous considérons un promoteur dont le graphe de transition correspond[34] à un cycle de m états (notés $s_0, s_1 \dots s_p \dots s_{m-2}, s_{m-1}$ dans l'ordre rencontré le long du cycle) où tous les taux de transition dans le sens du cycle k^{f} sont égaux, ainsi que tous ceux des transitions retour k^{b} (figure III.13). Ce cas particulier correspond au modèle de dynamique du promoteur étudié par Lemaire *et al.* (2006) dans lequel tous les événements ont lieu selon une séquence ordonnée prédéfinie (cf section 1.4).

Sa décomposition propre est (Cvetković *et al.*, 1979) :

$$\begin{cases} \lambda_i = k^{\mathrm{f}}(e^{-\frac{2\pi j}{m}i} - 1) + k^{\mathrm{b}}(e^{\frac{2\pi j}{m}i} - 1) \\ \Lambda_{s_p,i} = e^{p\frac{2\pi j}{m}i}/m \end{cases} \tag{III.45}$$

($j = \sqrt{-1}$ est l'unité imaginaire). Le spectre de $-\mathbf{M}$ est donc constitué de m valeurs propres réparties régulièrement sur une ellipse de centre $k^{\mathrm{f}} + k^{\mathrm{b}}$ et tangente à l'axe des imaginaires en 0 (figure III.13a). Cette ellipse est un cercle si $k^{\mathrm{b}} = 0$ et s'aplatit sur l'axe des réels à mesure que k^{b} augmente jusqu'à donner des valeurs propres exclusivement réelles pour $k^{\mathrm{b}} = k^{\mathrm{f}}$ (figure III.13b).

Le mode le plus lent, λ_1, oscille à une fréquence $\frac{|\mathrm{Im}(\lambda_1)|}{2\pi} \underset{m\gg 1}{\simeq} \frac{k^{\mathrm{f}}-k^{\mathrm{b}}}{m}$. C'est pour ce mode que le facteur de cohérence $\kappa(\lambda_1)$ (nombre d'oscillations pendant la durée de relaxation ; équation (III.17)) est maximale. Ainsi, d'après[35] (III.45)

$$\kappa \underset{m\gg 1}{\lesssim} \frac{m}{2\pi^2}\,\frac{k^{\mathrm{f}} - k^{\mathrm{b}}}{k^{\mathrm{f}} + k^{\mathrm{b}}} \tag{III.46}$$

La cohérence maximale est donc toujours majorée par l'expression (III.46) et tend vers cette valeur dans la limite des cycles longs (en pratique, on s'approche très rapidement de la limite[36]). Comme l'illustrent les figures III.13a et III.13b, la cohérence des oscillations (pente de la droite verte) est d'autant plus importante que le cycle est long et orienté.

Si l'on note $\overset{\circlearrowright}{E}$ la quantité d'énergie consommée par cycle, valant

$$\overset{\circlearrowright}{E} = m\,\mathrm{k_B}T \log \frac{k^{\mathrm{f}}}{k^{\mathrm{b}}} \tag{III.47}$$

d'après l'équation (III.9), on peut alors reformuler l'équation (III.46) en

$$\kappa \underset{m\gg 1}{\lesssim} \frac{m}{2\pi^2}\,\tanh\left(\frac{\overset{\circlearrowright}{E}/\mathrm{k_B}T}{2m}\right) \tag{III.48}$$

Cette expression montre que, quelle que soit leur fréquence, la cohérence des oscillations d'un cycle orienté homogène peut être déterminée uniquement à partir

[34] S'il existe d'autres états dans le système mais que toutes les transitions sortant du cycle sont nulles, les valeurs et vecteurs propres (III.45) restent valables. Cela conduit simplement à l'apparition de valeurs et vecteurs propres supplémentaires décrivant la transition vers le cycle depuis les états hors du cycle.

[35] La cohérence de λ_1 est, d'après (III.45), $\kappa = \frac{1}{2\pi}\frac{k^{\mathrm{f}}-k^{\mathrm{b}}}{k^{\mathrm{f}}+k^{\mathrm{b}}}\cot\frac{\pi}{m}$, ce qui se simplifie en (III.46).

[36] L'erreur est déjà de moins de 5% pour $m > 8$ et de moins de 0.5% pour $m > 25$.

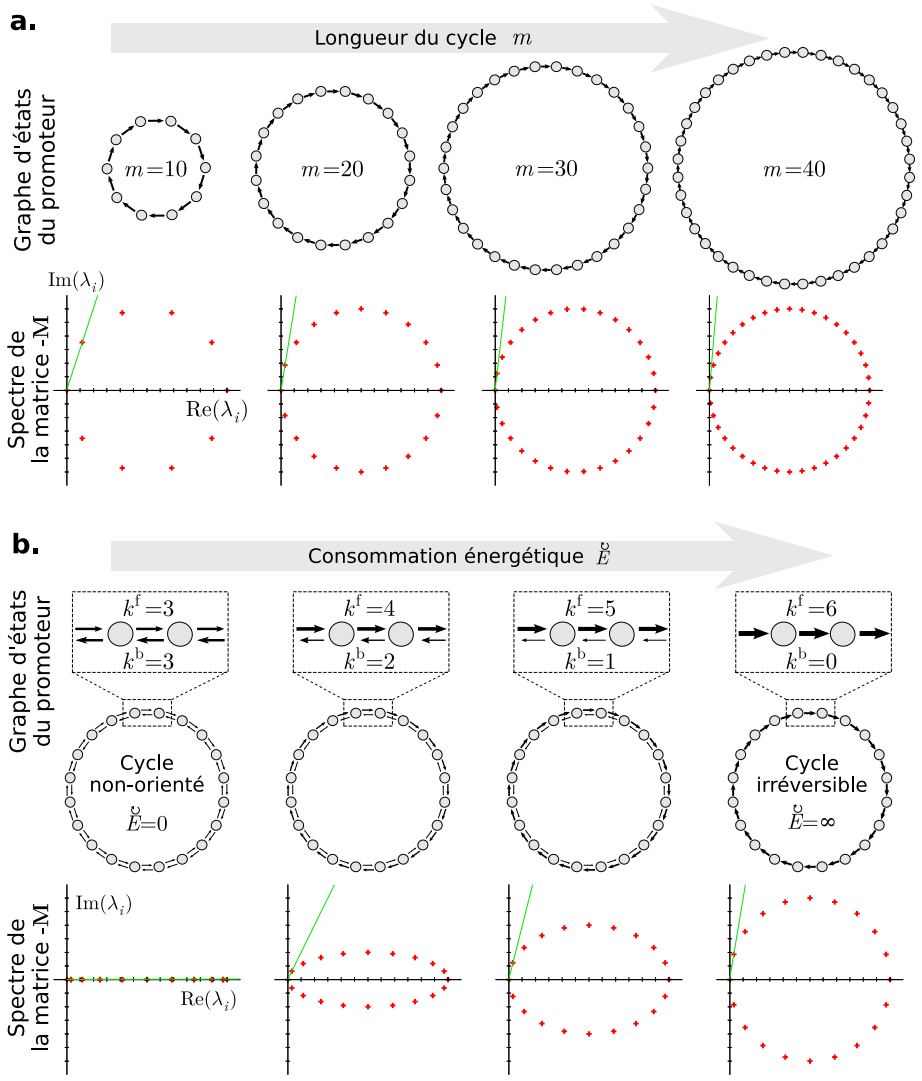


Fig. III.13: Spectre de M pour le cas du cycle orienté et homogène de transitions. Lorsque le graphe d'état du promoteur consiste en un cycle de transitions dont les taux dans un même sens sont égaux, les valeurs propres $-\lambda_i$ de la matrice $-\mathbf{M}$ sont placées sur le plan complexe de manière régulière sur une ellipse tangente en 0 à l'axe des imaginaires et ayant l'axe des réels pour grand axe. Pour un système donné, la cohérence maximale, notée κ, est proportionnelle à la pente la plus élevée des droites reliant les valeurs propres à l'origine du plan. Elle dépend alors à la fois du nombre m d'étapes du cycle et de la quantité d'énergie extérieure $\overset{\circlearrowleft}{E}$ reçue en un tour (cette dernière déterminant, pour une longueur de cycle m donnée, le rapport entre les taux de transition dans les deux sens k^{f} et k^{b} ; cf équation (III.47)). (a) Le nombre de valeurs propres correspondant au nombre d'états du graphe, la plus cohérente d'entre elles l'est d'autant plus que le cycle est long. La cohérence κ du système augmente donc avec le nombre m d'étapes du cycle. (b) La forme de l'ellipse sur laquelle se répartissent les valeurs propres dépend du rapport entre les taux de transition k^{f} et k^{b}. Plus le cycle est orienté (et donc plus la consommation énergétique $\overset{\circlearrowleft}{E}$ est élevée), plus l'ellipse se rapproche d'un cercle. Pour $\overset{\circlearrowleft}{E} = 0$, le cycle n'est pas orienté et l'ellipse est plate (toutes les valeurs propres sont réelles). À l'inverse, lorsque les transitions sont irréversibles ($k^{\mathrm{b}} = 0$), la consommation énergétique est infinie et l'ellipse est alors un cercle. Ainsi, la cohérence κ est d'autant plus élevée que la consommation d'énergie extérieure est importante.

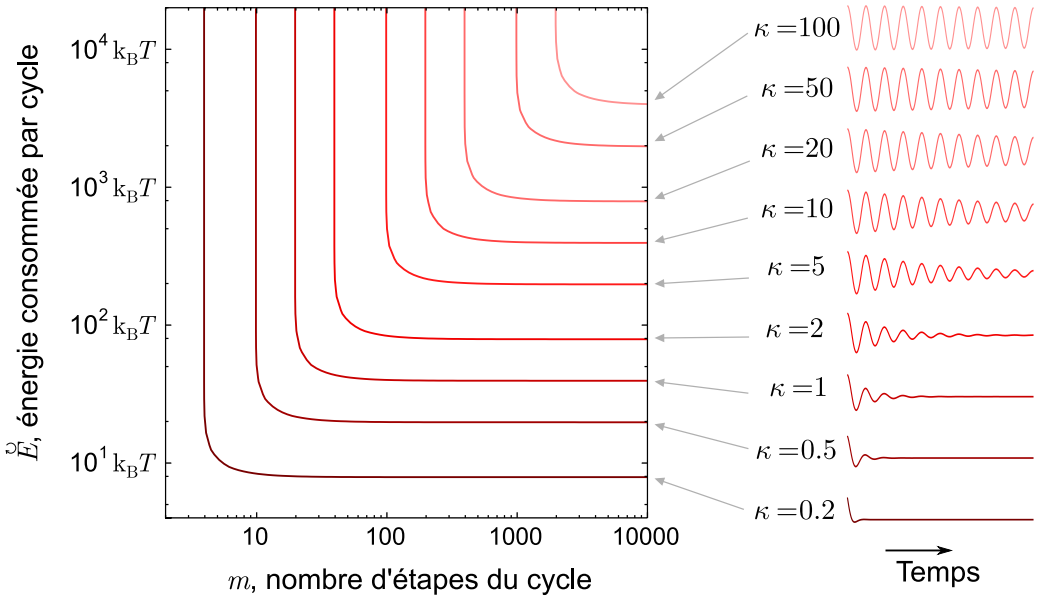


Fig. III.14: Relation entre cohérence, dépendance énergétique et longueur d'un cycle. L'expression (III.48), décrivant la cohérence κ des oscillations pour le cas idéalisé d'un cycle orienté et homogène de transitions où m représente la longueur du cycle et $\overset{\circlearrowright}{E}$ l'apport énergétique extérieur reçu par le système à chaque tour, permet de mettre en évidence certaines propriétés. Chaque courbe représente l'ensemble des combinaisons de m et de $\overset{\circlearrowright}{E}$ provoquant des oscillations de même cohérence[37]. Comme l'illustre intuitivement la figure III.13, on constate que la cohérence augmente avec la longueur du cycle et avec sa consommation énergétique. Cependant, ces courbes mettent également en évidences deux asymptotes (une horizontale et une verticale) indiquant qu'une cohérence donnée ne peut pas être atteinte en dessous d'une certaine longueur de cycle, quelle que soit la quantité d'énergie consommée, et *vice versa*. Par exemple, pour osciller avec une cohérence de $\kappa = 2$, le cycle ne peut pas avoir moins de $\sim$40 étapes et ne peut pas consommer moins de $\sim 80 k_BT$ par tour. Ces deux bornes peuvent être déterminées de façon analytique (équations (III.49) et (III.50)) et permettent, à partir d'une valeur de cohérence observée expérimentalement, de donner une borne inférieure à la longueur et à la consommation énergétique du cycle.

de la longueur et de la consommation énergétique du cycle. Chacune des courbes de la figure III.14 décrit l'ensemble des valeurs des paramètres m et $\overset{\circlearrowright}{E}$ conduisant à une même cohérence et correspondent simplement à des lignes de niveau[37] de l'équation (III.48). Les deux asymptotes de ces courbes révèlent une propriété particulière : une cohérence donné ne peut être atteinte que si le cycle est plus long qu'une certaine longueur et consomme plus qu'une certain quantité d'énergie. En d'autres termes, l'observation d'une cohérence donnée atteste d'une certaine longueur minimale et d'une certaine consommation énergétique minimale du cycle en dessous desquelles le système ne pourrait avoir un tel comportement. Ces deux mi-

[37]Notons que l'on peut donner une version paramétrique des lignes de niveau de l'équation (III.48) représentées sur la figure III.14 :

$$\frac{\overset{\circlearrowright}{E}/k_BT}{2\theta} = m = \frac{2\pi^2\kappa}{\tanh\theta}$$

avec $\theta > 0$. Les deux asymptotes correspondent à $\theta \to 0^+$ et à $\theta \to +\infty$, et le coude à $\theta = 1$.

norations correspondent simplement aux inégalités

$$m > 2\pi^2\kappa \tag{III.49}$$

$$\overset{\circ}{E}/\mathrm{k_B}T > 4\pi^2\kappa \tag{III.50}$$

obtenues, à partir de l'équation (III.48), pour $\overset{\circ}{E} \to \infty$ et $m \to \infty$ respectivement. Dans la section 4.4, nous montrons comment ces propriétés peuvent être généralisées aux cas de cycles présentant toutes sortes de déviations par rapport au cas idéalisé considéré ici (hétérogénéité des taux de transition, chemins alternatifs, ...) et peuvent être utilisées pour interpréter des données expérimentales telles que celles de Métivier *et al.* (2003) ou de Nagaich *et al.* (2004).

(iii) Sous-ensembles indépendants de facteurs : Si l'ensemble $\mathcal{F}$ des facteurs peut être partitionné en $\hat{n}$ sous-ensembles $\mathcal{F}^0 \cup \mathcal{F}^1 \cup \cdots \cup \mathcal{F}^{\hat{n}-1} = \mathcal{F}$ se comportant indépendamment les uns des autres (par exemple, des groupes de molécules ne formant aucun contact entre eux ; figure III.15a), alors la matrice $\mathbf{M}$ comporte beaucoup de redondances et peut être déterminée à partir de matrices $\mathbf{M}^{[p]}$, beaucoup plus petites, décrivant chacune la dynamique d'un sous-système. En effet, les facteurs de deux sous-ensembles différents n'interagissant pas entre eux, ils ne peuvent pas s'associer ou se dissocier simultanément (sous forme de complexe) et les transitions affectant les facteurs d'un sous-ensemble sont indépendantes de l'état des facteurs des autres sous-ensembles[38]. En termes de théorie des graphes, le graphe d'état du promoteur est alors le *produit cartésien* des graphes de chaque sous-système (figure III.15b) et le spectre de $\mathbf{M}$ peut être déterminé simplement à partir de ceux des différents sous-systèmes. Chaque valeur propre λ_i de $\mathbf{M}$ et son vecteur propre associé Λ_i sont déterminés par une unique combinaison $\{\lambda^{[0]}, \lambda^{[1]}, \ldots, \lambda^{[\hat{n}-1]}\}$ des valeurs propres des $\hat{n}$ sous-systèmes (une par sous-système) avec leurs vecteurs propres associés $\{\Lambda^{[0]}, \Lambda^{[1]}, \ldots, \Lambda^{[\hat{n}-1]}\}$ selon

$$\begin{cases} \lambda_i = \sum_{p=0}^{\hat{n}-1} \lambda^{[p]} \\ \Lambda_{s,i} = \prod_{p=0}^{\hat{n}-1} \Lambda^{[p]}_{s\cap\mathcal{F}^p} \end{cases} \tag{III.51}$$

Ainsi, le spectre de $\mathbf{M}$ est simplement la *convolution* des spectres de chaque sous-système (figure III.15c).

Toutes les dérivations que nous avons présentées jusqu'ici ne décrivent que la dynamique du promoteur. Le modèle introduit à la section 2 incluant cependant les processus de transcription et de traduction, nous pouvons maintenant décrire comment les événements stochastiques au niveau du promoteur d'un gène influencent son niveau d'expression.

[38] Plus formellement $\forall(s,s'), s \neq s'$, si $\exists p$ tel que $s \ominus s' \subseteq \mathcal{F}^p$ alors $M_{s',s} = M^{[p]}_{s'\cap\mathcal{F}^p, s\cap\mathcal{F}^p}$, sinon $M_{s',s} = 0$.

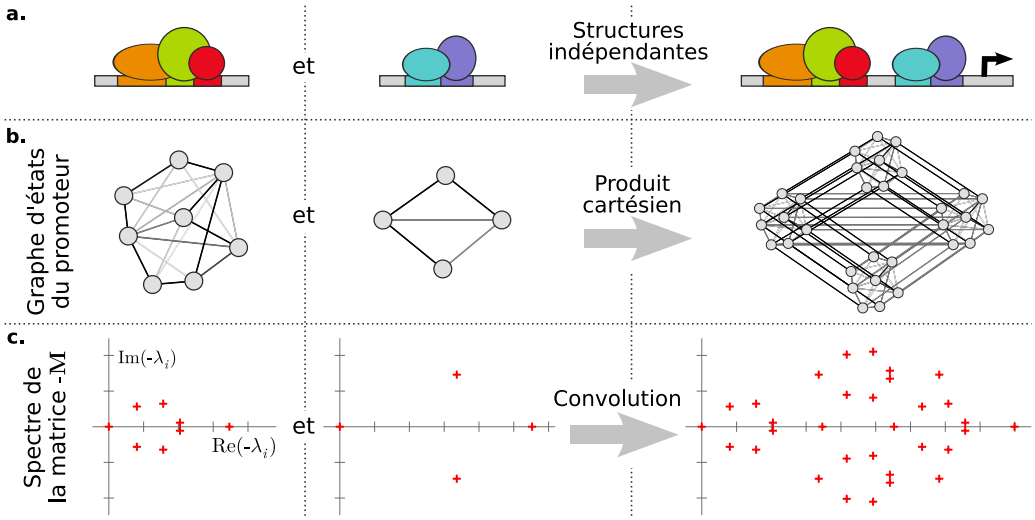


Fig. III.15: Graphe d'état et spectre de M pour deux structures indépendantes. (a) Lorsque le système est constitué de deux sous-ensembles de facteurs n'interagissant pas entre eux, le graphe d'état du promoteur, ainsi que son spectre, peuvent être déterminés simplement à partir ceux de ces deux structures prises séparément. (b) Le graphe d'état du système complet est alors le *produit cartésien* des graphes représentant chaque sous-système : chaque état de l'un de ces deux graphes est répété pour chacun des états de l'autre. (c) De plus, le spectre de la matrice $-\mathbf{M}$ qui décrit le système constitué des deux structures indépendantes s'obtient par la *convolution* des deux spectres correspondant à chacune de ces structures : chacun des deux spectres est reproduit à l'emplacement de chaque valeur propre de l'autre.

3.6 Transmission aux niveaux d'ARN et de protéines

Dans la section précédente, nous avons décrit la dynamique d'un promoteur arbitraire par divers indicateurs et interprété leur structure commune en termes de modes propres. On est ainsi en mesure de comprendre le lien entre des mesures expérimentales utilisées pour l'étude de la dynamique du promoteur, telles que le ChIP et le FRAP, et des indicateurs théoriques, tels que la corrélation croisée ou le spectre de puissances. C'est en particulier ce dernier qui va maintenant nous permettre de déterminer comment la dynamique stochastique du promoteur est transmise aux étapes successives de l'expression génique.

Nous décrivons d'abord de façon générale comment un processus de synthèse/dégradation est capable de transmettre les fluctuations de son taux de synthèse. En appliquant ce principe à notre modèle d'expression génique, nous déterminons les expressions analytiques des spectres de puissance et les variances normalisées des niveaux d'ARN et de protéines. L'analyse de la structure de ces équations nous conduit alors à proposer une généralisation du modèle permettant de considérer des durées de vie ainsi que des délais dans la synthèse des transcrits et des protéines de distributions arbitraires. Cela permet de rendre compte de très nombreux phénomènes – entre autres en relation avec notre modèle spatial présenté au chapitre II – et laisse entrevoir un certain nombre de propriétés. Nous terminons alors cette section en présentant une méthode numérique permettant de déterminer (dans le cas d'un modèle sans délai) la distribution stationnaire complète du niveau d'ARN pour

un promoteur quelconque.

Processus de synthèse/dégradation non-homogènes

Remarquons d'abord une propriété générale sur la quantité $B(t)$ de molécules $\mathcal{B}$ dont la synthèse est réalisée par l'intermédiaire d'une molécule $\mathcal{A}$, en quantité $A(t)$, selon

$$\begin{cases} \mathcal{A} & \xrightarrow{\hat{\rho}} & \mathcal{A}+\mathcal{B} \\ \mathcal{B} & \xrightarrow{\hat{\gamma}} & \varnothing \end{cases} \tag{III.52}$$

Ces deux processus sont notés $B^+(t)$ et $B^-(t)$ et correspondent à des sommes de Diracs (un à chaque instant où une réaction se produit). Ainsi,

$$\frac{\mathrm{d}B(t)}{\mathrm{d}t} = B^+(t) - B^-(t) \tag{III.53}$$

$B^+(t)$ et $B^-(t)$ sont des processus de Poisson ponctuels non-homogènes avec respectivement un taux instantané $\hat{\rho}A(t)$ et $\hat{\gamma}B(t)$. Leurs spectres de puissance sont donc[39]

$$\begin{cases} S_{B^+}(\omega) = \hat{\rho}^2 S_A(\omega) + \hat{\rho}\langle A\rangle \\ S_{B^-}(\omega) = \hat{\gamma}^2 S_B(\omega) + \hat{\gamma}\langle B\rangle \end{cases} \tag{III.54}$$

Il s'agit du spectre de puissance du taux instantané (par exemple $S_{\hat{\rho}A(t)}(\omega) = \hat{\rho}^2 S_A(\omega)$) auquel s'ajoute un bruit blanc dont la densité spectrale de puissance est égale au taux moyen de la réaction (par exemple $\langle B^+\rangle = \hat{\rho}\langle A\rangle$). Ce bruit est connu sous le nom de *bruit de grenaille* (ou *shot noise* en anglais) en raison de l'analogie avec l'arrivée stochastique des photons sur un dispositif optique ou des porteurs de charge dans un dispositif électronique. Augmentant comme $\hat{\rho}$, il masque d'autant plus le signal (qui augmente comme $\hat{\rho}^2$) que les événements sont rares ($\hat{\rho}$ faible).

Du fait de dépendances entre les processus $B^+(t)$ et $B^-(t)$, le spectre de puissance de (III.53) s'écrit $\omega^2 S_B(\omega) = \hat{\rho}^2 S_A(\omega) + \hat{\rho}\langle A\rangle - \hat{\gamma}^2 S_B(\omega) + \hat{\gamma}\langle B\rangle$ et, comme, à l'équilibre, $\hat{\rho}\langle A\rangle = \hat{\gamma}\langle B\rangle$, on obtient

$$S_B(\omega) = \overbrace{\underbrace{2\hat{\gamma}\langle B\rangle}_{\text{Shot noise}} \underbrace{(\hat{\gamma}^2+\omega^2)^{-1}}_{\text{Filtre passe-bas}}}^{B\text{-intrinsèque}} + \overbrace{\underbrace{\hat{\rho}^2 S_A(\omega)}_{\text{Fluctuation de } A} \underbrace{(\hat{\gamma}^2+\omega^2)^{-1}}_{\text{Filtre passe-bas}}}^{B\text{-extrinsèque}} \tag{III.55}$$

On retrouve alors une expression similaire à celles obtenues dans d'autres études (Warren *et al.*, 2006; Cox *et al.*, 2006), mais valable ici pour un signal $A(t)$ arbitraire. Elle permet de séparer les fluctuations provenant du système (B-intrinsèques, dues aux événements stochastiques de synthèse/dégradation de $\mathcal{B}$) de celles extérieures au système (B-extrinsèques, reflétant les variations de la quantité de $\mathcal{A}$).

[39] Pour $B^+(t)$ par exemple, pendant une unité de temps, on a en moyenne $\hat{\rho}\langle A\rangle$ événements (ce qui équivaut, du point de vue de l'auto-corrélation, à $\hat{\rho}\langle A\rangle$ paires d'événements avec un délai nul) et le nombre moyen d'événements séparés par un délai dans l'intervalle $[\tau;\tau+\mathrm{d}\tau]$ est $\hat{\rho}^2\langle A(t)A(t+\tau)\rangle_t \mathrm{d}\tau$. Donc l'auto-corrélation est $\langle B^+(t)B^+(t+\tau)\rangle_t = \hat{\rho}\langle A\rangle\delta(\tau) + \hat{\rho}^2\langle A(t)A(t+\tau)\rangle_t$ et sa transformée de Fourier donne l'équation (III.54).

Spectres de puissance et variances normalisées

Dans le modèle d'expression génique décrit au début de ce chapitre et illustré sur la figure III.16a, les variations de la quantité d'ARN et de protéines sont décrites par des processus simples de synthèse/dégradation, comme dans la plupart des modèles d'expression stochastique (section 2.1, chapitre I). L'efficacité transcriptionnelle du promoteur $X(t)$, dont nous avons obtenu une expression du spectre de puissance $S_X(\omega)$ (cf équation (III.26)), correspond au taux de synthèse des transcrits. Ainsi, l'application du principe énoncé précédemment aux deux processus de synthèse/dégradation $R(t)$ et $P(t)$ permet d'obtenir leurs spectres de puissance respectifs :

$$S_R(\omega) = \left[2\gamma\langle R\rangle + \sum_i \frac{-2\lambda_i {\beta_i}^2}{{\lambda_i}^2 + \omega^2}\right](\gamma^2 + \omega^2)^{-1} \tag{III.56}$$

$$S_P(\omega) = \left[\overbrace{2\tilde{\gamma}\langle P\rangle}^{\substack{\text{Shot noise}\\ \text{protéique}}} + \tilde{\rho}^2\left[\overbrace{2\gamma\langle R\rangle}^{\substack{\text{Shot noise}\\ \text{ARN}}} + \overbrace{\sum_i \frac{-2\lambda_i {\beta_i}^2}{{\lambda_i}^2 + \omega^2}}^{\text{Dynamique du promoteur}}\right]\overbrace{(\gamma^2 + \omega^2)^{-1}}^{\substack{\text{Filtre passe-bas}\\ \text{Promoteur}\to\text{ARN}}}\right]\overbrace{(\tilde{\gamma}^2 + \omega^2)^{-1}}^{\substack{\text{Filtre passe-bas}\\ \text{ARN}\to\text{protéines}}} \tag{III.57}$$

avec $\langle P\rangle = \tilde{\rho}\langle R\rangle/\tilde{\gamma}$ et $\langle R\rangle = \beta_0/\gamma$.

Les fluctuations du niveau d'ARN (resp. de protéines) correspondent ainsi simplement à celles du promoteur (resp. du niveau d'ARN) auxquelles viennent s'ajouter un shot noise, reflétant le caractère discret et plus ou moins rare des événements de synthèse/dégradation, puis en appliquant un filtre passe-bas, atténuant d'autant plus les fluctuations que celles-ci sont rapides par rapport à la durée de vie des molécules (cf figure III.16). La transmission de la dynamique du promoteur dépend donc à la fois de l'abondance moyenne des transcrits et des protéines et de leur durée de vie moyenne.

Mesurées chez *Saccharomyces cerevisiae* par des techniques haut-débit à l'échelle du transcriptome (Wang *et al.*, 2002) et du protéome (Belle *et al.*, 2006), ces durées de vie moyennes varient fortement d'un gène à l'autre : entre 3 min et 90 min pour la demi-vie des ARN et de moins de 4 min à plus d'un jour pour celle des protéines. On remarque alors, au regard de la période typique des oscillations observées sur les promoteurs eucaryotes (autour de 40∼60 min ; cf section 5.1, chapitre I), que ces durées de vie peuvent en théorie aussi bien masquer l'activité périodique du promoteur (durées de vie longues) que la laisser se propager jusqu'au niveau d'expression (durées de vie courtes). De fait, certaines des études mettant en évidence une activité périodique sur des promoteurs eucaryotes ont également observé ce comportement sur le niveau d'ARN du gène (Karpova *et al.*, 2008; Degenhardt *et al.*, 2009). Dans l'exemple de la figure III.16b, les durées de vie des transcrits et des protéines ont été choisies plutôt courtes (resp. $1/\gamma = 10$ min et $1/\tilde{\gamma} = 20$ min) et les oscillations du promoteur (dont la période la plus lente est ici de 40 min) sont transmises aux deux niveaux (bien que le niveau protéique présente également d'importantes fluctuations basses-fréquences apériodiques).

Une autre étude globale, également effectuée chez la levure, révèle que la quantité moyenne de transcrits peut varier de moins d'une copie par cellule pour certain gènes à une centaine de copies pour d'autres, et que la quantité de protéines est typiquement 1000 fois plus

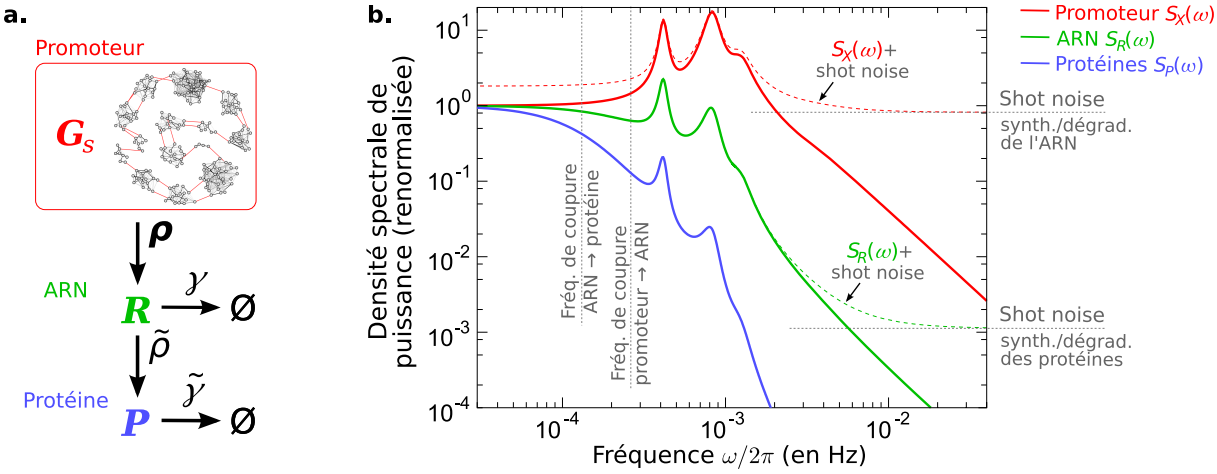


Fig. III.16: Propagation la stochasticité du promoteur jusqu'au niveau d'expression. (a) Bien qu'ayant adopté une description précise du promoteur, le modèle d'expression génique que nous présentons considère les étapes ultérieures du processus d'expression de façon très simple et comme la plupart des modèles d'expression stochastique (cf section 2.1 et section 2.1, chapitre I). Les molécules d'ARN et les protéines sont produites de manière stochastique à un taux dépendant respectivement de l'état courant du promoteur ($X(t)$) et de la quantité de transcrits ($\tilde{\rho}R(t)$). Chaque molécule a, à tout instant, une certaine probabilité d'être dégradée (décrite par un taux de dégradation γ pour les transcrits et $\tilde{\gamma}$ pour les protéines). (b) Dans ce cadre, on peut déterminer comment les fluctuations du promoteur, décrites par le spectre de puissance $S_X(\omega)$ de l'efficacité transcriptionnelle $X(t)$, sont transmises aux niveaux d'ARN et de protéines (cf équation (III.57)). Le spectre $S_R(\omega)$ du niveau d'ARN est obtenu par l'ajout à $S_X(\omega)$ d'un *shot noise*, puis par l'application d'un filtre passe-bas. Le shot noise correspond à une constante (ligne horizontale pointillée) d'autant plus élevée par rapport au signal $S_X(\omega)$ que le niveau moyen d'ARN est faible, masquant alors d'autant plus les détails de $S_X(\omega)$ de faible puissance. Le filtre passe-bas laisse passer les fluctuations sensiblement plus lentes que la *fréquence de coupure* $\omega = \gamma$ (déterminée par le taux de dégradation des ARN) et atténue d'autant plus les fluctuations que celles-ci sont rapides (à $\omega = \gamma$, l'atténuation est de $1/2$). Le spectre $S_P(\omega)$ du niveau d'expression s'obtient alors exactement de la même façon, à partir de celui du niveau d'ARN $S_R(\omega)$. Dans l'exemple présenté ici, la dynamique du promoteur est celle de l'exemple eucaryote présenté précédemment (figure III.11b3). Les durées de vie moyennes des transcrits et des protéines ($1/\gamma = 10$ min et $1/\tilde{\gamma} = 20$ min) ainsi que leur quantités moyennes ($\langle R \rangle = 30$ et $\langle P \rangle = 1000\,\langle R \rangle$) ont été choisies dans des plages de valeurs biologiquement réalistes (cf texte).

élevée que celle des transcrits (Lu *et al.*, 2007). Dans la figure III.16b, où les quantités moyennes de transcrits et de protéines sont $\langle R \rangle = 30$ et $\langle P \rangle = 1000\,\langle R \rangle$, le shot noise sur l'ARN reste moins important que les oscillations du promoteur. Cependant, plus le niveau moyen d'ARN est faible, plus le shot noise devient important face aux fluctuations du promoteur, conduisant alors au régime où le bruit de synthèse/dégradation de l'ARN domine.

On peut déduire, à partir des spectres de puissance $S_X(\omega)$, $S_R(\omega)$ et $S_P(\omega)$, les variances normalisées des processus $X(t)$, $R(t)$ et $P(t)$. En effet, la moyenne et la variance d'un

signal étant contenues dans son spectre de puissance[40], on obtient :

$$\frac{{\sigma_X}^2}{\langle X\rangle^2} = \sum_{i\neq 0} \frac{{\beta_i}^2}{{\beta_0}^2} \tag{III.58}$$

$$\frac{{\sigma_R}^2}{\langle R\rangle^2} = \frac{1}{\langle R\rangle} + \sum_{i\neq 0} \left[\frac{{\beta_i}^2}{{\beta_0}^2}\frac{\gamma}{\gamma-\lambda_i}\right] \tag{III.59}$$

$$\frac{{\sigma_P}^2}{\langle P\rangle^2} = \overbrace{\underbrace{\frac{1}{\langle P\rangle}}_{\text{Poisson}}}^{\text{Faible quantité de protéines}} + \overbrace{\underbrace{\frac{1}{\langle R\rangle}}_{\text{Poisson}}\ \underbrace{\frac{\tilde{\gamma}}{\gamma+\tilde{\gamma}}}_{\text{Lissage ARN}\to\text{prot.}}}^{\text{Faible quantité d'ARN}} + \overbrace{\sum_{i\neq 0}\left[\underbrace{\frac{{\beta_i}^2}{{\beta_0}^2}}_{\text{Modes propres}}\ \underbrace{\frac{\gamma}{\gamma-\lambda_i}\frac{\tilde{\gamma}}{\gamma+\tilde{\gamma}}\left(1+\frac{\gamma}{\tilde{\gamma}-\lambda_i}\right)}_{\text{Lissage promoteur}\to\text{ARN}\to\text{prot.}}\right]}^{\text{Dynamique du promoteur}} \tag{III.60}$$

On retrouve ici une expression similaire à celle que nous avions évoquée au chapitre I, section 2.4, permettant de séparer les différentes sources de stochasticité. On reconnaît en particulier les signatures en $1/\langle P\rangle$ caractéristiques des distributions de Poisson et les coefficients de lissage temporel. Ils proviennent respectivement des termes de shot noise et des filtres passe-bas mis en évidence dans le spectre de puissance (III.57). Cependant, le terme représentant la dynamique du promoteur est ici beaucoup plus riche que celui décrivant un promoteur à deux états[41].

Délais distribués : élongation, maturation, export, traduction et dégradations

Le modèle d'expression génique employé ici néglige un certain nombre de phénomènes pouvant avoir une influence sur les propriétés de transmission de la dynamique du promoteur. En effet, il considère comme instantané le délai entre l'initiation de la transcription et l'apparition dans le cytoplasme d'un ARN messager près à être traduit, de même que le délai entre l'initiation de la traduction et l'apparition d'une protéine entière et fonctionnelle. Or, ces délais (que nous appellerons par simplicité *délais de synthèse*) sont non seulement non nuls, mais sont en réalité également distribués (deux transcrits ne mettent pas nécessairement le même délai). De nombreux phénomènes de natures variées peuvent influencer ces délais et en particulier leur distribution. Par exemple, le délai de synthèse d'un transcrit dépend de sa dynamique d'élongation, de sa maturation, de sa diffusion au sein de l'espace nucléaire et de son expulsion hors du noyau par un pore nucléaire (cf sections 2.6 et 4.4 du chapitre I).

Une autre supposition forte du modèle est de considérer la dégradation des ARN et des protéines comme de simples réactions de premier ordre, impliquant donc des distributions exponentielles de la durée de vie de ces molécules. En effet, les mécanismes de dégradation de ces molécules font intervenir une machinerie moléculaire impliquant de nombreuses réactions, rendant alors non-exponentielles ces durées de vie.

[40] Pour tout signal $a(t)$, on a $\langle a\rangle^2 = \frac{1}{2\pi}\int_{0^-}^{0^+} S_a(\omega)\mathrm{d}\omega$ et $\langle a^2\rangle = \frac{1}{2\pi}\int_{-\infty}^{+\infty} S_a(\omega)\mathrm{d}\omega$.

[41] Par ailleurs, l'instanciation de cette expression avec le système idéalisé à deux états (section 3.5) reproduit l'expression du chapitre I, section 2.4 pour laquelle l'unique mode non-nul correspond au terme

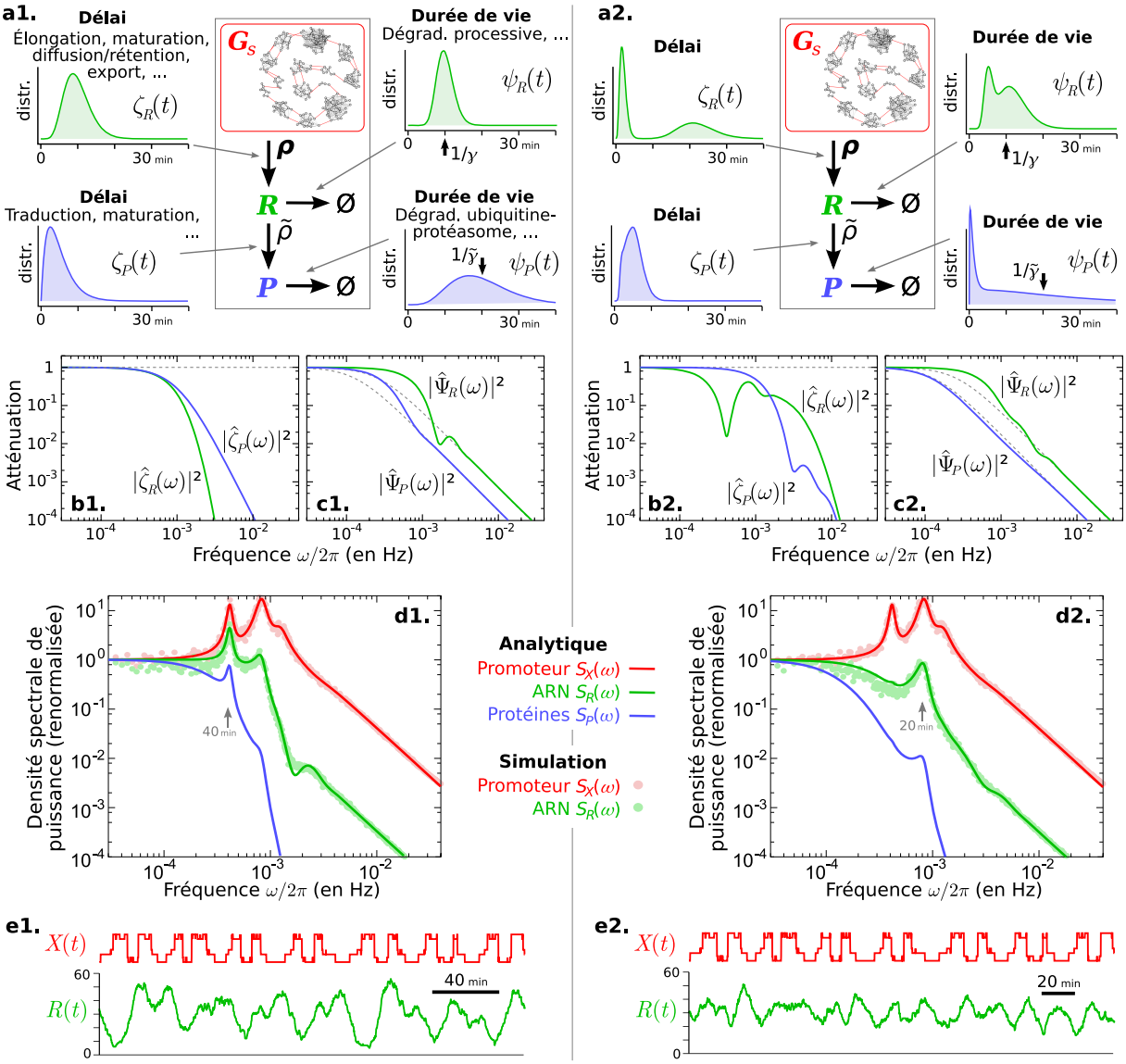


Fig. III.17: Délais de synthèse distribués et durées de vie non-exponentielles. La description classique de la synthèse et de la dégradation des ARN et des protéines (cf figure III.16a et section 2.1, chapitre I) suppose implicitement que ces molécules sont produites instantanément à l'instant de l'initiation de la transcription ou de la traduction et que leur durée de vie a une distribution exponentielle. Cette figure illustre comment des délais distribués dans la synthèse de ces molécules et des durées de vie non-exponentielles modifient les propriétés de transmission des fluctuations. (a1-2) Des phénomènes très divers peuvent être représentés par les quatre distributions décrites ici. Par exemple, la distribution du délai de synthèse d'un transcrit $\zeta_R(t)$ prend en compte des mécanismes tels que l'élongation de la transcription, la maturation de l'ARN (épissage, ...), son déplacement jusqu'au pore nucléaire et son export hors du noyau. Ou encore, la distribution de la durée de vie d'une protéine $\psi_P(t)$ peut refléter sa dégradation par un mécanisme de poly-ubiquitination et de reconnaissance par le protéasome. (b1-2) En comparaison avec le cas simple de la figure III.16, les délais de synthèse des transcrits $\zeta_R(t)$ et des protéines $\zeta_P(t)$ provoquent l'apparition de deux filtres supplémentaires, $|\hat{\zeta}_R(\omega)|^2$ et $|\hat{\zeta}_P(\omega)|^2$, tels que décrit par l'équation (III.61). (c1-2) De même, les filtres passe-bas décrits précédemment dans l'équation (III.57) dépendent des distributions de durée de vie des transcrits $\psi_R(t)$ et des protéines $\psi_P(t)$. Des distributions non-exponentielles provoquent alors pour ces filtres (notés $|\hat{\Psi}_R(\omega)|^2$ et $|\hat{\Psi}_P(\omega)|^2$ dans ce cadre plus général) des profils spécifiques

(Fig. III.17, cont. : cf page précédente)
démontrant des propriétés de filtrage différentes du cas exponentiel (lignes pointillées). (d1-2) Les spectres de puissance (dont la généralisation sous la forme de l'équation (III.61) est appuyée par les résultats de simulation) montrent alors que les délais de synthèse et les durées de vie des transcrits et des protéines permettent de contôler finement la transmission de la dynamique du promoteur vers les étapes successives du processus d'expression génique. Dans le cas (d1), la forme des filtres est telle que la périodicité lente de 40 min est beaucoup mieux transmise que celle à 20 min. Dans le cas (d2), la forme particulière du filtre $|\zeta_R(\omega)|^2$ (due au délai bimodal de synthèse des transcrits) provoque la disparition de la périodicité la plus lente au profit de la seconde. (e1-2) On observe en effet ce phénomène sur les fluctuations du niveau $R(t)$ de transcrits obtenu par simulation. Les deux cas présentés ici sont obtenus à partir de la même trajectoire $X(t)$ du promoteur, et en simulant le processus de synthèse/dégradation de l'ARN avec des distributions de synthèse $\zeta_R(t)$ et de durée de vie $\psi_R(t)$ différentes (a1 et a2).

Nous généralisons ici le résultat précédent en prenant en compte des distributions arbitraires pour les délais de synthèse des transcrits et de protéines, notées $\zeta_R(t)$ et $\zeta_P(t)$, et pour les durées de vies de ces molécules, notées $\psi_R(t)$ et $\psi_P(t)$, comme illustré sur la figure III.17a1-2. On note $\Psi_R(t)$ et $\Psi_P(t)$ les fonctions de répartition complémentaires de ces durées de vie, décrivant la probabilité qu'une molécule n'ait pas encore été dégradée au bout d'un temps t après sa synthèse[42].

Dans notre cas (où l'on suppose en particulier une absence de rétrocontrôles), l'application à la synthèse des transcrits d'un délai distribué revient simplement à utiliser comme taux de synthèse, non pas directement le processus $X(t)$ mais sa convolution avec la distribution $\zeta_R(t)$. Cela provoque, lors du passage à une description spectrale, un filtrage de $S_X(\omega)$ par $|\hat{\zeta}_R(\omega)|^2$, le carré du module de la transformée de Fourier de la distribution $\zeta_R(t)$. De même, le délai de synthèse protéique $\zeta_P(t)$ provoque un filtrage de $S_R(\omega)$ par $|\hat{\zeta}_P(\omega)|^2$. De plus, on peut interpréter les termes $2\gamma\langle R\rangle$ et $(\gamma^2+\omega^2)^{-1}$ de l'équation (III.57) comme représentant respectivement les événements discrets et stochastiques de synthèse et de dégradation (arrivant à un taux $2\gamma\langle R\rangle$) et comme la convolution du signal résultant par le profil moyen $\Psi_R(t)$. En effet, cette convolution revient à faire un filtrage par $|\hat{\Psi}_R(\omega)|^2$, correspondant simplement à $(\gamma^2+\omega^2)^{-1}$ dans le cas où $\psi_R(t)$ est une distribution exponentielle.

Ainsi, même si cette expression nécessite une preuve formelle que nous ne fournirons pas ici, on peut conjecturer que l'équation (III.57) se généralise en :

$$S_P(\omega) = \left[\overbrace{2\tilde{\gamma}\langle P\rangle}^{\substack{\text{Shot noise}\\ \text{protéique}}} + \overbrace{\left|\hat{\zeta}_P(\omega)\right|^2}^{\substack{\text{Traduction,}\\ \cdots}} \tilde{\rho}^2 \underbrace{\left[\overbrace{2\gamma\langle R\rangle}^{\substack{\text{Shot noise}\\ \text{ARN}}} + \overbrace{\left|\hat{\zeta}_R(\omega)\right|^2}^{\substack{\text{Élongation,}\\ \text{export, }\ldots}} \overbrace{S_X(\omega)}^{\substack{\text{Dynamique}\\ \text{du}\\ \text{promoteur}}} \right] \overbrace{\left|\hat{\Psi}_R(\omega)\right|^2}^{\substack{\text{Durée de vie}\\ \text{des ARN}}}}_{= S_R(\omega)} \right] \overbrace{\left|\hat{\Psi}_P(\omega)\right|^2}^{\substack{\text{Durée de vie}\\ \text{des protéines}}} \tag{III.61}$$

Comme le montre la figure III.17, cette expression est fortement appuyée par des résultats de simulations réalisées avec des distributions très différentes.

---

représentant la variance normalisée d'une loi de Bernoulli.

[42] $\Psi_R(t) = 1 - \int_0^t \psi_R(t')dt'$ et $\Psi_P(t) = 1 - \int_0^t \psi_P(t')dt'$.

La forme de ces quatre filtres est donc déterminée directement par l'allure des quatre distributions (cf figures III.17a, b et c). Cela permet alors un certain contrôle de la transmission de la dynamique du promoteur au niveau d'ARN et de protéine, par exemple en faisant ressortir ou disparaître certaines périodicités (figures III.17d1 et d2). Ces propriétés sont donc directement liées aux mécanismes sous-jacents pouvant provoquer diverses caractéristiques de ces distributions. Par exemple, plus la dégradation d'une molécule passe par de nombreuses réactions de durée moyenne similaire, plus sa durée de vie s'éloigne d'une distribution exponentielle. Ainsi, les molécules d'ARN, dont la dégradation commence par la destruction progressive de la queue polyA, peuvent avoir une durée de vie beaucoup moins variable qu'une distribution exponentielle[43]. Comme l'illustre la figure III.17a1 (où cette distribution est une loi gamma, représentant l'équivalent de 20 réactions successives de même taux), cela implique que le filtre passe-bas dû au processus de synthèse/dégradation des transcrits laisse passer les fluctuations jusqu'à des fréquences sensiblement plus rapides que dans le cas d'une durée de vie exponentielle de même moyenne (figure III.17c1).

Les distributions, utilisées ici comme paramètres du modèle, peuvent être le résultat d'autres études, telles que des modélisations de l'élongation (Roussel et Zhu, 2006; Voliotis *et al.*, 2008) par exemple, et/ou de notre modèle spatial, décrit au chapitre II. On peut en effet imaginer différents mécanismes susceptibles de provoquer la distribution assez particulière du délai $\zeta_R(t)$ présentée sur la figure III.17a2 et responsable de la disparition de la périodicité la plus lente du promoteur (figures III.17b2 et d2). On peut par exemple l'obtenir si, lorsqu'une molécule d'ARN est synthétisée, celle-ci peut soit être rapidement exportée hors du noyau (premier mode), soit être capturée par une structure nucléaire particulière (par exemple, un corps nucléaire ayant des propriétés de rétention des transcrits ; cf section 4.4, chapitre I) et relâchée progressivement (second mode). Un autre mécanisme simple pouvant conduire à une telle distribution peut également être envisagé si la transcription a lieu au voisinage d'un pore nucléaire : une fraction significative des transcrits diffusent jusqu'au pore et sont expulsés rapidement (premier mode), alors que les autres, diffusant dans l'espace nucléaire, peuvent sortir par n'importe quel pore et après un temps plus long et plus variable (second mode).

L'effet de ces mécanismes hypothétiques simples peut typiquement être testé à l'aide de notre modèle spatial présenté au chapitre II, celui-ci nous fournissant alors l'allure précise de cette distribution en fonction des différents paramètres (coefficients de diffusion, distances, facteur de rétention, ...).

Notons enfin que ces dérivations s'appliquent quelle que soit l'origine des fluctuations de $S_X(\omega)$. Ainsi, les propriétés de contrôle fin de la transmission des fluctuations par filtrages présentées ici s'appliquent également à la transmission des fluctuations gène-extrinsèques provenant des autres gènes du réseau.

[43] On connaît par ailleurs plusieurs mécanismes moléculaires responsables de la dégradation de la queue polyA (Xu *et al.*, 1997). Ceux ci influencent probablement la distribution de durée de vie des transcrits.

Distribution stationnaire du niveau d'ARN

La description des fluctuations des niveaux d'ARN ou de protéines par leur spectre de puissance rend compte de certains aspects dynamiques du signal. Mais n'étant qu'une statistique d'ordre 2, cet indicateur ne révèle pas certains détails relatifs aux ordres supérieurs (cf Box 1, page 22). À l'inverse, la distribution stationnaire des valeurs du signal – correspondant à la distribution sur une population de cellules telle qu'on peut l'observer par des techniques de cytométrie de flux – est un indicateur ne rendant pas compte des aspects dynamiques mais qui révèle des détails fins tels que la multimodalité et la position/taille/forme des différents modes. Sánchez et Kondev (2008) ont proposé une méthode pour déterminer les différents moments de la distribution stationnaire du niveau d'ARN sur un système similaire au nôtre, en se basant sur la résolution numérique de systèmes linéaires. Mais bien qu'il y ait théoriquement équivalence entre la distribution et l'ensemble (infini) des moments[44], ces derniers ne permettent pas une interprétation aussi intuitive que le permet la distribution. Dans le cas simple d'un promoteur à deux états, Innocentini et Hornos (2007) ainsi que Shahrezaei et Swain (2008a) ont récemment donné une expression analytique de la distribution stationnaire du niveau d'ARN. Mais la généralisation de ces méthodes dans le cas d'un nombre arbitraire d'états n'est pas possible simplement et une expression analytique (même basée sur la décomposition propre de $\mathbf{M}$) reste difficile à obtenir dans le cadre général de notre modèle. Nous proposons cependant ici une méthode numérique permettant de déterminer la distribution stationnaire du niveau d'ARN pour un promoteur arbitraire.

L'équation maîtresse décrivant les changement d'états du promoteur ainsi que la synthèse et la dégradation des molécules d'ARN (réactions (III.1), (III.2) et (III.3)) est

$$\frac{\mathrm{d}\phi_r}{\mathrm{d}t} = \mathbf{D}_{\boldsymbol{\rho}}(\phi_{r-1} - \phi_r) + \gamma((r+1)\phi_{r+1} - r\phi_r) + \mathbf{M}\phi_r \tag{III.62}$$

où $\mathbf{D}_{\boldsymbol{\rho}}$ est la représentation du vecteur $\boldsymbol{\rho}$ sous la forme d'une matrice diagonale et où $\phi_r(t)$ est le vecteur à N éléments (la dépendance temporelle est implicite dans l'équation) décrivant la probabilité de trouver, à l'instant t, le promoteur dans chacun des N états et avec r molécules d'ARN. En utilisant les matrices $\mathbf{M}$, $\mathbf{D}_{\boldsymbol{\rho}}$ et $\mathbf{I}$ (la matrice identité) comme blocs de dimension $N \times N$ pour construire la matrice $\mathbf{M}'$ suivante :

$$\mathbf{M}' = \begin{bmatrix} \mathbf{M} & \gamma\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{D}_{\boldsymbol{\rho}} & \mathbf{M} & 2\gamma\mathbf{I} & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{D}_{\boldsymbol{\rho}} & \mathbf{M} & 3\gamma\mathbf{I} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_{\boldsymbol{\rho}} & \mathbf{M} & 4\gamma\mathbf{I} & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_{\boldsymbol{\rho}} & \mathbf{M} & \dots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix} \tag{III.63}$$

l'équation maîtresse (III.62) peut être réécrite simplement

$$\frac{\mathrm{d}}{\mathrm{d}t}[\phi_0, \phi_1, \phi_2, \dots] = \mathbf{M}'[\phi_0, \phi_1, \phi_2, \dots] \tag{III.64}$$

[44] Il s'agit du théorème des moments (Papoulis, 1991), qui est applicable dans le cas considéré.

Bien que $\mathbf{M}'$ soit théoriquement de dimension infinie, sa définition peut, en pratique, être restreinte à des valeurs de r inférieures à un certain seuil, celui-ci devant être sensiblement supérieur à $\max(\rho_s)/\gamma$. En effet, la queue de la distribution du niveau d'ARN est nécessairement majorée par la queue d'une distribution de Poisson de moyenne $\max(\rho_s)/\gamma$. Ainsi, restreindre la définition de $\mathbf{M}'$ significativement plus haut que $\max(\rho_s)/\gamma$ garanti une erreur très faible[45]. L'état stationnaire du système correspond donc au vecteur $[\phi_0, \phi_1, \phi_2, \dots]$ de norme 1 qui vérifie[46] $\mathbf{M}'[\phi_0, \phi_1, \phi_2, \dots] = [0, 0, 0, \dots]$. La matrice $\mathbf{M}'$ étant très creuse (sparse), ce vecteur peut être obtenu assez efficacement par des méthodes itératives. Les marginaux de la distribution bivariée $[\phi_0, \phi_1, \phi_2, \dots]$ correspondent alors à $\Lambda_0 = \sum_r \phi_{s,r}$, la distribution stationnaire de l'état du promoteur que nous connaissons déjà et à $\mathrm{P}(R = r) = \sum_s \phi_{s,r}$, la distribution de la quantité de transcrits.

On pourrait en théorie appliquer la même méthode pour déterminer la distribution du niveau de protéines (construire une matrice $\mathbf{M}''$ avec des blocs $\mathbf{M}'$), mais la taille excessive du système rendrait sa résolution impossible en pratique. Ce problème peut être contourné en appliquant des méthodes de coarse-graining sur l'équation maîtresse (Lipniacki *et al.*, 2006), mais celles-ci imposent d'importantes simplifications au modèle, susceptibles de produire des artéfacts. Notons que certains auteurs argumentent que, lorsque $\gamma \gg \tilde{\gamma}$, transcription et traduction peuvent être approximées en une seule étape cinétique (Kepler et Elston, 2001; Hornos *et al.*, 2005). Sous cette condition, la distribution que nous prédisons est alors celle du niveau de protéines.

Remarquons que l'on peut très simplement représenter un promoteur auto-régulé en modifiant la définition de $\mathbf{M}'$. Si l'on remplace chaque bloc diagonal – correspondant à la même matrice $\mathbf{M}$ dans (III.63) – par les matrices $\mathbf{M}_{[A]=0}$, $\mathbf{M}_{[A]=1}$, $\mathbf{M}_{[A]=2}$, ...(obtenues avec l'équation (III.6)) $\mathbf{M}'$ représente alors un gène codant pour le FT A qui participe sa propre régulation.

Structure de la distribution *vs* dynamique du système

On peut comprendre la structure de la distribution de R en l'interprétant comme la somme des distributions partielles pour les différents états du promoteur $\mathrm{P}(R = r|s)$, pondérées par Λ_0. Les transients de $R(t)$ que l'on observe après un changement d'état du promoteur ont une durée caractéristique de $1/\gamma$, correspondant à la durée de vie moyenne des molécules d'ARN. Ainsi, la contribution d'un état s dans lequel le temps de résidence $(1/\sum_{s' \neq s} M_{s',s})$ est très supérieur à $1/\gamma$ est une distribution en loi de Poisson de moyenne ρ_s/γ. Lorsque le temps de résidence dans l'état s se rapproche de $1/\gamma$, cette distribution est altérée par les transients qui produisent des queues de distributions (celles-ci reflètent les valeurs de R des distributions antérieures à s). Quand les temps de transitions au sein d'un sous-ensemble $\mathcal{U} \subseteq \mathcal{S}$ sont très rapides par rapport à $1/\gamma$, ce sous-ensemble se comporte comme un état unique dont le taux de transcription est le taux moyen au sein

[45] Par exemple, si $\max(\rho_s)/\gamma = 200$, on sait que $\mathrm{P}(r > 400) < 2\,10^{-11}$ et que $\mathrm{P}(r > 250) < 3\,10^{-4}$.

[46] Notons que, ce système étant singulier (car $\det(\mathbf{M}') = 0$), chaque ligne est redondante avec toutes les autres. On peut supprimer ce degré de liberté par la condition de normalisation, en remplacant par exemple la dernière ligne de $\mathbf{M}'$ par $[1, 1, 1, \dots, 1, 1]$ et en cherchant le vecteur (unique) qui vérifie $\mathbf{M}'[\phi_0, \phi_1, \phi_2, \dots] = [0, 0, 0, \dots, 0, 1]$.

de $\mathcal{U}$. Si le temps de résidence dans $\mathcal{U}$ est très supérieur à $1/\gamma$, on retrouve alors une loi de Poisson.

4 Exploitation du modèle et résultats biologiques

Nous avons vu, dans la section 3, comment le comportement du promoteur était déterminé par la structure du système, elle-même définie par les interactions mutuelles entre molécules sur le promoteur et avec les modifications conformationnelles et covalentes. Le promoteur peut alors démontrer d'une activité riche avec, en particulier, des dynamiques multi-échelles : des phénomènes globaux lents (par exemple un comportement périodique) et des dynamiques locales rapides (par exemple l'association/dissociation des FT sur un promoteur dans un état chromatinien donné). De plus, l'activité du système peut être décrite comme une superposition de nombreux comportements simples, les *modes*, que l'on retrouve alors dans les différents indicateurs décrivant la dynamique du promoteur et sa transmission au niveau d'expression. Nous montrons ainsi que l'activité spontané d'un unique gène dont tous les régulateurs sont en concentration constante peut démontrer de comportements dynamiques et stochastiques relativement complexes. On retrouve par ailleurs des similitudes avec des comportement dynamiques mis en évidence dans la littérature à une échelle supérieur, en considérant l'interaction de plusieurs gènes représentés de manière beaucoup plus simple (Simpson *et al.*, 2004; Cox *et al.*, 2006). Les spectres de puissance des niveaux d'expression témoignent en effet, dans ces études, de comportements stochastiques périodiques analogues à ceux que nous mettons en évidence au niveau du promoteur.

Ainsi, le gène, considéré depuis de nombreuses années en *Biologie des Systèmes* comme la brique élémentaire des réseaux de régulation d'où seuls peuvent naître des comportements complexes, se révèle être en réalité une source potentiellement très riche de complexité. Dans cette section, nous illustrons tout d'abord sur un exemple très simple comment la prise en compte des interactions combinatoires entre molécules et avec les modifications conformationnelles et covalentes peut conduire à des comportements inattendus de prime abord. Nous montrons ensuite comment cette dynamique permet une grande plasticité de comportements avec un contrôle fin des caractéristiques stochastiques de l'expression génique. Pour cela, nous en examinons d'abord les propriétés stationnaires (variance normalisée, distribution, ...) sur un exemple procaryote (avec des mécanismes typique de ces promoteurs et des paramètres biologiquement réalistes), puis les propriétés dynamiques (spectres de puissance, spectre de $\mathbf{M}$, ...) sur un exemple eucaryote (démontrant un comportement périodique, mais défini plus simplement que dans l'exemple présenté précédemment)[47]. Nous montrons ensuite comment certains résultats concernant le comportement périodique des promoteurs eucaryotes obtenus précédemment sur un système idéalisé peuvent être généralisés, permettant alors l'interprétation de données expérimentales de la littérature. Enfin, nous comparons plusieurs hypothèses concernant les mécanismes d'interactions des FT sur le promoteur et leur action sur l'état de la chromatine, au regard de leur cohérence avec différents types d'observations expérimentales.

[47]Ces travaux ont fait l'objet d'une publication dans *BMC Systems Biology* (Coulon *et al.*, 2010)

4.1 Variabilité induite par une concentration élevée d'un FT

Une approche classique en modélisation est d'identifier un phénomène d'intérêt et d'isoler les conditions minimales requises pour sa réalisation, puis de tester sa généralisation dans un contexte plus réaliste. Cette démarche fait souvent intervenir plusieurs modèles et/ou plusieurs formalismes de modélisation. Par exemple, un résultat peut être mis en évidence sur un modèle minimaliste en faisant appel à une résolution analytique puis généralisé sur un modèle plus complet par l'intermédiaire de la simulation. En effet, un intérêt majeur de la simulation, comparée aux méthodes analytiques, est que la complexification du modèle ne proscrit pas son étude (même si l'accroissement de l'espace paramétrique en résultant peut proscrire une exploration systématique). Le modèle que nous avons proposé permet de représenter un promoteur arbitraire – aussi bien minimaliste que très complet – et les dérivations s'appliquent dans tous les cas. Nous donnons ici un exemple de l'application d'une telle démarche.

La stochasticité de l'expression génique étant due à la rareté de certains événements discrets, elle est souvent imputée *a priori* aux faibles concentrations des espèces chimiques. Nous montrons ici que les interactions moléculaires au niveau du promoteur peuvent conduire à la situation inverse dans laquelle la variabilité de l'expression d'un gène est

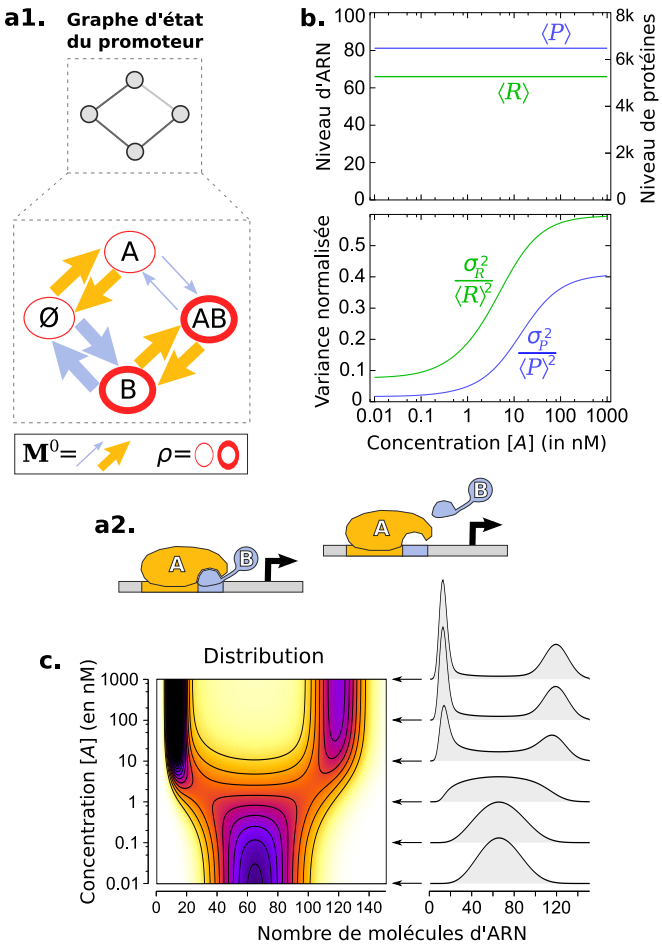


Fig. III.18: Variabilité induite par une forte concentration de FT : système minimal. (a1) Le promoteur décrit sur cette figure comporte $N = 2$ facteurs A et B dont la seule influence mutuelle est que la présence de A sur le promoteur rend plus rare l'association et la dissociation de B. Le taux de transcription ne dépend pas de la présence A. (tous les paramètres sont pris dans des plages de valeurs biologiement réaliste[48] et sont résumés dans le tableau B.3, annexe B) (a2) Un tel système peut être obtenu très simplement, par exemple si la forme du FT A est telle qu'il retient le FT B s'il est associé et l'empêche de s'associer s'il ne l'est pas. (b) L'évolution de la moyenne et de la variance normalisée des niveaux d'ARN et de protéines en fonction de la concentration $[A]$ révèle qu'une augmentation de la concentration d'un FT peut parfois accroître la variabilité de l'expression. (c) une description plus détaillée du comportement du système peut être obtenue en observant comme la distribution du niveau de transcrits change avec la concentration de A.

induite par une concentration élevée d'un de ses FT.

Le système le plus simple donnant lieu à cette propriété correspond à un promoteur sur lequel deux facteurs A et B interagissent de façon à ce que B s'associe et se dissocie plus lentement lorsque A est présent sur le promoteur et où le taux de transcription dépend simplement de la présence de B (figure III.18a1). Les différents paramètres de ce système ont été choisis dans des plages de valeurs biologiquement réalistes[48]. La variabilité de l'expression en fonction de la concentration de A (figure III.18b) montre, non seulement qu'un FT peut réguler la variabilité de l'expression génique sans influencer son niveau moyen (un phénomène déjà identifié à différents niveaux (Thattai et van Oudenaarden, 2001; Sánchez et Kondev, 2008)), mais surtout que la variabilité peut augmenter avec la concentration d'un FT. On passe en effet d'une distribution monomodale à une distribution bimodale lorsque $[A]$ augmente (figure III.18c). En effet, c'est ici pour une forte concentration de A que les événement influençant la transcription – l'association et la dissociation de B – sont rares. Et, comme l'ont montré de très nombreux auteurs[49], une dynamique lente du promoteur provoque une forte variabilité. Ainsi, la vision simpliste selon laquelle une augmentation la concentration d'un FT réduit nécessairement l'hétérogénéité transcriptionnelle n'est pas toujours juste.

L'exploration du domaine de validité de cette propriété lorsque l'on s'éloigne du système idéal décrit précédemment montre qu'elle reste valable pour d'autres concentrations de B, si l'on introduit un influence de B sur A, ou encore si le taux de transcription dépend également de A (figures III.19a1-4). Dans un cadre plus général, si les paramètres du système sont tirés aléatoirement selon des distributions telles que toutes les transitions de B sont sensiblement moins fréquentes quand A est présent et que le taux de transcription dépend essentiellement de B (figure III.19b1), la propriété est vérifiée, même pour un système à plus de deux facteurs (figure III.19b2).

De nombreuses situations peuvent correspondre à un tel système. Par exemple, la forme du FT A peut être telle qu'il immobilise le FT B sur le promoteur si celui-ci y est présent et qu'il l'empêche de s'associer sinon (figure III.18a2). Ou bien B peut représenter l'état ouvert ou fermé de la chromatine et A un FT ayant le même site de fixation que les complexes de remodelage chromatinien.

[48]Des valeurs similaires à celles rapportées dans des travaux portant sur la régulation procaryote, employant des techniques de thermodynamique quantitative (Ackers *et al.*, 1982; Bintu *et al.*, 2005b) et de cinétique sur molécule unique (Wang *et al.*, 2009), ont été utilisées pour les concentrations des FT ($[A] \in [10^{-2}; 10^{3}]$ nM et $[B] = 5$ nM), pour les temps de résidence et les constantes d'équilibre de dissociation des interactions bimoléculaires FT-ADN ($1/k_A^{\text{off}} = 30$ s, $1/k_B^{\text{off}} = 60$ s, $K_A^{\text{d}} = 0.5$ nM et $K_B^{\text{d}} = 5$ nM) et pour la modification de l'énergie d'activation de l'association/dissociation de B lors de la présence de A ($\Delta E = 2.5$ kcal.mol^{-1}). Les durées de vie moyennes des transcrits et des protéines sont respectivement de 5 et 20 min (Bernstein *et al.*, 2002). Les quantités moyennes de transcrits et de protéines sont $\langle R \rangle = 66$ et $\langle P \rangle = 100\langle R \rangle$ (Lu *et al.*, 2007). Tous les paramètres sont résumés dans le tableau B.3 (annexe B).

[49]Voir par exemple (Ko, 1991; Peccoud et Ycart, 1995; Rigney et Schieve, 1977; Kepler et Elston, 2001; Blake *et al.*, 2003; Raser et O'Shea, 2004; Simpson *et al.*, 2004; Paulsson, 2004, 2005a; Cox *et al.*, 2006; Lipniacki *et al.*, 2006; Paszek, 2007; Kaufmann et van Oudenaarden, 2007; Innocentini et Hornos, 2007; Shahrezaei et Swain, 2008a)

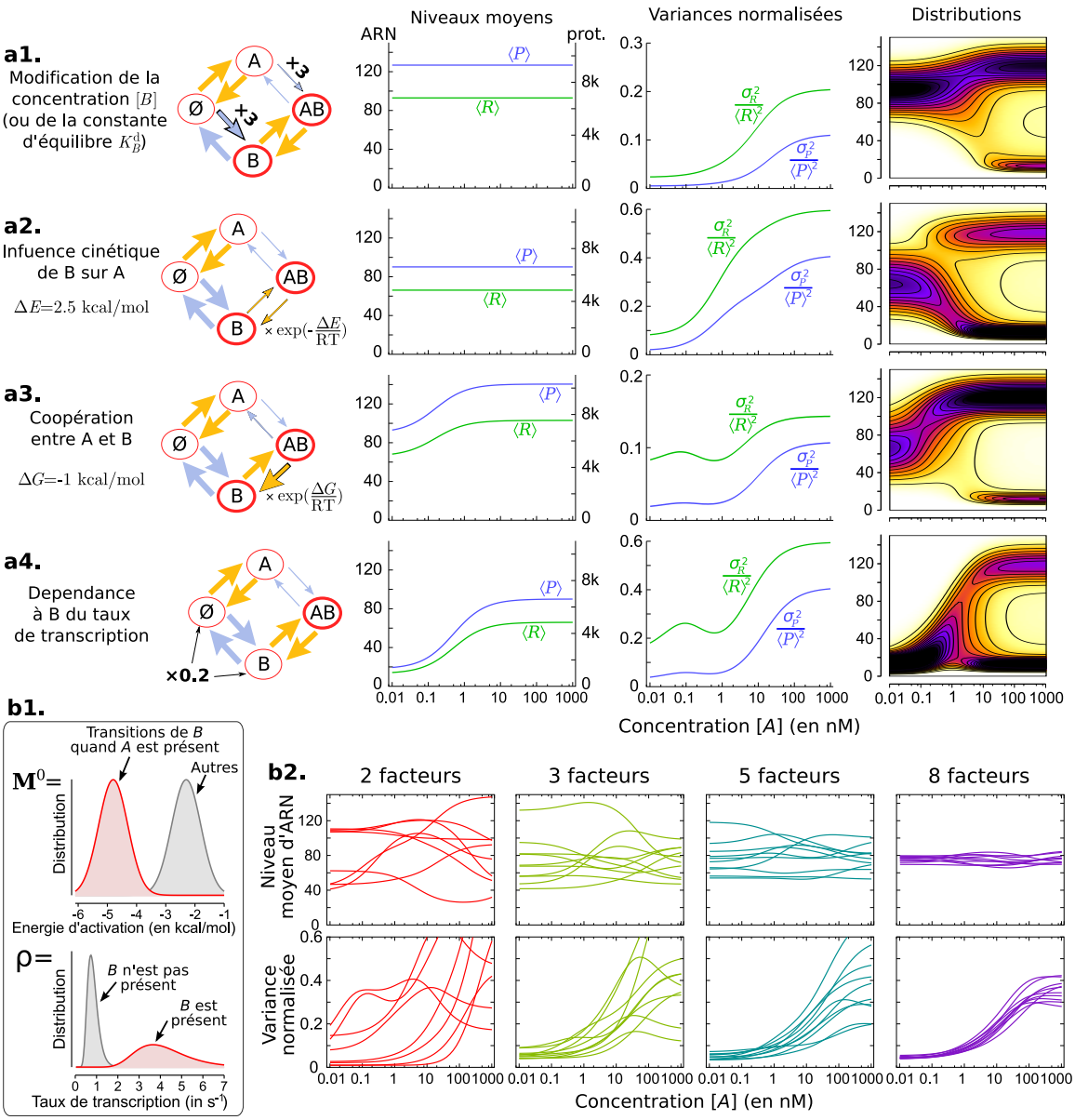


Fig. III.19: Domaine de validité : généralisation par rapport au système minimal. La robustesse de la propriété mise en évidence sur la figure III.18 peut être étudiée en testant sa validité dans des situations différentes du cas idéal. (a1-4) Lorsque l'on relâche des propriétés de symétrie et/ou d'indépendance entre paramètres, on constate que la propriété d'augmentation de la variance normalisée avec $[A]$ reste vérifiée. On peut, de plus, comprendre l'effet précis de ces modifications en observant le profil de la distribution en fonction de $[A]$. (b) Dans le cas plus général où les taux de transition $\mathbf{M}$ et les taux de transcription $\boldsymbol{\rho}$ sont tirés aléatoirement selon des lois de probabilité log-normales (b1 ; décrivant que A s'oppose aux associations/dissociation de B et que le taux de transcription dépend essentiellement de B), on observe également le maintient de cette propriété, et ce, pour un nombre de facteurs supérieur à deux (b2).

4.2 Contrôle transcriptionnel de l'hétérogénéité de l'expression

Le niveau moyen d'expression d'un gène en fonction des concentrations des FT est classiquement représenté par une combinaison de fonctions de Hill, ce qui concorde souvent avec les données expérimentales (Setty *et al.*, 2003; Kim et O'Shea, 2008). Nous montrons ici que derrière cette apparente simplicité du comportement moyen peut se cacher un contrôle fin des aspects stochastiques de l'expression génique.

Nous prenons comme exemple le système relativement simple décrit sur la figure III.20a, comportant des caractéristiques structurelles typiques de la régulation procaryote (telles que décrites par Bintu *et al.* (2005a,b); Vilar et Saiz (2005) par exemple) et avec des valeurs de paramètres typiques de ces promoteurs, telles que rapportées dans la littérature.

Sur ce promoteur, un FT A peut s'associer coopérativement en deux sites de fixation (ΔG_{A-A}=-2 kcal/mol) et est en compétition partielle avec un FT C pour l'un de ses deux sites (ΔG_{A-C}=1.5 kcal/mol) (Ackers *et al.*, 1982; Saiz et Vilar, 2006a). Le coût énergétique que représente le repliement de l'ADN sous forme de boucle (dont la valeur, typiquement entre 8 et 10 kcal/mol, dépend de la longueur et du type de la boucle) vaut ici ΔG_{loop}=9 kcal/mol et est sur-compensé par l'énergie d'interaction avec les deux FT de type A qui maintiennent la boucle fermée ($\Delta G_{\text{loop}-A}$=-5.5 kcal/mol pour chaque site) (Vilar et Saiz, 2005; Saiz et Vilar, 2006b; Saiz *et al.*, 2005). La boucle d'ADN s'oppose alors à l'association et à la dissociation du FT C ($\Delta E_{\text{loop}-C} = 2.5$ kcal/mol). Les temps de résidence bimoléculaires FT-ADN sont fixés à des valeurs plutôt courtes parmi celles rapportées par Wang *et al.* (2009) ($1/k_A^{\text{off}} = 20$ s à chaque site et $1/k_C^{\text{off}} = 60$ s) et le temps pour que l'ADN forme une boucle lorsque les deux sites de A sont occupés est très rapide ($1/k^{\text{close}} = 1$ s) (Saiz et Vilar, 2006b). Les plages de concentration (de 10^{-2} à 10^3 nM) et les constantes d'équilibre de dissociation ($K_A^{\text{d}} = 20$ nM et $K_C^{\text{d}} = 1$ nM) correspondent à des valeurs physiologiques typiques (Ackers *et al.*, 1982; Bintu *et al.*, 2005b). La transcription est réprimée par l'état fermé de la boucle d'ADN et favorisée par la présence de C et de A sur l'un de ses deux sites. La durée de vie des transcrits (5 min) et leur quantité moyenne (entre 10 et 70 copies par cellule) ont été choisies telles que rapportées par Bernstein *et al.* (2002) et Lu *et al.* (2007) respectivement. L'ensemble des paramètres de ce système est décrit dans le tableau B.4 (annexe B). Ce système est défini dans le cadre de la formulation simplifiée du modèle (section 2.5) et est énergie-indépendant ($\dot{E} = 0$).

Le niveau moyen d'ARN en fonction des concentrations des FT A et C (figures III.20b1-2) reproduit le profil classique à quatre plateaux observé *in vivo* (voir par exemple celui obtenu par Setty *et al.* (2003) pour l'opéron lactose chez *Escherichia coli* ; figure III.20c1) et prédit par les modèles thermodynamiques (figure III.20c2, d'après Bintu *et al.*, 2005b). Cependant, les aspects stochastiques de l'expression tels que la variance normalisée (figure III.20b3) et la distribution complète de la quantité d'ARN (figure III.20d) en fonction des concentrations de FT révèlent des profils beaucoup plus riches et caractéristiques. Par exemple, le nombre, la position, la forme et la taille relative des différents modes sont très variables.

Remarquons que le comportement moyen est totalement déterminé par les énergies associées aux états $\mathbf{G^0}$. En effet, le système étant énergie indépendant, on peut, comme

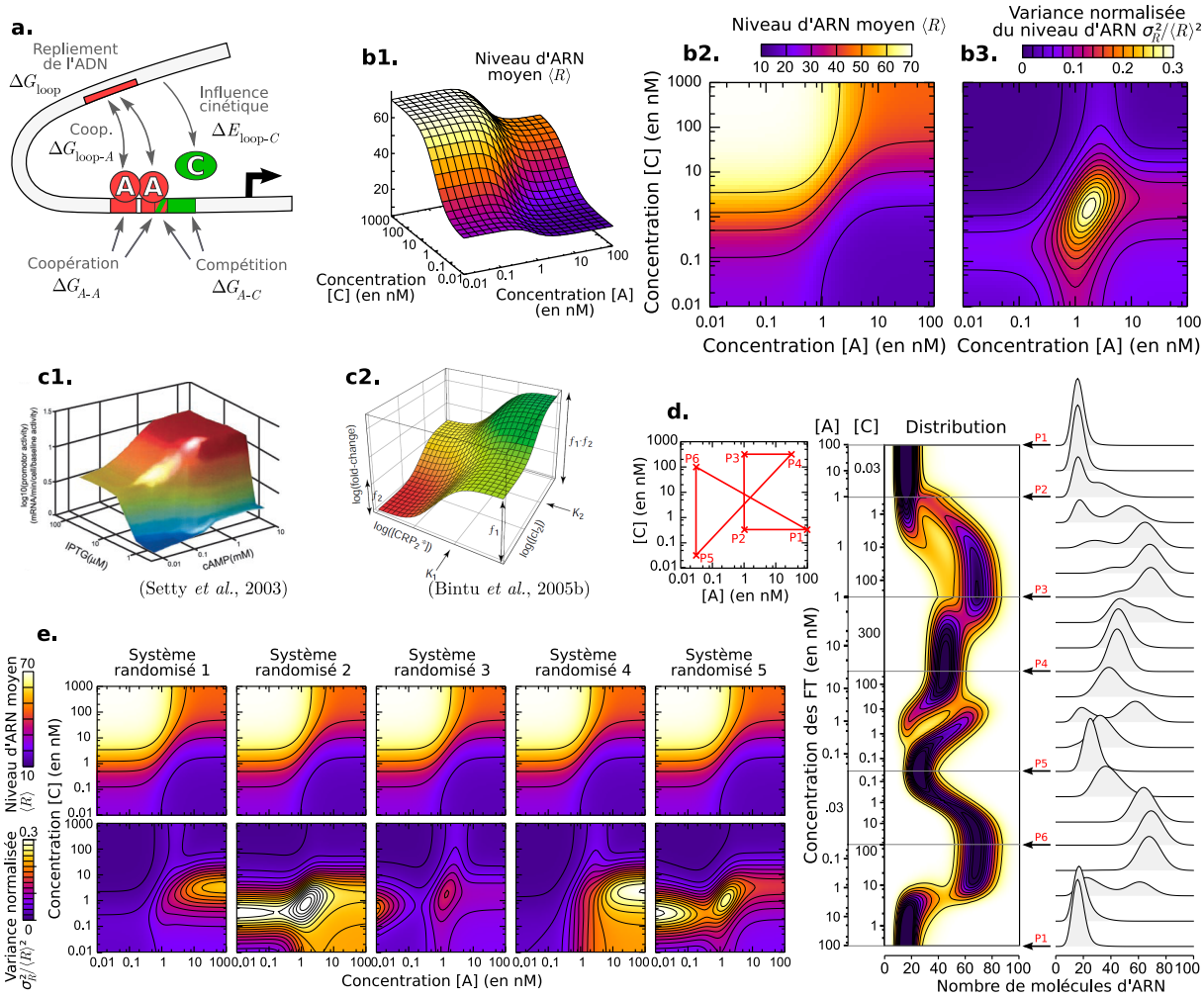


Fig. III.20: Régulation des propriétés stationnaires de la variabilité de l'expression. (a) Le promoteur représenté sur cette figure correspond à un système relativement simple présentant des caractéristiques et des valeurs paramétriques typiques des promoteurs procaryotes (cf texte et tableau B.4, annexe B). Le FT A possède deux sites de fixation sur lesquels deux molécules peuvent s'associer de façon coopérative. L'un des deux sites est en compétition partielle avec un site de fixation pour un FT C. Et enfin, le repliement de l'ADN sous la forme d'une boucle est influencé par la présence de A et influence en retour la dynamique d'association/dissociation de tous les FT. (b) L'exploration, en fonction des concentrations $[A]$ et $[C]$, de la moyenne (b1-2) et de la variance normalisée (b3) de la quantité de transcrits révèle qu'un profil stochastique caractéristique peut se cacher derrière un comportement simple du niveau moyen. (c) Ce dernier, descriptible par des fonctions de Hill et similaire aux observations expérimentales (c1 ; d'après Setty *et al.* (2003)) et aux prédictions des modèles thermodynamiques (c2 ; d'après Bintu *et al.* (2005b)), n'est donc que peu révélateur du comportement du promoteur. (d) La distribution complète du niveau d'ARN (représentée ici le long d'un chemin arbitraire car de dimensionnalité trop élevée) montre quant à elle un certain nombre de caractéristiques beaucoup plus spécifiques (nombre, taille, position et forme des différents modes). (e) La modification des énergies d'activation $\mathbf{E}^0$ (par l'ajout d'une variable aléatoire de distribution normale centrée d'écart type 3 kcal/mol et identique dans les deux sens de chaque réaction, afin de préserver le caractère énergie-indépendant du système) n'a aucune incidence sur l'expression moyenne tant que les énergies libres $\mathbf{G}^0$ restent inchangées. En revanche, sa profonde influence sur la variabilité de l'expression révèle que des mesures stationnaires de cette dernière (données de cytométrie de flux par exemple) contiennent des informations dynamiques sur l'activité du promoteur.

le font les modèles thermodynamiques (Ackers *et al.*, 1982; Bintu *et al.*, 2005a), déterminer la probabilité que le promoteur soit dans chacun des différents état à partir de la seule définition de $\mathbf{G^0}$ (cf équation (III.10) et section 3.4) et en déduire alors le niveau moyen de l'expression. En effet, si l'on modifie le système en randomisant les énergies d'activation $\mathbf{E^0}$ sans changer $\mathbf{G^0}$, le profil de l'expression moyenne reste parfaitement identique (figure III.20e). On constate en revanche que la variance normalisée décrit des profils remarquablement variés, une observation ayant d'importantes implications à la fois expérimentales et théoriques.

D'une part, cette observation montre, d'un point de vue expérimental, que des mesures, même *stationnaires*, de la variabilité de l'expression témoignent des aspects *dynamiques* de l'activité promoteur, contrairement aux mesures de l'expression moyenne. L'activité *in vivo* d'un promoteur peut être explorée expérimentalement en fonction de la concentration de ses FT (Setty *et al.*, 2003; Kim et O'Shea, 2008). Par exemple, les mesures effectuées par Setty *et al.* (2003) (figureIII.20c1) du niveau moyen de transcrits de l'opéron lactose chez *E. coli* pour différentes concentrations de cAMP et d'IPTG ont été interprétées par ces mêmes auteurs en termes de fonctions de Hill. Comme l'illustrent les figures III.20d et III.20e, les aspects stationnaires de la variabilité de l'expression (mesurables très simplement par cytométrie de flux) constituent un réelle signature du système sous-jacent et reflètent en particulier sa dynamique. Avec de telles données obtenues pour des promoteurs au moins partiellement connus (pour lesquels on connaît par exemple le nombre et la nature des facteurs), on peut donc espérer, à l'aide de notre modèle et de ses dérivations, pouvoir inférer les paramètres cinétiques du système – des informations pourtant reconnues comme difficiles à obtenir expérimentalement (Saiz et Vilar, 2006a).

D'autre part, la grande variété des profils de la variance normalisée représentés sur la figure III.20e, ainsi que les modifications fines de la distribution en fonction des concentration des FT illustrées par la figure III.20d, montrent que la régulation transcriptionnelle est un moyen très flexible pour la cellule de contrôler précisément la variabilité son expression génique. Par exemple, la répartition des phénotypes au sein d'une population clonale peut être régulée de façon très précise en fonction de la concentration de certaines molécules dans le milieu.

Ainsi, même s'il reste relativement simple dans cet exemple, un promoteur faisant intervenir quelques facteurs peut facilement démontrer d'une grande plasticité dans la modulation de la variabilité de son expression. On peut alors se demander si une telle capacité, montrée ici sur les propriétés stationnaires de l'expression, peut également avoir lieu au niveau des propriétés dynamiques du promoteur. Nous tentons maintenant de répondre à cette question en considérant un système représentant un promoteur eucaryote, dans lequel la dynamique tient une place particulièrement importante.

4.3 Régulation complexe de la dynamique du promoteur

Les aspects dynamiques de l'activité du promoteur, et en particulier le comportement périodique de son occupation par des FT et des modifications de la chromatine, peuvent également démontrer des propriétés complexes de contrôle par la concentration des FT. Comme nous l'avons montré précédemment, la répartition sur le plan complexe des valeurs propres λ_i de la matrice $\mathbf{M}$, appelée *spectre de* $\mathbf{M}$, reflète toute la dynamique de la structure (cf section 3.2). Chaque valeur propre λ_i correspond à une composante élémentaire, appelée *mode*, que l'on retrouve dans les différents indicateurs (ChIP, corrélations croisées, ... ; cf section 3.3). La position d'une valeur propre λ_i sur le plan complexe détermine la fréquence et la cohérence du mode correspondant, s'il s'agit d'un mode périodique, ou son temps de relaxation s'il s'agit d'un mode apériodique (figure III.8).

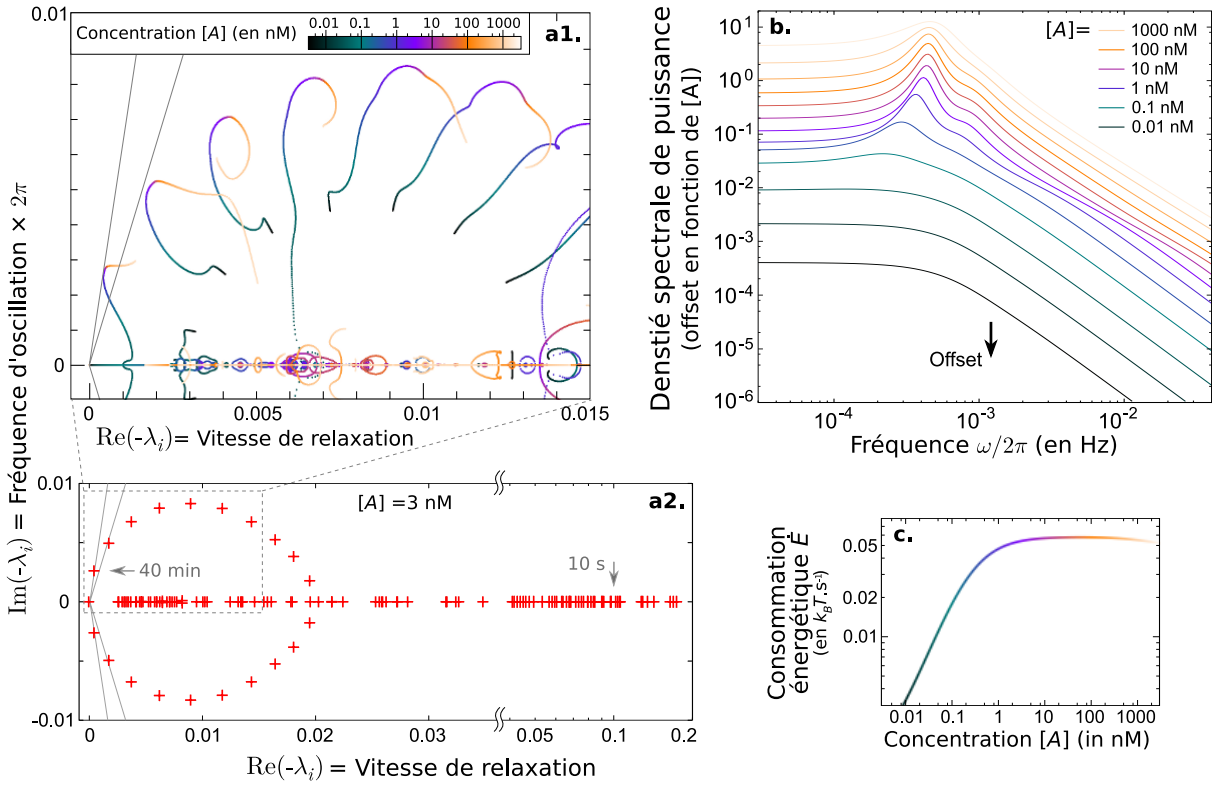


Fig. III.21: Modulation de la dynamique du promoteur par la concentration d'un FT. (a1) Les trajectoires des valeurs propres $-\lambda_i$ de la matrice $-\mathbf{M}$ sur le plan complexe en réponse à une modification de la concentration du FT A décrivent comment la dynamique du système peut être contrôlée par cette molécule. À chaque valeur de la concentration $[A]$ correspond une répartition donnée de ces valeurs propres, décrivant les caractéristiques temporelles des différentes composantes de la dynamique du système (cf section 3.2). Par exemple, la figure (a2) correspond à la concentration $[A]$=3 nM. (b) Ces trajectoires se traduisent par une modification du spectre de puissance $S_X(\omega)$ de l'efficacité transcriptionnelle du promoteur (par exemple l'apparition/disparition ou la modulation de la fréquence de comportements périodiques) pouvant alors se répercuter sur le niveau d'expression (cf section 3.6). (c) La consommation énergétique $\dot{E}$ du système est également fonction de $[A]$. Elle est faible lorsque l'absence de ce FT empêche l'activité périodique.

Les éléments de la matrice **M** étant fonction des concentrations des FT, la position de ses valeurs propres sur le plan complexe l'est également. Pour illustrer ce phénomène, nous faisons appel à un exemple de promoteur eucaryote différent de celui défini précédemment dans ce chapitre (figure III.4). En effet, dans cet exemple (figure III.21), le graphe d'état du promoteur n'est pas pré-structuré par la supposition de certaines actions réciproques entre FT et FC. Nous considérons à l'inverse toutes les combinaisons possibles de 7 facteurs dans le but de ne pas contraindre le système *a priori* à une certaine structure. Un tel promoteur n'ayant alors pas spontanément une activité cyclique, nous avons fixé les pondérations du graphe d'état de façon à obtenir ce type de comportement. Pour cela, la matrice **M** a été obtenue à l'aide d'un algorithme d'optimisation (annexe B.2), conduisant le système à avoir un comportement périodique cohérent et significativement plus lent que le reste de la dynamique. On obtient ainsi un système présentant une activité cyclique lente de période 40 min et dont les dynamiques les plus rapides ont une durée caractéristique allant jusqu'à moins de 10 s (cf flèches sur la figure III.21a2).

Dépendance de la dynamique à la concentration des FT

Comme l'illustre la figure III.21a1, les valeurs propres λ_i peuvent décrire sur le plan complexe des trajectoires non triviales en fonction de la concentration $[A]$ d'un FT. Celles-ci déterminent comment changent les caractéristiques temporelles des différents phénomènes au sein de l'activité multi-échelle du promoteur : de l'occupation périodique lente du promoteur par les FT à la dynamique rapide de leurs associations/dissocations. Ces modifications ont par exemple un impact direct sur l'activité transcriptionnelle $S_X(\omega)$, se traduisant entre autres par l'apparition/disparition ou le changement de fréquence des modes périodiques (figure III.21b). On constate par ailleurs que la consommation énergétique $\dot{E}$ du système est significativement plus basse pour des valeurs de $[A]$ ne donnant pas lieu à un comportement périodique (figure III.21c), indiquant alors que l'absence du FT A contraint le système à emprunter des transitions consommant globalement peu d'énergie.

Le simple fait que la concentration d'un FT influence son taux d'association affecte non seulement localement la durée de certaines étapes spécifiques au sein de la dynamique (par exemple certaines étapes d'un cycle), mais modifie également les probabilités relatives des différentes transitions quittant ces états et affecte ainsi potentiellement toute la dynamique du système. Cela peut alors avoir de multiples conséquences, comme celles d'influencer la probabilité des transitions inverses au sein d'un cycle orienté (ou des transitions quittant un cycle), la probabilité relative de différents chemins alternatifs dans le graphe d'état, ou encore l'engagement du système dans différentes dynamiques. Chacun de ces phénomènes a sa propre influence sur le comportement du système et pourrait être étudié dans des cas idéalisés (tels que ceux de la section 3.5). Mais dans un système non stéréotypé, tel que celui présenté sur la figure III.21, tous ces phénomènes ont lieu simultanément, en différents points et dans des mesures différentes, résultant alors en une dépendance non-triviale de la dynamique du promoteur à la concentration des FT. En effet, les valeurs propres se déplacent le long de trajectoires pouvant présenter des coudes plus ou moins anguleux et des bifurcations (résultant en de probables effets de seuil) et avec des vitesses variables (impliquant par exemple des sensibilités différentes

dans différentes plages de concentration). Ces propriétés de modulation de la dynamique du promoteur par la concentration des FT peuvent alors constituer un moyen pour la cellule de contrôler de façon précise et avec une importante flexibilité les caractéristiques temporelles stochastiques de l'expression de ses gènes.

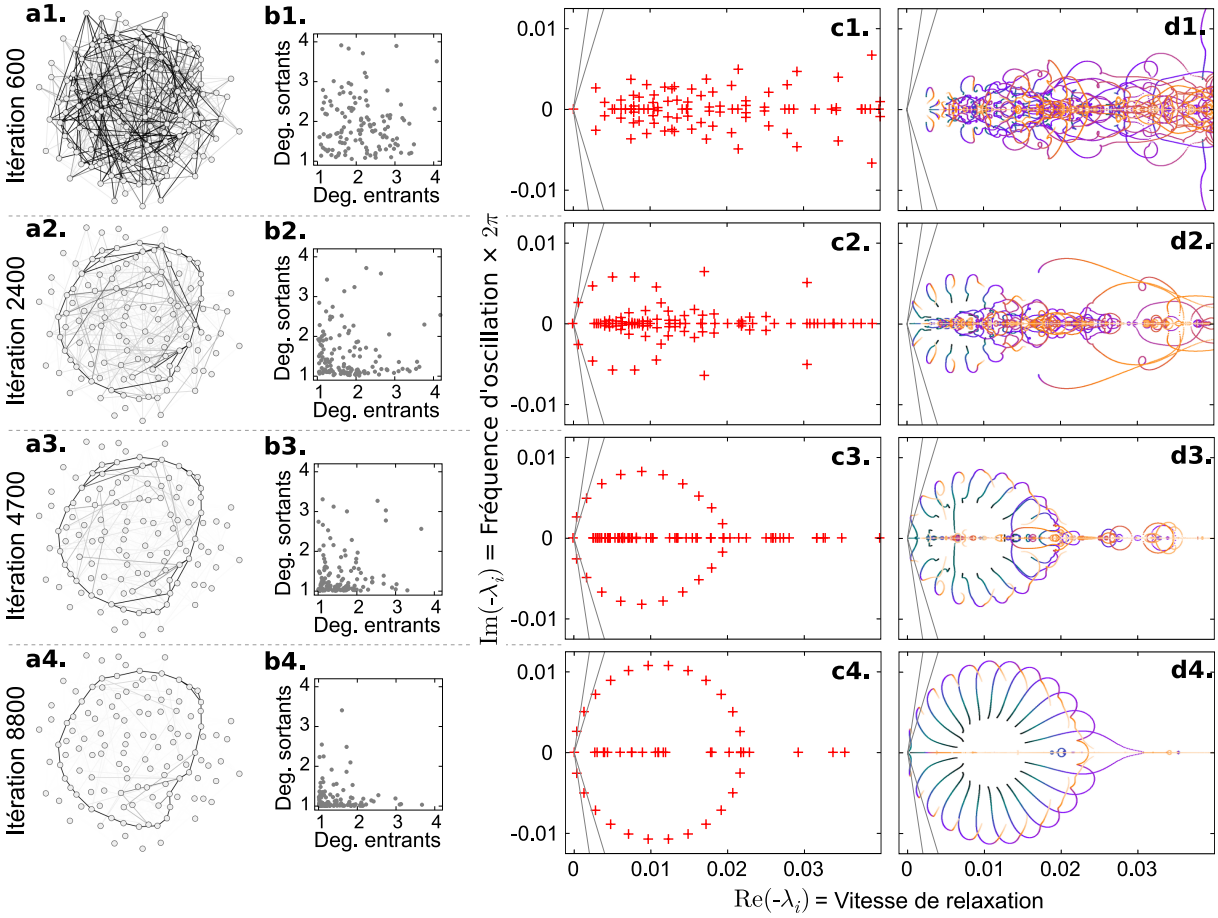


Fig. III.22: Structurations parallèles du graphe d'état du promoteur et de son spectre. Cette figure décrit la structuration progressive du système par l'algorithme d'optimisation décrit dans l'annexe B.2, induisant un comportement périodique. Cela permet alors d'évaluer l'influence du degré d'organisation sur le comportement du promoteur, en particulier en réponse à différentes concentrations des FT. (a1-4) Au sein du graphe d'état du promoteur apparaît spontanément un cycle orienté comportant d'abord de nombreux chemins alternatifs, puis tendant progressivement vers une unique séquence cyclique d'événements. (b1-4) Les degrés entrants et sortants des nœuds du graphe témoignent également de cette structuration (pour l'état s, ils sont définis respectivement par $\sum_{s'} M_{s,s'}/\max_{s'}(M_{s,s'})$ et $\sum_{s'} M_{s',s}/\max_{s'}(M_{s',s})$). (c1-4) Le spectre de la matrice $-\mathbf{M}$ (sur lequel se base le critère d'optimisation) montre l'apparition puis l'expansion d'un cercle de valeurs propres, semblable au cas idéalisé du cycle orienté homogène (cf section 3.5). (d1-4) Les trajectoires que décrivent ces valeurs propres en fonction de la concentrations du FT A montrent que, plus le système tend vers le cas du cycle homogène, plus le contrôle de sa dynamique devient contrainte et stéréotypée. (NB : Pour des questions de clarté, l'échelle de gris des transitions dans les graphes d'états ne représente pas les taux de transition $\mathbf{M}$, mais les flux à l'état stationnaire $\mathbf{M}\,\mathbf{D}_{\Lambda_0}$.)

Flexibilité et plasticité *vs* structuration du graphe d'état

La figure III.22 représente différentes étapes du processus d'optimisation employé pour obtenir l'exemple que nous considérons ici. Bien qu'uniquement basé sur des critères spectraux, cette optimisation structure progressivement le graphe d'état du promoteur : on passe d'un système peu structuré où les états présentent plusieurs transitions entrantes et sortantes (figures III.22a1 et III.22b1) à un système relativement stéréotypé, proche du cycle orienté homogène (tel que décrit à la section 3.5), comme en témoigne le cercle de valeurs propres dans le spectre de **M** (figure III.22c4) ainsi que le fait que la plupart des états n'aient qu'une unique transition sortante (figure III.22b4). La complexité décroissante que l'on observe dans la trajectoire des valeurs propres au cours du processus d'optimisation suggère alors, pour une taille de structure donnée, un compromis entre la cohérence de l'activité périodique et la flexibilité de son contrôle par la concentration des FT. C'est en effet à un niveau intermédiaire de structuration du graphe d'état, où le système garde un certain niveau de désordre, que le promoteur peut présenter une activité cyclique significativement cohérente démontrant toutefois un comportement complexe en réponse à une modification des concentrations des FT. À ce niveau intermédiaire (figure III.22a2 ou III.22a3 par exemple), le système possède un cycle significativement orienté qui n'est toutefois pas restreint à une unique séquence homogène d'événement mais peut comporter des hétérogénéités ainsi que de nombreux chemins alternatifs (de longueurs et de caractéristiques cinétiques différentes).

États alternatifs du promoteur : mobilité des molécules, redondances fonctionnelles et code des histones

Les propriétés de structuration partielle du graphe d'état énoncées précédemment correspondent précisément à ce qui est observé expérimentalement sur les promoteurs eucaryotes. En effet, du fait du caractère fortement transitoire des interactions entre molécules, les complexes multiprotéiques sur le promoteur sont relativement instables et peuvent être dans une grande variété de composition alternatives (Hager *et al.*, 2006; Métivier *et al.*, 2006). Cela correspond alors à des chemins alternatifs locaux dans le graphe d'état. Dans le cadre du graphe structuré en états chromatiniens, présenté à la figure III.4c, il s'agit des différents chemins au sein d'un même état chromatinien. De plus, les nombreuses espèces moléculaires recrutées sur le promoteur à différents instants au cours du cycle peuvent témoigner de mobilités différentes, impliquant alors des hétérogénéités cinétiques le long du cycle.

Par ailleurs, des expériences de re-ChIP ont permis de montrer que des complexes ayant une même fonction (réalisant par exemple une même modification de la chromatine) et recrutés aux mêmes étapes du cycle lent sur des cinétiques de ChIP, pouvaient en réalité s'exclure mutuellement sur un même promoteur (Métivier *et al.*, 2003, 2006). Ce phénomène correspond typiquement à un engagement du système dans différentes voies alternatives, sur la base des événements stochastiques que constitue la formation de l'un ou de l'autre de ces deux complexes. Mais, ces derniers réalisant la même fonction, cet engagement n'est que temporaire et le système abouti alors au même état chromatinien. Dans le

cadre du graphe structuré de la figure III.4c, une telle redondance fonctionnelle se traduit par une structuration supplémentaire : le sous-graphe correspondant à un état chromatinien comporte alors plusieurs sous-ensembles peu connectés entre eux, mais présentant chacun des transitions énergie-dépendantes aboutissant dans le même état chromatinien.

À une échelle supérieure, on imagine aisément que le code des histones, défini précédemment de façon simpliste par les figures III.4a1-3 de sorte que les modifications de la chromatine suivent une unique séquence d'états, soit en réalité bien plus complexe et puisse provoquer de nombreux chemins alternatifs dans la succession des états chromatiniens, ayant chacun des caractéristiques temporelles différentes. La probabilité que le système emprunte ces différentes voies dépend alors de la concentration des molécules constituant les complexes de modification de la chromatine, ainsi que de celles des cofacteurs participant au recrutement de ces derniers.

Tous ces phénomènes observés expérimentalement se traduisent par des structurations particulières du graphe d'état du promoteur à des échelles différentes et semblent être les caractéristiques fondamentales permettant un contrôle précis et flexible de sa dynamique.

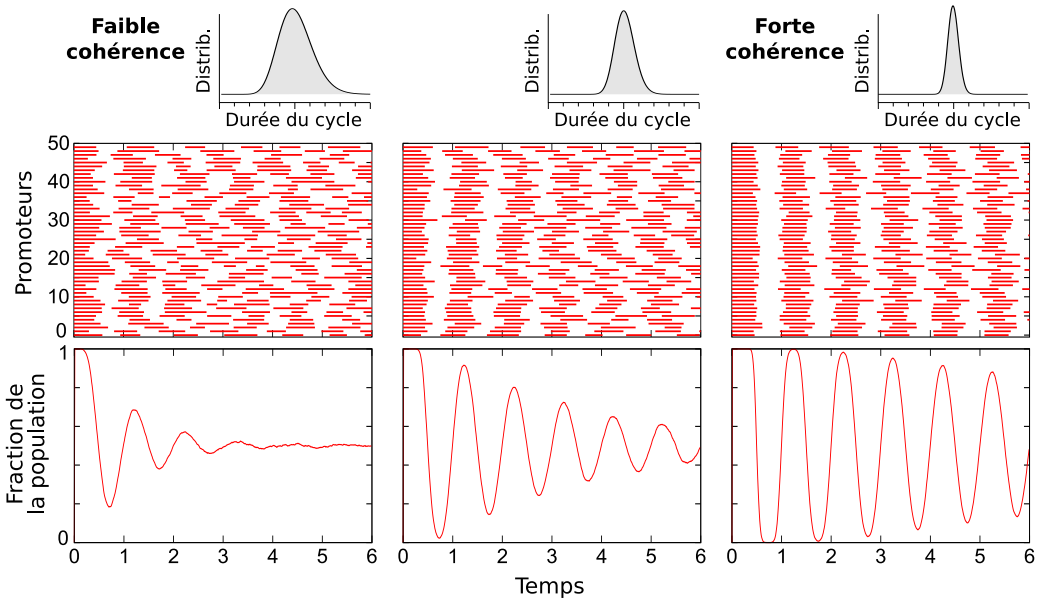


Fig. III.23: Désynchronisation au sein d'une population et cohérence des oscillations. Cette figure illustre intuitivement comment la désynchronisation progressive des différents promoteurs d'une population initialement synchronisée (comme c'est le cas lors d'une cinétique de ChIP) est à l'origine de l'amortissement des oscillations que l'on observe globalement sur la population au cours du temps. Pour une durée du cycle très variable, la fraction de la population présentant par exemple une modification donnée d'une queue d'histone tend rapidement vers une valeur moyenne, indiquant que la population est alors complètement désynchronisée. À l'inverse, si le cycle a une durée relativement peu variable, il faudra attendre un temps beaucoup plus long pour arriver à cet état moyen de la population. Ainsi, la vitesse d'amortissement des oscillations observées sur la population, que l'on peut quantifier par un facteur de cohérence κ (section 3.2), reflète le caractère stochastique du comportement périodique individuel de chaque promoteur.

4.4 Activité périodique des promoteurs eucaryotes : consommation énergétique et longueur effective du cycle

L'amortissement des oscillations observées sur une cinétique de ChIP reflète la désynchronisation des différents promoteurs au sein d'une population initialement synchronisée (Degenhardt *et al.*, 2009). La distribution du temps nécessaire pour effectuer un cycle détermine à quel point certains promoteurs seront en avance ou en retard par rapport à la moyenne de la population. Ainsi, moins cette durée est variable, plus les cellules restent longtemps synchronisées (Figure III.23).

La distribution de durée d'une séquence de p réactions de même durée moyenne τ (chacune étant distribuée exponentiellement) est simplement une distribution gamma[50] $\Gamma(p, \tau)$. Dans le cas d'un cycle irréversible et homogène à m étapes et dont le taux de transition est k^{f}, Degenhardt *et al.* (2009) déduisent que la distribution de la durée du cycle est simplement $\Gamma(m, 1/k^{\mathrm{f}})$. La moyenne et la variance de cette distribution étant respectivement m/k^{f} et $m/k^{\mathrm{f}2}$, la variance du temps nécessaire pour effectuer q cycles, exprimée en unité de la durée moyenne d'un cycle, est q/m. Les auteurs concluent donc que la population se désynchronise au fil des cycles et ce d'autant moins rapidement que le cycle comporte beaucoup d'étapes. Comme nous l'avons montré dans la section 3.2, l'amortissement des oscillations que reflète cette désynchronisation progressive peut être quantifié par un facteur de cohérence κ (équation (III.18)). Pour le système particulier qu'est le cycle homogène irréversible, il vaut par exemple $\kappa = m/2\pi^2$, indiquant en effet que la cohérence augmente avec le nombre d'étapes du cycle. Dans le cas plus général où l'on considère des réactions réversibles, nous avons donné un seuil minimal pour le nombre d'étapes et pour la consommation énergétique par cycle en dessous desquels une cohérence donnée ne peut être atteinte (cf section 3.5) :

$$m > 2\pi^2\kappa \qquad \text{et} \qquad \frac{\overset{\circ}{E}}{\mathrm{k_B}T} > 4\pi^2\kappa \tag{III.65}$$

(notons que π^2=9.87..., de sorte que ces deux bornes sont approximativement 20κ et 40κ).

De plus, on remarque que la présence d'hétérogénéités dans une séquence de transitions ainsi que l'ajout de chemins alternatifs (de taille au plus aussi longue) ne peut qu'augmenter la variabilité de la durée nécessaire pour traverser le motif résultant et donc diminuer la cohérence d'un cycle formé à partir de ce motif (figure III.24). Ainsi, les deux minorations (III.65) s'appliquent dans le cas général au chemin le plus long ou, plus précisément, à la longueur moyenne des différents chemins, pondérée suivant leurs probabilités respectives. On peut alors définir la notion de *longueur effective* d'un cycle ou d'un ensemble de transitions, correspondant à la longueur de la séquence homogène de transitions qu'il faudrait pour obtenir la même cohérence (ou la même variabilité de durée) avec une consommation énergétique identique (figure III.24). Une propriété fondamentale est que la longueur réelle (qu'il s'agisse du nombre moyen ou du nombre maximal de transitions) ne peut jamais être plus courte que la *longueur effective*.

[50]Plus précisement, il s'agit d'une distribution d'Erlang, la distribution gamma en étant une généralisation à des valeur de p non-entières.

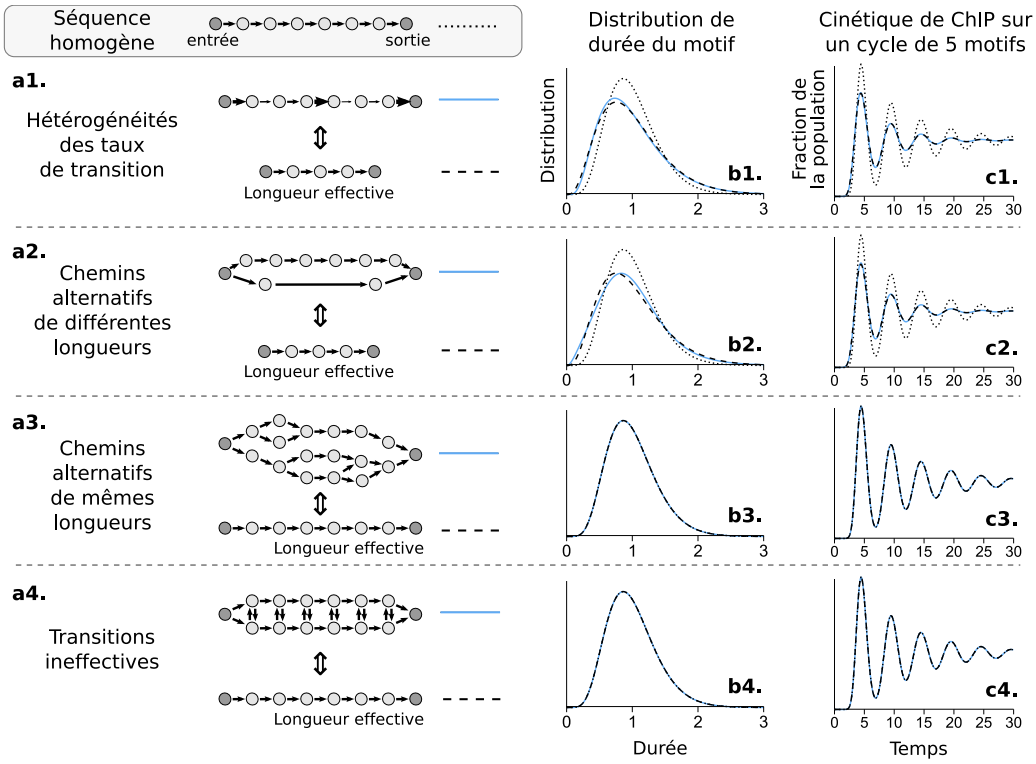


Fig. III.24: Longueur effective d'un motif : hétérogénéités et chemins alternatifs. La durée que le système met pour traverser un sous-graphe donné (appelé *motif*) du graphe d'état du promoteur dépend de l'ensemble des transitions entre les états d'entrée et de sortie (cf figure III.12 et section 3.3). En comparaison avec le cas idéalisé de la séquence homogène de transitions, la présence d'hétérogénéités ou de chemins alternatifs (a1-4) modifie la distribution de la durée nécessaire pour traverser le motifs (b1-4) ainsi que le profil de cinétique de ChIP que l'on observerait sur un cycle constitué par plusieurs instances de ce motif (c1-4). On constate que l'ajout, à une séquence homogène d'une longueur donnée, d'hétérogénéités dans les taux de transition (a1), de chemins alternatifs – qu'ils soient plus courts (a2) ou de même longueur (a3) que la séquence de référence – ou encore l'ajout de transitions entre les états analogues de chemins alternatifs de même longueur (a4) ne peut que rendre plus variable la durée du motif et moins cohérentes les oscillations sur une cinétique de ChIP. Pour un motif quelconque, on peut alors qualifier de *longueur effective* la longueur de la séquence homogène qui conduirait à la même variabilité de durée et à la même cohérence que le motif considéré (pour une même différence d'énergie libre entre les états d'entrée et de sortie ; cf section 2.4). Ainsi, la longueur effective d'un motif est toujours inférieure ou égale à sa longueur réelle, que celle-ci soit définie comme la longueur du chemin le plus long ou comme la longueur moyenne.

Ainsi, mesurer la cohérence des oscillations sur des données expérimentales permet de déduire des informations structurelles sur le système en question. Par exemple, les oscillations observées sur des promoteurs naturels au sein de cellules vivantes (figure III.26a, Métivier *et al.*, 2003) sont significativement plus cohérentes que celles observées *in vitro* par Nagaich *et al.* (2004) sur un promoteur minimal[51] (figure III.26b). Cela indique en

[51]Il s'agit d'un système minimal capable de reproduire *in vitro* un comportement périodique à partir d'un promoteur sur lequel sont assemblés des nucléosomes, de recepteur au glucocorticoïde (GR), de son ligand, du complexe ATP-dépendant de remodelage chromatinien SWI/SNF et d'ATP.

effet un nombre d'étapes et une consommation énergétique moins importants pour ce système minimal, propriété qui concorde également avec la période sensiblement plus courte des oscillations.

Remarquons que l'énergie extérieure reçue par un promoteur peut provenir de diverses sources : non seulement de l'activité des complexes ATP-dépendants de remodelage chromatiniens, mais également de celle des enzymes de modification d'histones. En effet, lors de ces réactions, l'hydrolyse des co-enzymes fournit de l'énergie au système. Par exemple, une molécule d'acétyl-CoA est utilisée comme donneur d'un groupe acétyl lors de réactions d'acétylation et son hydrolyse libère une énergie de -7.5 kcal.mol^{-1}, très comparable à celle d'une molécule d'ATP (-7.3 kcal.mol^{-1}) (Burton, 1955; Berg *et al.*, 2002; Brooks, 2004). La réaction inverse, la désacétylation, utilise le cofacteur NAD$^+$ et fournit une énergie de -4 kcal.mol^{-1} (Sauve *et al.*, 2006). De même, l'ATP est un donneur pour les réactions de phosphorylation (Berg *et al.*, 2002; Brooks, 2004). Ainsi, les réactions de modification covalente de la chromatine fournissent une énergie du même ordre de grandeur que celle d'une hydrolyse d'ATP. Cependant, celle-ci dépend des conditions ioniques et du pH du milieu. Bien qu'elle soit caractérisée de façon précise dans l'eau (ΔG_{ATP}=-7.3 kcal.mol^{-1}, soit environ -12 $\mathrm{k_B}T$), sa valeur en condition intra-cellulaire est plus approximative. Elle est estimée typiquement à ΔG_{ATP}=-12 kcal.mol^{-1}, soit environ -20 $\mathrm{k_B}T$ (Berg *et al.*, 2002).

Le compromis énergie *vs* longueur effective

Bien que l'intérêt biologique que représente le comportement périodique des nombreux promoteurs eucaryotes reste une question encore très ouverte (Ford et Thanos, 2009), le caractère particulièrement cohérent des oscillations observées sur cellules vivantes suggère fortement qu'il s'agisse d'un trait positivement sélectionné. Si l'on part du principe que des oscillations cohérentes constituent un avantage évolutif, alors notre étude permet de mettre à jour un compromis important entre la consommation énergétique le long du cycle $\overset{\circ}{E}$ et le nombre effectif d'étapes du cycle m.

La consommation énergétique constitue un coût direct évident pour la cellule qui tend souvent à optimiser cette ressource limitée. Quant au nombre effectif d'étapes du cycle, il résulte de plusieurs facteurs tels que le nombre de molécules participant à la régulation et leur façon d'interagir sur le promoteur (nous reviendrons plus en détail sur ces deux aspects). Il représente donc également un coût, bien que plus indirect : chaque molécule doit être synthétisée, les interactions moléculaires doivent évoluer vers des combinatoires spécifiques, ...

Comme le décrivent les inégalités (III.65) et la figure III.25a, une cohérence de $\kappa = 10$ par exemple peut être atteinte avec une longueur effective de seulement $m = 198$ mais cela nécessite une consommation énergétique $\overset{\circ}{E}$ particulièrement élevée. Augmenter très légèrement la longueur effective permet alors d'obtenir cette même cohérence pour une consommation énergétique beaucoup moins importante. Mais plus cette dernière se rapproche de $400\,\mathrm{k_B}T$ (cf inégalité (III.65)), plus une augmentation de m (même importante) ne permet qu'une diminution très marginale de $\overset{\circ}{E}$ (figure III.25a). La région intermédiaire

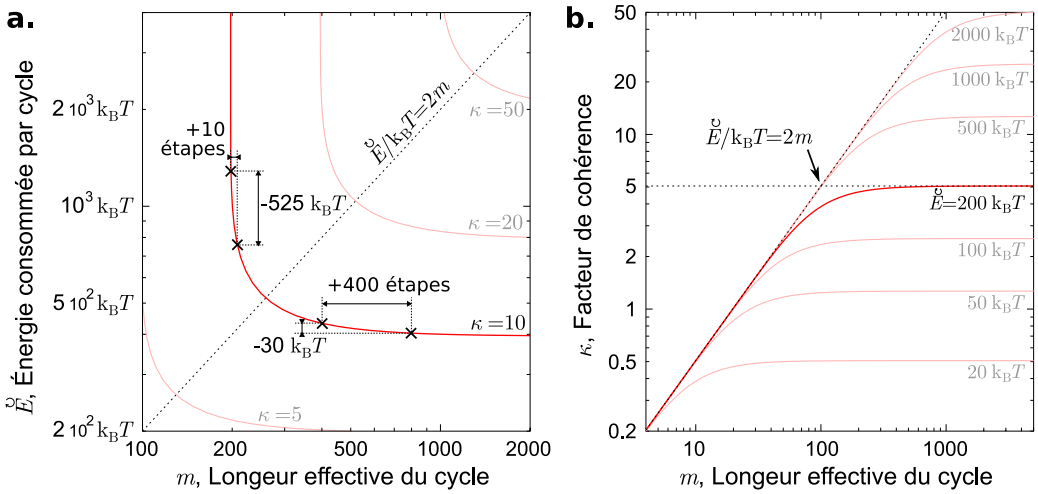


Fig. III.25: Compromis entre longueur effective et consommation énergétique. (a) Une même valeur de cohérence κ peut être obtenue avec des cycles ayant des longueurs effectives m et des consommations énergétiques $\overset{\circ}{E}$ très différentes. La courbe rouge décrit par exemple l'ensemble des valeurs de m et $\overset{\circ}{E}$ provoquant une cohérence de $\kappa = 10$ (figure identique à la figure III.14). Ainsi, plus la consommation énergétique est élevée, plus le cycle peut être court et réciproquement. Mais les deux asymptotes, verticale et horizontale (correspondant aux inégalités (III.65)), définissent, pour m et pour $\overset{\circ}{E}$, une borne en dessous de laquelle une cohérence donnée ne peut être atteinte. On constate alors que l'économie énergétique que permet le rallongement d'un cycle est très significative si celui-ci est proche de sa longueur minimale et très modeste s'il en est loin. De façon analogue, la réduction de la longueur du cycle que permet une augmentation de la consommation énergétique $\overset{\circ}{E}$ est très importante si cette dernière est proche de son seuil minimal et très marginale si elle en est loin. Le compromis entre ces deux régimes correspond aux valeurs proches de la droite $\overset{\circ}{E}/k_BT = 2m$. (b) Par le même mécanisme, l'amélioration de la cohérence d'un cycle par l'augmentation de sa longueur effective est très coûteuse et peu efficace si celle-ci est déjà très supérieure à $\overset{\circ}{E}/2k_BT$, de même que par l'augmentation de sa consommation énergétique si celle-ci est déjà très supérieure à $2mk_BT$. Ce compromis maintient alors probablement le système dans le régime intermédiaire (proche de $\overset{\circ}{E}/k_BT = 2m$) où une augmentation aussi bien de m que de $\overset{\circ}{E}$ provoque une amélioration significative de la cohérence.

entre ces deux régimes extrêmes se situe aux alentours de $\overset{\circ}{E}/k_BT = 2m$. Pour améliorer la cohérence d'un système donné, si le rapport entre $\overset{\circ}{E}/k_BT$ et $2m$ est très supérieur à 1, une augmentation de la longueur du cycle sera très efficace alors qu'une augmentation de la consommation énergétique le sera très peu, et réciproquement (figure III.25b).

Ainsi, même si la consommation énergétique et la longueur effective du cycle représentent des coûts très différents, il paraît peu vraisemblable que le rapport entre $\overset{\circ}{E}/k_BT$ et $2m$ soit très éloigné[52] de 1. Si l'on considère que les transitions énergie-dépendantes fournissent chacune une énergie proche de $\Delta G_{ATP} = 20\,k_BT$, une valeur de 1 du précédent rapport signifie alors que la longueur effective moyenne entre deux consommations énergétiques est de 10 étapes.

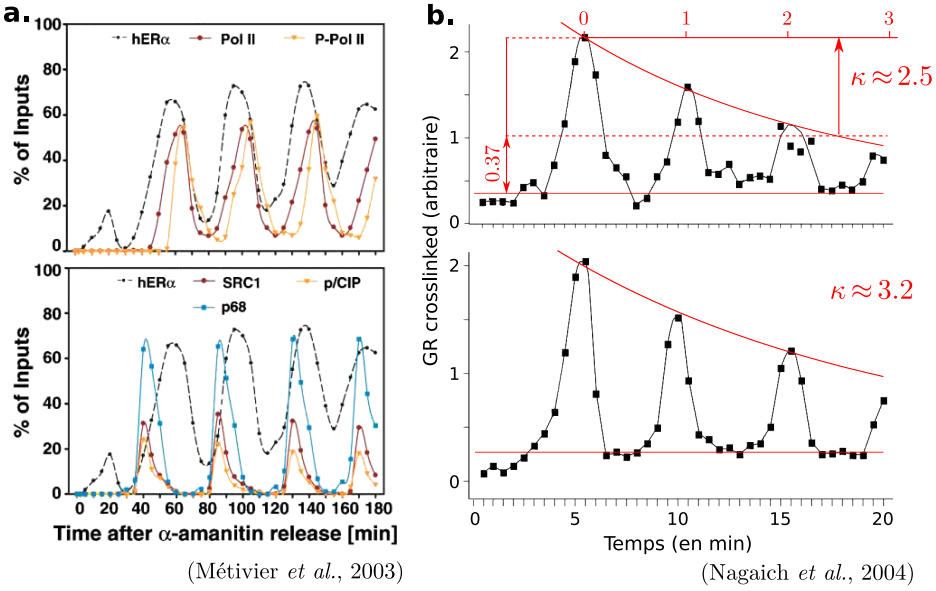
[52]Si l'on imagine, de façon simpliste, un coût proportionnel à m et à $\overset{\circ}{E}/k_BT$, mais qui soit 50 fois plus important pour m (resp. 50 fois moins important), le rapport optimal est alors de 2.35 (resp. 0.332).

4.5 Quantification sur données expérimentales

Notre modèle permet de faire un lien quantitatif entre des mesures expérimentales telles que des cinétiques de ChIP et des caractéristiques structurelles du système sous-jacent telles que la longueur et la consommation énergétique du cycle. Mais les données disponibles dans la littérature ne permettent souvent pas d'estimer ces valeurs de façon précise et ne sont pas obtenues dans les conditions nécessaires pour pouvoir en déduire ces informations de façon rigoureuse. Nous présentons ici toutefois deux méthodes permettant d'estimer ces valeurs que nous appliquons à des données de la littérature. Ces résultats sont donc plutôt à prendre à titre indicatif et comme une illustration de ce que l'on pourrait faire en réalisant des mesures dédiées.

Quantification par la cohérence des oscillations

L'estimation précise du facteur de cohérence sur des données de cinétique de ChIP (ou sur des données de même nature, issues d'autres techniques expérimentales ; cf figure I.9, page 57) requiert des mesures d'une durée au moins de l'ordre de la durée de cohérence et une résolution temporelle significativement supérieure à la période des oscillations. Par exemple, les données de Métivier *et al.* (2003) (figure III.26a), de Karpova *et al.* (2008) ou de Degenhardt *et al.* (2009) nécessiteraient d'être recueillies sur une durée plus longue pour déterminer correctement l'enveloppe exponentielle des oscillations (notons qu'il suffirait en réalité d'observer une seule oscillation après un délai long, typiquement 24 ou 48 h). Seules les données de Nagaich *et al.* (2004) (figure III.26b) permettent une estimation



(Métivier *et al.*, 2003) (Nagaich *et al.*, 2004)

Fig. III.26: Cohérence des oscillations sur des données expérimentales. (a) Pour pouvoir en estimer la cohérence des oscillations sur les données de Métivier *et al.* (2003), ce dernières nécessiteraient d'être obtenues sur une durée plus longue afin de mesurer précisément leur amortissement exponentiel. Elle semble cependant être de l'ordre d'au moins de plusieurs dizaines. (b) Sur les données de Nagaich *et al.* (2004), la décroissance exponentielle des oscillations a une durée caractéristique (durée nécessaire pour que l'enveloppe ait réduit d'un facteur $e^{-1} \simeq 0.37$; ligne pointillée) qui, exprimée en nombre de période des oscillations, indique un facteur de cohérence de 2.5 ou de 3.2.

relativement grossière du facteur de cohérence. Sur ces données, la durée caractéristique de l'enveloppe exponentielle exprimée en nombre de périodes donne un facteur de cohérence κ de l'ordre de quelques unités : 2.5 dans un cas et 3.2 dans un autre (figure III.26b). La plus faible de ces deux valeurs, $\kappa = 2.5$, indique alors un nombre d'étapes supérieur à 50 et une consommation énergétique d'au moins 100 $k_B T$ par cycle, soit au moins 8 hydrolyses d'ATP. La cohérence des données de Métivier *et al.* (2003) ne peut pas être mesurée, mais semble être de l'ordre de quelques dizaines, ce qui indiquerait une borne minimale de quelques centaines voire du millier pour le nombre d'étapes du cycle et une consommation énergétique d'au moins l'équivalent de quelques dizaines voire de la centaine d'ATP.

Ces estimations paraissent raisonnables si l'on compte le nombre de réactions de modifications de la chromatine apportant de l'énergie au système qui peuvent avoir lieu sur ces promoteurs. En effet, le promoteur ps2 étudié par (Métivier *et al.*, 2003) comporte deux nucléosomes et 13 motifs CpG (Reid *et al.*, 2009) (cf figures I.6b1-2, page 49). Chaque nucléosome possède 8 queues d'histones, chacune comportant plusieurs sites pouvant subir des modifications covalentes (Verdone *et al.*, 2005; Mellor, 2006; Reid *et al.*, 2009). De plus, il a été montré que la méthylation/déméthylation des motifs CpG participait au comportement cyclique et pouvait avoir lieu dans les deux sens du brin d'ADN (Métivier *et al.*, 2008). Enfin, l'activité cyclique de ce promoteur implique des remodelages par différents complexes, tel que NuRD et SWI/SNF, l'action de ce dernier nécessitant par exemple l'hydrolyse d'un nombre d'ATP de l'ordre de la dizaine (Kassabov *et al.*, 2003; Zofall *et al.*, 2006). De plus, l'estimation d'au moins 8 hydrolyses d'ATP pour le système minimal *in vitro* de Nagaich *et al.* (2004) paraît également réaliste puisque le seul apport énergétique possible sur ce promoteur est un remodelage par le complexe SWI/SNF.

Ainsi, les estimations obtenues à partir de la cohérence des oscillations semblent plausibles au regard des mécanismes moléculaires en jeu sur ces promoteurs. Mais notons toutefois que la cohérence des oscillations sur une cinétique de ChIP peut ne pas uniquement refléter le processus stochastique local au promoteur mais également résulter en partie d'un couplage entre plusieurs gènes de la cellule. En effet, dans le cas du ChIP sur cellules vivantes, si la méthode utilisée pour synchroniser la population à l'instant initial de la cinétique n'agit pas de manière spécifique sur le gène d'intérêt (c'est en particulier le cas de la synchronisation par l'α-amanitine), il est alors possible que plusieurs gènes aient simultanément un comportement périodique. Si les différentes fréquences de ces processus sont proches, que les gènes se régulent mutuellement et que les comportements périodiques de chaque promoteur se transmettent significativement à leurs niveaux protéiques respectifs, alors les différentes activités périodiques peuvent être couplés et peuvent ainsi présenter une cohérence plus élevée que si elles étaient observées individuellement. Même si les concentrations libres des FT dans les cellules ne présentent pas le caractère périodique observé sur le promoteur[53] de tels phénomènes ne peuvent cependant pas être exclus. Ainsi, pour dériver de manière rigoureuse les informations de longueur et de consommation énergétique, il serait nécessaire de faire des expériences complémentaires en ne synchronisant les cellules que pour le gène étudié.

[53] R. Métivier, communication personnelle.

Quantification par la forme précise des profils de ChIP

Une autre méthode permet d'estimer la cohérence de manière plus précise à partir des données de Métivier *et al.* (2003). Afin de déterminer précisément l'ordre des différentes étapes, les auteurs ont réalisé des cinétiques de ChIP à une résolution temporelle fine de 1 min et sur des fenêtres de temps restreintes décrivant, pour chaque anticorps, l'augmentation du signal depuis 0 (cf figure I.9a6, page 57). Ces cinétiques montrent pour certaines molécules un profil assez caractéristique : une augmentation brutale de leur probabilité de présence sur le promoteur.

Dans une séquence homogène de transitions irréversibles de durées moyenne τ_e, la probabilité d'avoir atteint la q-ième étape au bout d'un temps t est simplement la fonction de répartition d'une distribution gamma, soit la fonction gamma incomplète inférieure régularisée[54] (Abramowitz et Stegun, 1964), notée $P_\Gamma(q, t/\tau_e)$.

Une molécule dont le turnover est rapide et qui ne pourrait s'accrocher qu'à partir d'une étape donnée dans le cycle (l'étape en question correspondant par exemple à un remodelage chromatinien mettant à jour le site de fixation du facteur ; cf figure III.27a1) suivrait ce profil $\alpha\, P_\Gamma(q, t/\tau_e)$ à un coefficient α près. Ce dernier, reflétant à la fois la fraction du temps où le site est occupé et l'efficacité de l'anticorps utilisé, est nécessairement $0 \leq \alpha \leq 1$. Si l'affinité du facteur le long du cycle n'est pas binaire mais augmente de façon plus ou moins progressive en plusieurs étapes (figure III.27a2), le profil de ChIP observé au cours du temps correspond à la somme de plusieurs profils $\alpha\, P_\Gamma(q, t/\tau_e)$ et ne peut donc être que moins abrupt.

L'allure d'un profil binaire $P_\Gamma(q, t/\tau_e)$ observé sur une population renseigne donc sur la dynamique stochastique des différents promoteurs. Une augmentation plus ou moins abrupte indique à quel point certains promoteurs sont en avance ou en retard par rapport à la moyenne de la population $t = q\tau_e$ et reflète ainsi la même information que la cohérence. Pour une durée par transition fixée, les profils correspondant au passage de différentes étapes le long de la séquence de transitions sont d'autant moins abruptes que l'étape à lieu tard (figure III.27b1). À l'inverse, une augmentation observée au bout d'un délai donné est d'autant plus abrupte que le nombre de transitions pour y arriver est important (figure III.27b2). Ces deux tendances opposées sont analogues au fait que la variance augmente lorsque l'on fait la somme de nombres aléatoires, et diminue lorsque l'on en fait la moyenne. De plus, on peut généraliser ce résultat au cas où l'on considère des réactions réversibles, prenant alors en compte non seulement le nombre q de transitions, mais également l'énergie ΔG consommée le long de ces q transitions. En effet, comme l'illustre la figure III.27c, les profils obtenus pour des valeurs très différentes de q et ΔG (qui conduiraient à une même valeur de κ si ces q transitions constituaient un cycle) sont identiques dans la limite de $\kappa \gg 1$. Par exemple, q=2592 et ΔG=5184 $\mathrm{k_B}T$ provoque le même profil que q=65820 et ΔG=3949 $\mathrm{k_B}T$ bien que ce second système soit 25 fois plus long que le premier, car ces situations correspondent toutes deux à $\kappa = 100$ (figures III.27c1-2). Cette propriété montre l'équivalence entre la cohérence et la variabilité du durée du cycle comme nous l'évoquions déjà à la figure III.24. Toutes valeurs de m et $\overset{\circ}{E}$ correspondant à

[54] Sa définition est $P_\Gamma(q,t) = \int_0^t x^{q-1}e^{-x}dx / \int_0^\infty x^{q-1}e^{-x}dx$.

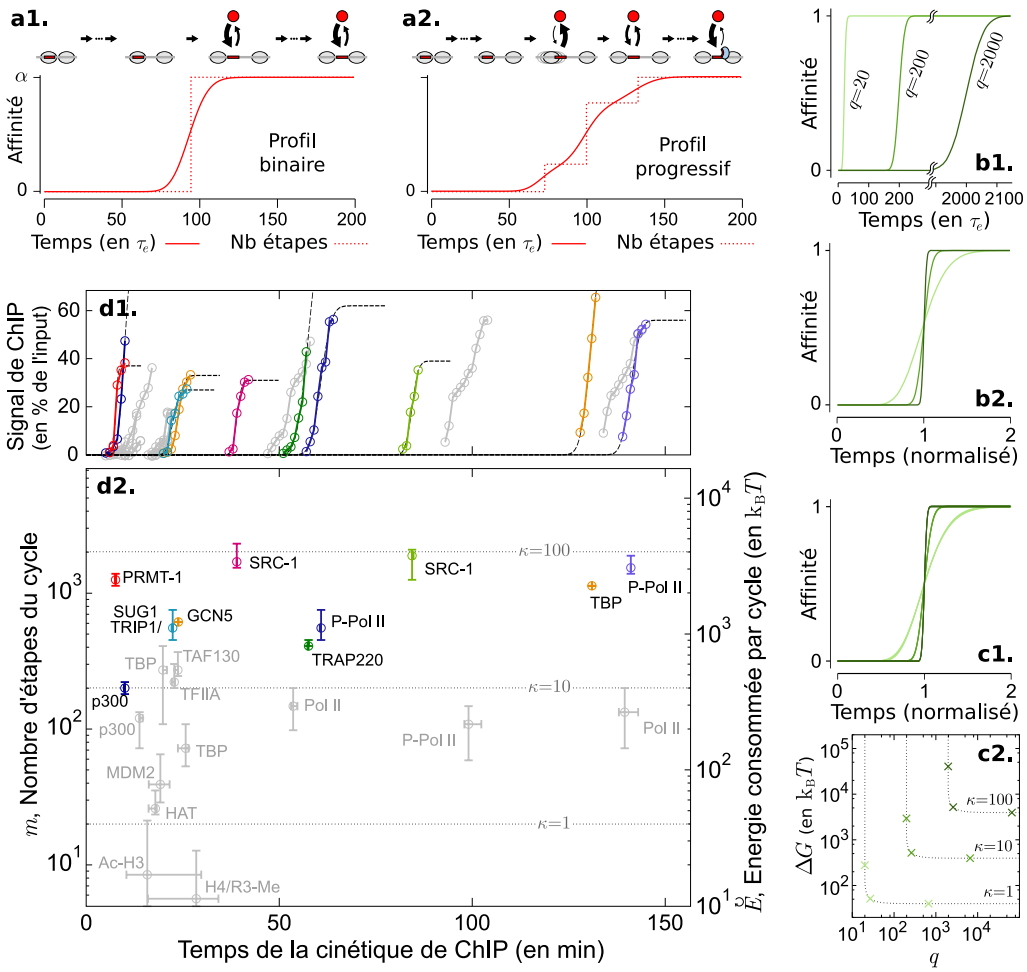


Fig. III.27: Forme des profils de ChIP et cohérence du système. Cette figure illustre comment la forme précise des profils de ChIP peut comporter les mêmes informations que la cohérence des oscillations. (a1) Un FT ne pouvant s'associer qu'à partir d'une étape donnée de la progression du promoteur (par exemple après une modification particulière de la chromatine) a un profil de ChIP assez caractéristique, correspondant à la fonction de répartition d'une loi gamma (en considérant tout d'abord les transitions comme irréversibles). (a2) Si, à l'inverse, l'affinité d'un FT augmente progressivement en plusieurs paliers au cours de la progression du promoteur, son profil de ChIP ne peut alors être que moins abrupts que dans le cas précédent. (b1) Pour une durée d'étape τ_e donnée, les profils de ChIP correspondant à des FT ne pouvant s'associer qu'après différentes étapes de la progression du promoteur sont d'autant moins abrupt que le nombre q d'étapes pour que le FT s'associe est élevé. (b2) À l'inverse, un profil observé à un instant donné est d'autant plus abrupt que le nombre q d'étape pour l'atteindre est important. (c) Ces relations peuvent être généralisées au cas où l'on considère les transitions réversibles (prenant alors en compte la consommation énergétique ΔG le long des q transitions) : la forme du profil (c1) est alors entièrement déterminée par le facteur de cohérence que produirait un cycle de q transitions consommant une énergie ΔG (c2). La forme d'un profil de ChIP reflète donc la même information que la cohérence. (d1) Ce profil peut alors être utilisé pour fitter les cinétiques de ChIP réalisées à une résolution de 1 min par Métivier *et al.* (2003), en fonction de q, de ΔG et de α (un coefficient représentant à la fois l'affinité maximal du FT et l'efficacité de l'anticorps utilisé pour le ChIP). (d2) On en déduit alors pour chaque FT des valeurs de la longueur effective m et de la consommation énergétique $\overset{\circlearrowleft}{E}$ du cycle en dessous desquelles le profil ne pourrait théoriquement pas être obtenu. Les valeurs les plus élevées indiquent un cycle de plus d'un millier d'étapes consommant l'équivalent de plus de 100 molécules d'ATP. (les fits de mauvaise qualité sont grisés et les barres d'erreur représentent des valeurs produisant des fits de qualités similaires).

une même cohérence κ produisent des distributions de durée du cycle identiques (puisque leurs fonctions de répartition sont identiques). Il s'agit donc d'une loi Gamma $\Gamma(m, t/\tau_e)$ dont le paramètre de forme est $m = 2\pi^2\kappa$ (sa variance normalisée est donc $(2\pi^2\kappa)^{-1}$). Enfin, comme le montre la figure III.24, introduire des hétérogénéités dans les taux de transition ainsi que des chemins alternatifs ne peut qu'augmenter la variabilité du délai pour atteindre la q-ième étape et donc rendre le profile moins abrupt. Les mesures de longueur obtenues par cette méthode décrivent donc la longueur effective du cycle.

En exploitant les dépendances illustrées sur les figures III.27b et III.27c, il est possible de donner une borne inférieure à la longueur effective et à la consommation énergétique du cycle. La figure III.27d2 représente l'estimation de ces deux bornes obtenues sur les données de Métivier *et al.* (2003)[55] en fittant chacun des profils de ChIP à 1 min sur α $P_\Gamma(q, t/\tau_e)$ (figure III.27d1) et en considérant une période du cycle de 40 min. Les profils pour lesquels la précision temporelle est mauvaise ont été grisés et correspondent probablement à des molécules dont le profil d'affinité est progressif (figure III.27a2). Les profils pour lesquels le fit est relativement précis indiquent une longueur effective du cycle au moins de l'ordre du millier et une consommation énergétique équivalente à au moins 100 hydrolyses d'ATP.

Notons cependant que ces valeurs sont assez approximatives et ne donnent qu'un ordre de grandeur. Les barres d'erreur n'indiquent, pour chaque profil, que les intervalles sur lesquels les fits sont de qualité comparable[56] et ne prennent pas en compte d'éventuelles imprécisions de mesure, ni le faible nombre de points expérimentaux utilisés pour chaque fit. L'amélioration de ces résultats nécessiteraient idéalement (en plus de l'utilisation d'une méthode de synchronisation spécifique au gène d'intérêt), non seulement un nombre plus important de points de mesures par profil, mais également d'avoir leur forme complète jusqu'à leurs valeurs maximales. En effet, un profil incomplet (comme c'est le cas de TRAP220 vers 60 min) ne permet pas une bonne précision sur l'estimation de α et rend le résultat assez peu fiable. En revanche, remarquons que les profils PRMT-1, SRC-1 à 40 min et P-Pol-II à 140 min sont relativement complets de ce point de vue et donnent des minorations de m et de $\overset{\circ}{E}$ parmi les plus élevées.

4.6 Progression de l'état chromatinien et mécanismes d'interaction des molécules sur/avec la chromatine

Les facteurs définissant l'état du promoteur sont de natures très variées. Notre modèle permet de représenter dans un même formalisme aussi bien les molécules pouvant se fixer sur leur site cible que les modifications conformationnelles et covalentes de la chromatine, que nous avons qualifiées respectivement de facteurs de transcription (FT) et de facteurs chromatiniens (FC) (cf section 2). La modification de l'état de la plupart des FC nécessite l'action de complexes de remodelage ou d'enzymes de modification d'histones et correspondent donc à des transitions où des molécules énergétiques sont consommées

[55] Nous remercions Raphaël Métivier pour nous avoir fourni ces données.

[56] Les différents profils ne comportant pas le même nombre de point de mesure, un critère direct sur la qualité du fit produirait artificiellement des barres d'erreur étroites pour les profils à peu de points.

(hydrolyses d'ATP ou de co-enzymes). Ces transitions étant très peu réversibles spontanément et nécessitant le recrutement de complexes différents pour être inversées, elles ont été proposées par différents auteurs comme constituant la base de la progression cyclique du promoteur (Métivier *et al.*, 2006; Hager *et al.*, 2006; Benecke, 2006; Karpova *et al.*, 2008; Degenhardt *et al.*, 2009; Reid *et al.*, 2009). En effet, contrairement aux FT qui ont une forte mobilité, les FC semblent plus à même d'encoder un état stable dans le temps et donc de permettre des oscillations lentes. Par exemple, le turnover de l'acétylation des queues d'histones va de quelques minutes à quelques dizaines de minutes (Waterborg, 2001, 2002).

Plusieurs hypothèses concernant la dynamique du promoteur ont été formulées à partir de cette progression selon une suite d'états méta-stables de la chromatine (Hager *et al.*, 2006; Métivier *et al.*, 2006; Karpova *et al.*, 2008; Degenhardt *et al.*, 2009; Reid *et al.*, 2009). Elles permettent en particulier d'expliquer la différence d'échelle de temps observée avec les techniques de FRAP et de ChIP. En effet, la première de ces techniques montre des temps de résidence des molécules allant de la seconde à la minute (McNally *et al.*, 2000; Becker *et al.*, 2002; Phair *et al.*, 2004; Rayasam *et al.*, 2005; Bosisio *et al.*, 2006; Karpova *et al.*, 2004, 2008) alors que la seconde témoigne d'une activité périodique de l'ordre de l'heure (Shang *et al.*, 2000; Métivier *et al.*, 2003; Karpova *et al.*, 2008; Degenhardt *et al.*, 2009).

Pour expliquer ces différences, il est parfois proposé que ces deux techniques ne reflètent pas le même type d'événements d'association/dissociation (Hager *et al.*, 2006; Métivier *et al.*, 2006). Le ChIP, nécessitant la fixation des cellules au formaldéhyde (cf Box 2, page 52), ne serait sensible qu'à de rares associations spécifiques et stables des molécules alors que le FRAP serait également sensible à leurs associations non-spécifiques et transitoires, plus nombreuses. Remarquons ici que, sans nécessairement faire appel à de telles différences de sensiblité, ces deux techniques mesurent en réalité deux aspects bien différents de la dynamique des FT. Le FRAP mesure le temps de résidence des molécules (en quelque sorte leur k^{off}) et ce, toutes étapes du cycle confondues, alors que le ChIP rend compte de la progression au cours du temps de leur probabilité de présence (en quelque sorte l'évolution du rapport $k^{\text{on}}/k^{\text{off}}$). La durée d'un motif sur une cinétique de ChIP ne révèle ainsi pas nécessairement le temps de résidence de la molécule. On peut en effet construire deux système différents produisant exactement les mêmes cinétiques de ChIP avec des courbes de FRAP très différentes (figure III.28). Dans un cas, la récupération de fluorescence de la courbe de FRAP est de même durée qu'un motif de ChIP, dans l'autre elle est beaucoup plus courte. Remarquons enfin que, même si elles sont très nombreuses, les associations non-spécifiques influencent d'autant moins la courbe de FRAP qu'elles sont transitoires. On peut par exemple montrer (sur un système très simple à trois états : FT dissocié, FT associé de façon non-spécifique et FT associé de façon spécifique) que si une molécule s'associe 99 fois sur 100 de façon non-spécifique, restant en moyenne 1 s, et 1 fois sur 100 de façon stable, restant alors en moyenne 100 s, la composante lente de la courbe de FRAP aura la même amplitude que la composante rapide (et non une amplitude 100 plus faible). Ainsi, les associations stables, bien que rares, doivent être détectable par le FRAP. La supposition que le ChIP et le FRAP ne reflètent pas le même type d'associations ne semble donc pas être valide.

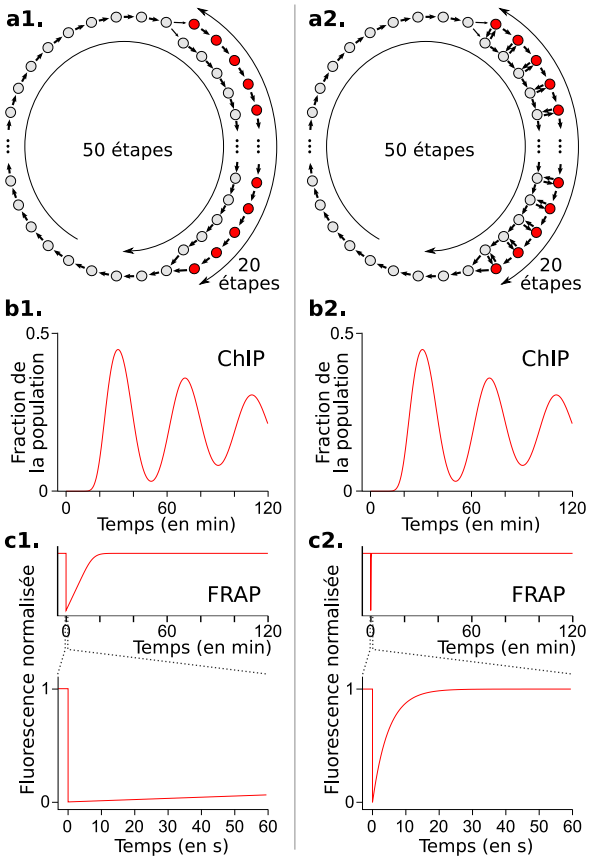


Fig. III.28: ChIP et FRAP : deux mesures complémentaires de la dynamique d'un FT. La durée d'un pic sur une cinétique de ChIP ne reflète pas nécessairement le temps de résidence d'un FT. En effet, comme l'illustre cette figure, une même courbe de ChIP peut correspondre à deux situations pour lesquelles les dynamiques du FT sont très différentes. (a) Les deux exemples comparés ici correspondent à deux cycles de même longueur au sein desquels un FT ne peut s'associer au promoteur que pendant une fenêtre temporelle donnée. Dans le premier cas, le FT ne peut s'associer qu'au début de la fenêtre et ne peut se dissocier qu'à la fin (a1). Dans le second, il peut s'associer et se dissocier à tout instant pendant la fenêtre (a2). (b) Dans les deux cas, les cinétiques de ChIP sont parfaitement identiques. (c) Mais, comme en témoigent les courbes de FRAP, le temps de résidence du FT est de l'ordre de la durée d'un pic de la cinétique dans le premier cas (c1) et est beaucoup plus court dans le second (c2). Ainsi, le ChIP et le FRAP révèlent des information différents sur la dynamique du FT, à savoir sa probabilité de présence au cours du cycle et sa durée de résidence toutes étapes du cycle confondues.

Plusieurs auteurs proposent que l'état de la chromatine, évoluant lentement et de manière cyclique selon une suite d'états méta-stables, détermine le spectre de FT ou de complexes pouvant s'associer ou non au promoteur, ces associations étant alors essentiellement transitoires (Métivier *et al.*, 2006; Hager *et al.*, 2006; Karpova *et al.*, 2008; Reid *et al.*, 2009; Hager *et al.*, 2009). Il s'agit là en partie du principe que nous avons instancié dans l'exemple de promoteur eucaryote présenté au début de ce chapitre (figure III.4). Les divers points de vue de la littérature diffèrent en revanche quant à la façon dont les molécules sont recrutées sur le promoteur, interagissent entre elles et effectuent les transitions vers d'autres états de la chromatine. Alors que certains auteurs, tels que Karpova *et al.* (2008), ne présentent l'association/dissociation des FT sur le promoteur que comme un phénomène passif (ne faisant que refléter l'évolution de la chromatine), d'autres proposent des mécanismes plus précis. En particulier, Hager *et al.* (2006) et Métivier *et al.* (2006) proposent que de nombreux complexes pré-assemblés s'associent au promoteur de façon transitoire jusqu'à ce qu'une association dite « productive » du complexe permettant la transition vers l'état chromatinien suivant ait lieu. Nous montrons maintenant que, pour reproduire le caractère fortement cohérent des périodicités observées sur les cinétiques de ChIP, la dynamique combinatoire d'assemblage/désassemblage des complexes sur le promoteur est essentielle car c'est elle qui détermine la longueur effective entre chaque événement énergie-dépendant de modification de la chromatine.

Afin d'étudier l'interaction entre la dynamique des associations/dissociations de molécules et les modifications de l'état de la chromatine, nous instancions notre modèle en considérant des interactions réciproques entre FT et FC (de façon similaires à l'exemple du promoteur eucaryote présenté au début de ce chapitre ; figures III.4a2-3). Nous ne décrivons cependant ici qu'un seul état chromatinien (figure III.29a) et considérerons par la suite celui-ci répété 100 fois sous la forme d'une séquence cyclique de modifications telle que celle de l'exemple eucaryote de la figure III.4c. Au sein du sous-graphe correspondant à cet état chromatinien, on peut remarquer trois états particulier (figure III.29a). Dans l'état d'entrée dans ce sous-graphe, atteint juste après la réaction énergie-dépendant amenant le système dans l'état chromatinien considéré, le promoteur est occupé par le complexe, que nous appelons *complexe d'entrée*, ayant réalisé cette transition. De même, juste avant de quitter l'état chromatinien, le promoteur, dans l'état de sortie, est occupé par le *complexe de sortie*, réalisant la modification de la chromatine suivante[57]. Enfin, un troisième état particulier, le *template*, est celui où le promoteur n'est occupé par aucun FT et ne comporte que les molécules constituant la chromatine. Nous nous intéressons alors à l'influence des interactions moléculaires au sein de cet état chromatinien sur les dynamiques locales et globales du promoteur.

Complexes préassemblés ou assemblage/désassemblage sur le promoteur ?

Au regard du caractère particulièrement cohérent des oscillations observées sur les cinétiques de ChIP, la durée du cycle doit être très peu variable, ce qui implique alors que la durée entre deux changements d'état de la chromatine le soit également. Pour ce faire, il est nécessaire que les associations/dissociations des molécules suivant immédiatement une réaction énergie-dépendante et aboutissant à la réaction énergie-dépendante suivante (dissociation/désassemblage du complexe d'entrée et association/assemblage du complexe de sortie) constituent un ensemble de transitions dont la longueur effective soit suffisamment importante. Cela implique en particulier que les transitions soient nombreuses, orientées et homogènes et que les chemins alternatifs soient de longueur et de durées similaires. Rappelons que nous avons montré qu'une longueur effective moyenne entre deux réactions énergétique était vraisemblablement sous-optimale si sensiblement inférieure à 10 et ne provoquait qu'une amélioration marginale si sensiblement supérieure à 10.

Nous montrons tout d'abord que, pour que le nombre de transitions soit suffisant, il est nécessaire que les complexes se forment et se déforment essentiellement sur le promoteur et non qu'il s'associent et se dissocient sous forme préassemblée. La figure III.29b instancie le modèle du « return-to-template » (Hager *et al.*, 2006) qui propose que différents complexes existent sous forme préassemblée dans le nucléoplasme et s'associent et se dissocient en permanence sur le promoteur, celui-ci retournant alors, après chaque dissociation, à son état vide, le *template*, prêt à accueillir un complexe différent. Tous les complexes

[57] On considère ici par simplicité que les complexes d'entrée et de sortie ne sont constitués que des FT réalisant la modification de la chromatine. On peut cependant prendre en compte le fait que d'autres FT peuvent être ou non présents sur le promoteur sans modifier l'activité des complexes de modification de la chromatine (c'est par exemple le cas des 7 derniers FT dans la figure III.4a). On obtient alors plusieurs états d'entrée et de sortie par état chromatinien (cf figure III.4c) et les complexe d'entrée et de sortie ont alors plusieurs compositions possibles.

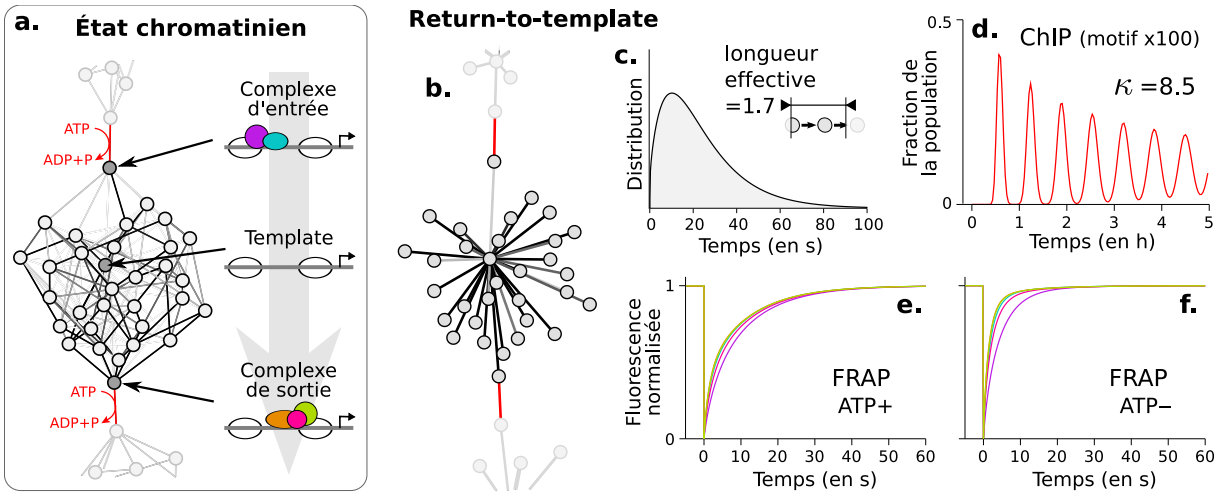


Fig. III.29: Interaction des FT sur la chromatine : instanciation du *return-to-template*. (a) Nous étudions ici en détail la dynamique du promoteur au sein du sous-graphe correspondant à un état chromatinien, c'est-à-dire l'association/dissociation de FT et de complexes entre deux événements consécutifs de modification énergie-dépendante de la chromatine. Les complexes présents sur le promoteur lors de son entrée et lors de sa sortie de cet état chromatinien – appelés respectivement *complexe d'entrée* et *complexe de sortie* – sont ceux provoquant les deux modifications de la chromatine[57]. Parmi tous les états du sous-graphe considéré, correspondant aux différentes combinaisons des FT pouvant s'associer sur la chromatine, l'état particulier où le promoteur n'est occupé par aucun FT est appelé le *template*. Ce système peut être instancié à l'aide de notre modèle afin d'étudier différentes hypothèses concernant l'interaction des FT sur/avec la chromatine. (b) Par exemple, le modèle du « return-to-template » (Hager *et al.*, 2006, cf texte) correspond à une instanciation où seuls les transitions entre le template et les autres états sont possibles. (c) Même s'il est optimisé à l'aide d'un algorithme induisant un comportement périodique cohérent (cf annexe B.2), un état chromatinien ayant une telle structure a nécessairement une distribution de durée relativement variable et correspondant donc à une longueur effective courte (ici de 1.7 et nécessairement inférieure à 3 ; cf texte). (d) Un système dans lequel cet état chromatinien serait répété 100 fois sous la forme d'une séquence cyclique ne permet alors d'atteindre qu'une cohérence sensiblement inférieure à celle de données réelles de cinétique de ChIP (ici $\kappa = 8.5$). (e) Les courbes de FRAP montrent que les temps de résidence des FT sont de l'ordre de la dizaine de secondes et (f) que ces dernier ne sont pas rallongés par une déplétion du système en ATP.

n'interagissent que de façon transitoire à l'exception de celui permettant la transition vers l'état chromatinien suivant. Dans un tel système, la longueur effective entre deux réactions énergie-dépendantes est nécessairement inférieure à 3 transitions (elle vaut par exemple 1.7 dans l'exemple présenté ici, ne permettant alors d'atteindre qu'une cohérence de 8.5 ; figures III.29c et III.29d). En effet, dans le meilleur des cas, une modification de l'état chromatinien est suivie de (i) la dissociation du complexe l'ayant provoquée, (ii) l'association du complexe capable de réaliser la modification chromatinienne suivante et (iii) la réaction énergie-dépendante en question. L'association/dissociation d'autres complexes ne peut faire qu'augmenter la variabilité de la durée de l'état chromatinien et donc diminuer la longueur effective. En effet, si un mauvais complexe s'associe, la durée supplémentaire nécessaire à sa dissociation rend le délai entre les étapes (i) et (ii) encore plus variable que s'il ne s'agissait directement de la bonne association. De façon générale, tout

événement qui ramène le promoteur dans un état déjà atteint antérieurement, ne peut que faire augmenter la variabilité de la durée de l'état chromatinien et ce d'autant plus que cet événement est long. Si les mauvaises associations sont très transitoires ce phénomène est peu important, mais la longueur effective reste dans tous les cas inférieure à 3. Un tel système étant donc sous-optimal, il demanderait une consommation énergétique très importante et semble alors peu plausible. Il paraît donc plus efficace que les complexes se forment, au moins en partie, sur le promoteur. En effet, dans ce cas, chaque association d'un FT ou d'un sous-complexe constitue une étape cinétique. Cette vision concorde avec les travaux de Gorski *et al.* (2008) montrant que l'assemblage du complexe de pré-initiation de la polymérase I se fait essentiellement sur le promoteur, permettant alors une régulation plus précise de la transcription que s'il s'associait préassemblé (Darzacq et Singer, 2008; Hager *et al.*, 2009). Bien qu'il ne s'agisse pas d'un complexe de modification de la chromatine, ses propriétés d'assemblage sur le promoteur semblent similaires à celle que nous proposons.

Dynamique hors d'équilibre et rôle moteur des réactions énergie-dépendantes

Une deuxième caractéristique essentielle pour reproduire des périodicités cohérentes est que les réactions doivent être orientées (c'est-à-dire plus probables dans un sens que dans l'autre) et ce de façon homogène. Rappelons qu'une réaction est d'autant plus orientée qu'elle provoque une diminution importante d'énergie libre (figure III.5 et section 2.4). Les transitions énergie-dépendantes de modification de la chromatine sont par nature très orientées, mais si l'on considère que les associations/dissociations de molécules le sont peu (figures III.30a1 et b1), la distribution de durée de l'état chromatinien reste assez variable. En effet, dans ce système, la formation du complexe capable de réaliser la modification de la chromatine suivante peut aussi bien se faire en peu d'étapes qu'en beaucoup d'étapes (considérant les retours en arrière). L'état chromatinien ne peut donc pas atteindre une longueur effective conséquente et les périodicités de l'activité globale ne sont que modérément cohérentes (figures III.30c1 et d1).

Ainsi, une suite de transitions fortement orientées décrit nécessairement une descente graduelle d'énergie libre, partant d'un état de haute énergie et aboutissant à un état de basse énergie. Cela signifie que la transition énergie-dépendante de l'état chromatinien doit mettre le système dans un état énergétiquement peu favorable et que les interactions moléculaires doivent être telles qu'il n'y ait pas de dissociation immédiate du complexe ou de réarrangement conformationnel faisant retomber le système vers des états faiblement énergétiques. Au contraire, les interactions combinatoires doivent accompagner cette descente graduelle en amenant le complexe d'entrée à se désassembler en plusieurs étapes successives et en provoquant la formation, également progressive, du complexe de sortie, permettant alors la transition vers l'état chromatinien suivant (figures III.30a2 et b2). Le système doit rester en permanence sur une dynamique transitoire hors d'équilibre, cette dernière n'étant possible que grâce à l'apport énergétique que constitue la réaction de modification de la chromatine réalisée par le complexe d'entrée. Si le promoteur se met rapidement dans une conformation favorable ou que le complexe d'entrée se dissocie immédiatement, le reste de la dynamique s'effectue nécessairement par des transitions entre

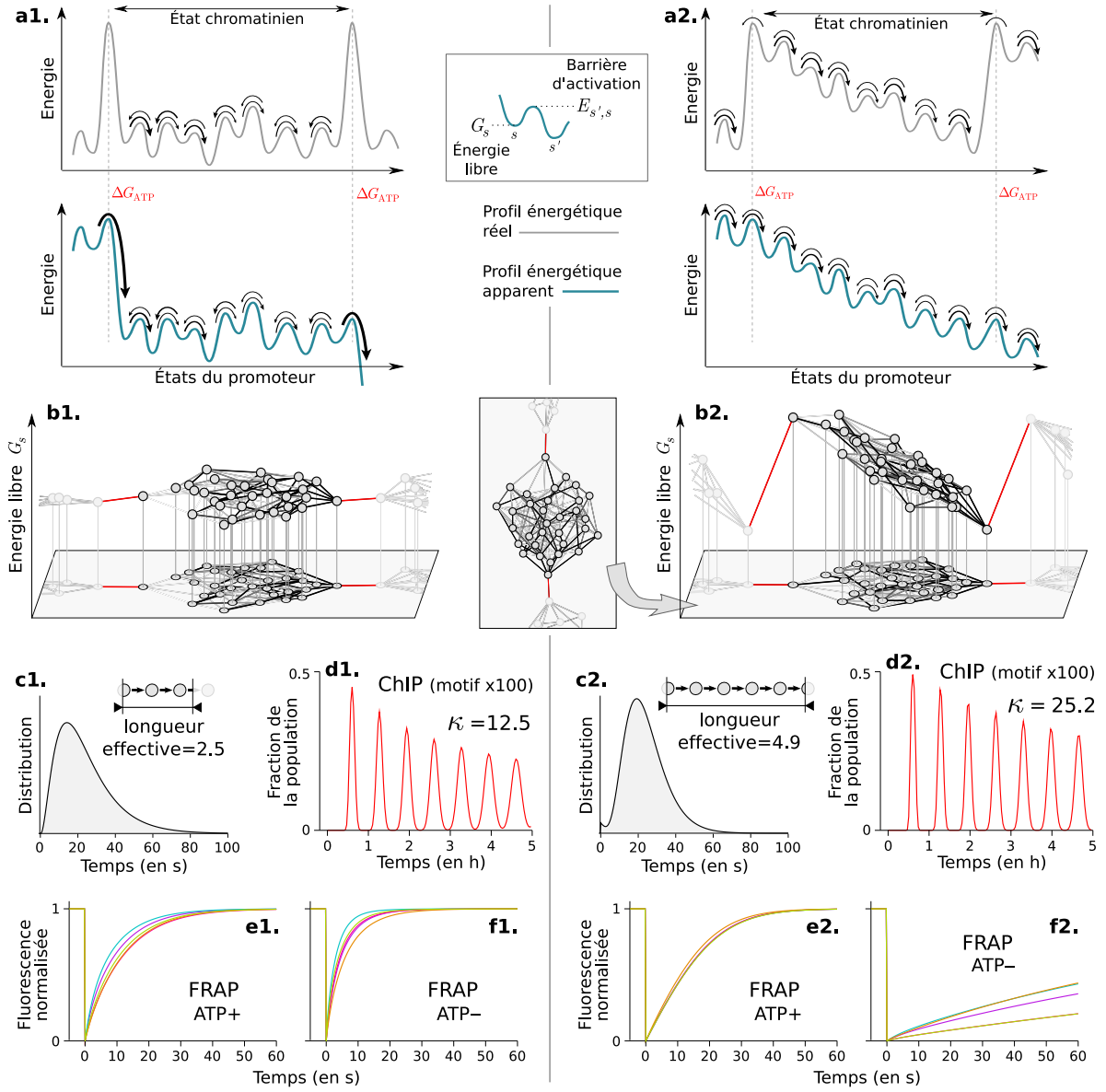


Fig. III.30: Décroissance énergétique progressive et dynamique hors d'équilibre. Cette figure présente la comparaison deux modes d'interaction des FT sur/avec la chromatine. (a1) Dans le premier cas, les états du promoteur, décrivant la présence/absence des FT, ont des énergies libres similaires et les réactions énergie-dépendantes permettent de passer des barrières d'activation élevées. (a2) Dans le second cas, le désassemblage du complexe d'entrée et l'assemblage du complexe de sortie constituent au contraire une diminution progressive de l'énergie libre depuis une valeur élevée, atteinte grâce à une réaction énergie-dépendante. Les schéma présentés ici emploient la même représentation que la figure III.5. (b) L'instanciation de ces deux mécanismes sur le système de la figure III.29a (présentée ici uniquement en termes des énergies libres G_s associées aux états et non des énergies d'activation $E_{s',s}$) permet de comparer ces deux mode d'interaction (cf annexe B.2 et tableau B.1 pour plus de détails sur la paramétrisation de ces deux exemples). (c1-2) Le second cas, où l'énergie libre décroît progressivement, permet une durée d'état chromatinien sensiblement moins variable que le premier cas, ce qui correspond à une longueur effective plus élevé et (d1-2) permet d'atteindre une périodicité plus cohérente sur les cinétiques de ChIP. (e1-2) Les courbes de FRAP montrent que pour ces deux systèmes les temps de résidence des FT sont de l'ordre de quelques dizaines de secondes. (f1-2) Cependant, seul le second cas reproduit l'observation expérimentale très singulière, décrite par la figure I.8b (page 54), que la déplétion en ATP rallonge très sensiblement ces temps de résidence.

états d'énergie similaires et les transitions ne sont alors pas globalement orientées vers l'état permettant la modification chromatinienne suivante.

Dans ce cadre, certains types de modifications conformationnelles peuvent jouer un rôle important dans la dynamique au sein d'un même état chromatinien. En effet, nous avons défini l'état chromatinien comme une combinaison de facteurs méta-stables tels que les modifications covalentes des histones et de l'ADN et les modifications conformationnelles nécessitant un apport d'énergie extérieure. Néanmoins, certaines modifications conformationnelles peuvent parfois être rapides et spontanées. En effet, en fonction de son profil énergétique le long de la séquence d'ADN, un nucléosome peut avoir différents comportements. Alors que d'importants minima locaux d'énergie peuvent immobiliser un nucléosome à certains emplacements et que son passage par-dessus certaines barrières d'énergie nécessite l'apport d'énergie extérieure (hydrolyse d'ATP), un nucléosome peut toutefois se déplacer spontanément dans certaines régions d'énergie équivalente, en allant préférentiellement vers les états de plus basse énergie. On peut alors imaginer des mécanismes réalisant une descente graduelle d'énergie entre deux modifications énergie-dépendantes de la chromatine, basées en partie sur le déplacement spontané des nucléosomes. Par exemple, après une telle modification, un nucléosome peut se trouver en haut d'un gradient d'énergie. Si de plus son déplacement le long de la séquence est obstrué par les molécules s'y trouvant associées (par exemple le complexe d'entrée), celles-ci sont alors en position énergétiquement défavorable et peuvent se dissocier progressivement au fur et à mesure que le nucléosome se déplace. Ce déplacement peut alors provoquer en parallèle la mise à jour progressive de site de fixation pour les FT du complexe de sortie. Ainsi, des réarrangement conformationnels spontanés de la chromatine peuvent participer à la dynamique intra-état chromatinien.

On remarque que, parmi les différentes hypothèses testées ici, le cas le plus vraisemblable est celui où le promoteur a une dynamique en permanence hors d'équilibre (figure III.30a2), contrairement à l'hypothèse souvent faite par la plupart des modèles d'organisation de la chromatine et d'occupation du promoteur par les FT (Segal et Widom, 2009; Jiang et Pugh, 2009; Raveh-Sadka *et al.*, 2009). En effet, dans ces modèles, les auteurs justifient l'application de méthodes thermodynamiques (imposant de considérer un système ne consommant pas d'énergie) en avançant que les modifications énergie-dépendantes ne font que fluidifier la structure en abaissant des barrières énergétiques élevées (cf section 1.3). C'est pourtant dans le cas de la figure III.30a2 que l'application d'une telle hypothèse est la moins justifiée. Ainsi, même si les méthodes thermodynamiques donnent parfois de bons résultats, la compréhension de l'organisation de la chromatine au niveau d'un promoteur peut nécessiter de considérer le système comme ouvert et hors d'équilibre.

La comparaison des courbes de ChIP et de FRAP dans les trois cas présentés sur les figures III.29 et III.30 démontre que le temps de résidence des FT est sensiblement plus court que la période des oscillations. Cependant, seul le cas où les complexes se forment sur le promoteur par une descente progressive d'énergie libre (figure III.30b2) permet de reproduire une propriété très singulière de la dynamique des FT. Le turnover des molécules est en effet très lié à l'activité ATP-dépendante des complexes de remodelages. Comme décrit plus en détail à la section 5.2 du chapitre I (et en particulier sur la figure I.8b,

page 54), plusieurs études montrent que l'inhibition de cette activité ATP-dépendante augmente drastiquement le temps de résidence des FT (Fletcher *et al.*, 2002; Elbi *et al.*, 2004; Karpova *et al.*, 2004; Rayasam *et al.*, 2005; Agresti *et al.*, 2005), expliquant par ailleurs pourquoi les anciennes mesures *in vitro* (Perlmann *et al.*, 1990) rapportaient des temps de résidence beaucoup plus long que les études récentes *in vivo* (McNally *et al.*, 2000; Becker *et al.*, 2002; Phair *et al.*, 2004). On remarque que c'est ici uniquement dans le cas du système de la figure III.30b2 qu'une déplétion en ATP provoque des temps de résidence très longs des FT. En effet, en l'absence d'énergie extérieure, le système est alors fermé (au sens décrit dans la section 2.4) et le profil apparent devient le même que le profil réel (figures III.30a1-2). La répartition entre les différents états favorise alors les états d'énergies les plus basses. Dans le cas de la figure III.30b2, il s'agit des complexes de sortie entièrement formés et la dissociation des molécules, impliquant de remonter la descente énergétique en sens inverse, est alors rare. Dans le cas de la figure III.30b1, tous les états ont des énergies comparables et les associations/dissociations de molécules se font aussi rapidement que lorsque l'ATP est présent.

Notons que la dynamique de désassemblage du complexe d'entrée est aussi importante que la dynamique d'assemblage du complexe de sortie. Ces deux dynamiques peuvent par ailleurs être entremêlées : le complexe de sortie peut commencer à s'assembler alors que le complexe d'entrée est encore partiellement présent et des interactions combinatoires peuvent être telles que la dissociation de certains FT du complexe d'entrée est facilitée par la présence de certains FT du complexe de sortie. De plus, des molécules ne faisant pas partie de ces deux complexes peuvent participer à la dynamique en facilitant par exemple le recrutement ou la dissociation de certains sous-complexes.

Bien que les mécanismes d'assemblage du complexe de transcription ne soient pas nécessairement les mêmes que ceux des complexes de modification de la chromatine, leur comparaison n'en est pas moins instructive. Dans le mode d'assemblage dénommé « random assembly » par Hager *et al.* (2009), les molécules s'associent de façon transitoire et *indépendantes* et forment occasionnellement un complexe complet et productif. Dans ce cadre, si pour chaque molécule le temps de résidence moyen est du même ordre que le temps d'absence, les énergies libres associées aux différents états sont comparables et l'on est dans le cadre du système de la figure III.30b1. Si les temps de résidence sont plus importants que les temps d'absence, cela signifie que l'énergie libre diminue à chaque association d'une molécule et l'on est alors dans le cadre du système de la figure III.30b2. Dans le mode d'assemblage dénommé « regulated assembly » par Hager *et al.* (2009), les molécules s'associent de façon coopérative et interstabilisatrices, permettant alors de meilleurs propriétés de régulation (Gorski *et al.*, 2008; Darzacq et Singer, 2008). En effet, une interstabilisation des molécules sur le promoteur signifie que la co-occurrence de deux molécules provoque une diminution d'énergie libre plus importante que la somme des diminutions correspondant à leurs associations séparées. L'énergie libre diminue donc au fur et à mesure que le complexe se forme. Ainsi, ces deux modes d'assemblage – « random assembly » et « regulated assembly » – peuvent tous les deux rentrer dans le cadre de la figure III.30b2. Mais, comme nous le montrons maintenant, des molécules totalement indépendantes les unes des autres ne permettent cependant pas d'atteindre une longueur effective élevée. Des interactions combinatoires, même modérées, sont alors nécessaires.

Le compromis nombre de molécules *vs* combinatoire des interactions

Dans le cas où le désassemblage/assemblage des complexes sur le promoteur se fait de façon progressive et par une descente graduelle d'énergie libre, on peut mettre en évidence un compromis entre le nombre de molécules des différents complexes et la combinatoire de leurs interactions. En effet, une même longueur effective peut aussi bien être obtenue avec peu de molécules et des interactions combinatoires très fortes qu'avec beaucoup de molécules et des combinatoires faibles. Par exemple, seulement 4 molécules (impliquant $2^4 = 16$ états) permettent de faire une séquence très contrainte de 10 événements, alors que 10 molécules permettent de faire un ensemble de transitions très peu combinatoires contenant de nombreux chemins alternatifs chacun de longueur 10 (comme dans le cas de la figure III.24a3). Dans le premier cas les molécules s'associent et se dissocient plusieurs fois au sein de l'état chromatinien, alors que dans le deuxième, chaque molécule peut ne s'associer qu'une fois. Remarquons cependant que, dans ce second cas, des molécules totalement indépendantes les unes des autres, même nombreuses, ne permettent pas une longueur effective importante[58]. Quelques interactions combinatoires (par exemple de simples exclusions mutuelles entre des molécules du complexe d'entrée et d'autres du complexe de sortie) peuvent suffire à donner un ordre préférentiel à l'assemblage et à augmenter la longueur effective de l'état chromatinien. La combinatoire des interactions ainsi que le nombre de FT impliquées dans la dynamique d'un état chromatinien jouent donc tous deux sur la longueur effective de ce dernier.

*
* *

Le modèle de dynamique stochastique de la régulation transcriptionnelle présenté au début de ce chapitre peut donc être utilisé pour montrer des propriétés de différentes natures et à différents niveaux. Par exemple, concernant la régulation eucaryote, nous avons pu montrer que le degré de structuration du graphe d'état du système, pouvant prendre place à différentes échelles, déterminait la flexibilité du contrôle de la dynamique par la concentration des FT. Nous avons également vu, à un niveau global, que la consommation énergétique d'un graphe d'état à structure cyclique, ainsi que sa *longueur effective* (une notion de longueur généralisée au cas où le cycle comporte diverses sortes d'hétérogénéités et de chemins alternatifs) déterminaient la cohérence des oscillations observées sur une cinétique de ChIP. Ainsi, une cohérence donnée, mesurable par différentes techniques sur des données expérimentales, permet d'estimer des bornes inférieures pour ces mesures de longueur et de consommation énergétique. De plus, nous avons mis à jour un compromis entre ces deux quantités indiquant alors, à un niveau local, que le nombre effectif d'événements entre deux modifications énergie-dépendantes de la chromatine était vraisemblablement peu éloigné de la dizaine. Enfin, à cette même échelle locale, nous avons comparé différentes hypothèses pour l'interaction des FT avec la chromatine provoquant sa modification. Il apparaît alors que les complexes auraient tendance à s'assembler/désassembler

[58] Même si les dissociations des molécules du complexe d'entrée et les associations de celles du complexe de sortie ont toutes lieu avec la même durée moyenne τ, les événements successifs auront des durées exponentielles de moyenne $\tau/10$, $\tau/9$, $\tau/8$ ... impliquant une longueur effective de l'état chromatinien, par exemple, de 2.93 pour 10 molécules, 5.18 pour 100 molécules ou 7.49 pour 1000 molécules.

essentiellement sur le promoteur (et non à s'associer/dissocier sous forme préassemblée) et ce, par étapes successives correspondant à l'assemblage d'une structure de plus en plus énergétiquement favorable, depuis un état rendu peu favorable par une réaction énergie dépendante. Ce mode d'interaction est en effet en mesure de reproduire plusieurs caractéristiques observées expérimentalement par diverses techniques (cohérence des cinétiques de ChIP, dépendance à la déplétion en ATP des courbes de FRAP, ...).

Ce modèle générique et ses dérivations peuvent ainsi servir de base pour de nombreux travaux, tels que ceux que nous venons de présenter, ainsi que pour des investigations futures, comme nous en évoquons maintenant quelques pistes.

5 Perspectives

Comme l'illustrent les différents résultats présentés dans ce chapitre, le modèle proposé ici est susceptible d'avoir des applications très diverses. Il peut en effet aussi bien représenter des systèmes minimaux que des systèmes très complets et peut à la fois être employé pour étudier des questions très fondamentales ou pour interpréter des données expérimentales. Ainsi, ce modèle ouvre de nombreuses perspectives de natures très variées.

5.1 Analyse commune de données de ChIP, de FRAP et de FRET

Une caractéristique particulière de notre modèle est qu'il permet de reproduire des données expérimentales obtenues par des techniques très différentes, en particulier les techniques de ChIP, de FRAP et de FRET, donnant chacune un point de vue différent sur la dynamique des complexes de régulation et d'activation transcriptionnelle (cf Box 2, page 52). Le modèle employant une description générique ne présupposant aucune structure au système, son utilisation pour l'interprétation commune et quantitative de ce type de données expérimentales semble donc être une application attrayante. Cependant, contrairement aux interprétations de mesures expérimentales que nous avons pu faire dans ce manuscrit (pour lesquelles nous avions par exemple présupposé un cycle global), l'analyse de données sans information *a priori* sur la structure du système constitue un défi de taille pour lequel nous ne pouvons ici qu'évoquer des pistes.

D'un point de vue computationnel, une stratégie d'analyse serait de s'appuyer sur la décomposition modale du système. En effet, nous avons donné une formulation des trois indicateurs que nous considérons ici sous la forme d'une somme de fonctions élémentaires, des exponentielles complexes, décrivant chacune un mode du système. On pourrait alors être tenté d'appliquer des techniques d'analyse multi-exponentielle (Istratov et Vyvenko, 1999; Kryzhniy, 2004) afin de déterminer les valeurs propres du système, puis d'employer des méthodes utilisées pour les *inverse eigenvalue problems* (problèmes inverses des valeurs propres) (Chu, 1998) afin d'en déduire la matrice $\mathbf{M}$ des taux de transition. Cependant, ces deux problèmes sont particulièrement délicats d'un point de vue computationnel (Chu, 1998; Istratov et Vyvenko, 1999) et peuvent difficilement s'appliquer ici. Il semble donc plus pertinent d'employer d'autres méthodes, ne passant pas par l'étape de la décompo-

sition modale, comme des techniques d'algèbre linéaire s'appuyant par exemple sur des pseudo-inversions de matrices.

Sur le plan expérimental, on peut également proposer plusieurs pistes pour extraire des informations pertinentes pour ce type d'analyses. Classiquement, le FRAP mesure les temps de résidence sur un ensemble de promoteurs désynchronisés et donc dans des étapes différentes de leur éventuelle progression lente. Une mesure permettant probablement d'extraire des informations importantes serait de réaliser des expériences de FRAP sur une population initialement synchonisée et à différents instants au cours de l'activation transcriptionnelle, comme dans les travaux de Karpova *et al.* (2008). Cela permettrait de connaître plus précisément l'évolution de la mobilité d'une protéine au cours de la progression du promoteur. D'autre part, la réalisation de cinétiques de ChIP dans différentes conditions, par exemple pour différentes concentrations d'un FT, permettrait également d'obtenir des informations importantes. En effet, dans notre modèle, la façon dont changent les propriétés dynamiques en fonction des concentrations est révélateur des la structure du système sous-jacent. Enfin, dans les approches de ChIP, certains auteurs ont effectué des expériences dites de Re-ChIP consistant à utiliser successivement deux anti-corps afin de déterminer si deux molécules peuvent être présentes simultanément sur un même promoteur (Métivier *et al.*, 2003). L'utilisation de combinaisons d'anti-corps ouvre d'importantes possibilités pour obtenir une description précise de l'état du promoteur. En particulier, il paraît possible, sur la base de notre modèle de déterminer des combinaisons optimales d'anti-corps (deux ou plus) permettant de révéler des informations importantes. Il semblerait alors pertinent d'utiliser ces combinaisons d'anti-corps dans le cadre de cinétiques de Re-ChIP.

5.2 Fluctuations extrinsèques et résonance stochastique

Dans la définition du modèle présenté dans ce chapitre ainsi que dans son utilisation, nous n'avons jusqu'ici pris en compte que les fluctuations gène-intrinsèques (cf section 1.2, chapitre I). En effet, toute fluctuation gène-extrinsèque a été éliminée *a priori* en considérant les concentrations de tous les FT constantes au cours du temps. Il est pourtant très probable que les promoteurs puissent démontrer de propriétés particulières face à des fluctuations extrinsèques. En effet, un phénomène bien connu en physique statistique et en traitement du signal peut avoir ici d'importantes implications : le phénomène de *résonance stochastique* (Hänggi, 2002; Lindner *et al.*, 2004; Rouvas-Nicolis et Nicolis, 2007; McDonnell et Abbott, 2009).

Dans un cadre général, on observe ce phénomène lorsque la présence de fluctuations aléatoires dans un système non linéaire améliore sa capacité à transmettre un signal (McDonnell et Abbott, 2009). Un cas typique relativement simple est celui d'un système bistable basculant stochastiquement d'un état à l'autre. Ce système présente spontanément un comportement apériodique, mais, lorsque les probabilités de transition varient périodiquement au cours du temps, on observe une composante périodique dans le comportement du système, et ce, de façon optimale pour une quantité intermédiaire de bruit (Rouvas-Nicolis et Nicolis, 2007). L'association/dissociation sur son site de fixation d'un FT dont la concentration varie au cours du temps est sujette à ce phénomène. Si l'on considère un

promoteur complexe dont la dynamique spontanée présente déjà diverses périodicités et diverses relaxations, on peut s'attendre à des propriétés particulières de transmission d'un signal (filtrage/amplification de certaines fréquences, non-additivité entre sources, ...), celles-ci étant éventuellement contrôlables par la concentration d'autres FT.

Bien que notre modèle suppose une quantité constante de tous les FT, il est cependant facile de considérer certains types de fluctuations extrinsèques afin d'étudier les propriétés de transmission du gène. En effet, on peut par exemple considérer dans un premier temps de manière simpliste que la concentration du FT A bascule de façon discrète et périodique d'une valeur à une autre. En combinant les exponentielles matricielles des deux matrices $\mathbf{M}$ correspondant à ces deux concentrations de $[A]$, on peut déterminer toutes les mesures de spectres de puissance et d'autocorrélations (voire également les corrélations aux ordres supérieures) décrivant comment le signal d'entrée est transmis par le gène. On peut de plus généraliser cette méthode à tout signal correspondant à un nombre fini de paliers[59] et étudier la transmission de fluctuations gène-extrinsèques en fonction de leurs caractéristiques spectrales. Cette méthode permet aussi de prendre en compte l'évolution simultanée de plusieurs concentrations de FT, ouvrant alors de nombreuses possibilités pour étudier comment des corrélations entre plusieurs concentrations (telles que décrites à la section 3.3, chapitre I) peuvent, dans un tel système, donner lieu à des propriétés constructives ou destructives à différentes fréquences. Comme l'illustre la figure III.21a, la position des valeurs propres de la matrice $\mathbf{M}$ change avec la concentration des FT. Ainsi, le comportement d'un gène face à des fluctuations au cours du temps de la concentration d'un FT dépend de l'ensemble des positions des valeurs propres correspondant à l'ensemble des valeurs prises par cette concentration. On peut alors s'attendre à des comportements particuliers en fonctions des trajectoires des valeurs propres (par exemple au niveau des bifurcations et dans les régions où la vitesse de déplacement est très variable), sachant de plus que celles-ci dépendent fortement de la concentration des autres FT.

Les propriétés de transmission de fluctuations gène-extrinsèques par un gène tel que nous l'avons décrit n'ont pas été étudiées dans le cadre de cette thèse. Il apparaît cependant clairement que le cadre théorique et les dérivations analytiques développées ici pour étudier le comportement spontané d'un gène sont tout à fait adaptés pour en étudier les propriétés de transmission d'un signal extrinsèque. De plus, nos travaux suggèrent fortement des propriétés particulières et donnent certaines pistes pour l'études de ces mécanismes.

5.3 Oscillateurs stochastiques couplés et spécificité du code des histones

Parce qu'il permet de représenter de façon générique les influences réciproques entre l'état de la chromatine et les enzymes provoquant des transitions entre ces états, ce modèle constitue un outil particulièrement adapté pour aborder des questions relatives au code des histones. Il permet en effet de décrire toute sorte de structure pour ce code, d'un

[59] C'est d'ailleurs toujours le cas lorsque les variations d'une concentration sont dues à la synthèse/dégradation des molécules dans un volume donné fixe. La généralisation à des signaux continus est possible mais fait intervenir une intégrale impossible à résoudre analytiquement dans le cas général.

système très contraint et combinatoire s'apparentant à un automate à base de règles à un système très peu spécifique où chaque modification peut avoir lieu et n'est que peu influencée par l'état courant de la chromatine. Les influences mutuelles entre FT et FC que nous avons implémentées dans l'exemple très simple de promoteur eucaryote présenté sur les figures III.4a1-3 correspondent à un code des histones simpliste et très contraint. En effet, nous considérons alors que l'état de la chromatine détermine de façon binaire la possibilité pour un FT de s'associer ou non. De plus, le système est construit de façon à ne provoquer qu'une unique séquence possible d'état chromatiniens. Cette vision d'un code des histones très spécifique et combinatoire n'est cependant probablement pas très réaliste et la nature de ce code reste un sujet de débat dans la littérature actuelle. Certaines études montrent par exemple que l'on peut expliquer les changements des propriétés biochimiques des histones suite à leurs modifications post-traductionnelles par un simple modèle cumulatif plutôt que par un principe de combinatoire complexe (Dion *et al.*, 2005; Henikoff, 2005).

Une vision sensiblement différente de celle du code très contraint que nous avons employé jusqu'ici peut cependant être développée. Elle permet une activité périodique sensiblement cohérente en se basant toutefois sur un code des histones très peu contraint et très peu combinatoire.

La modification d'un site d'une queue d'histone et sa modification inverse font intervenir au moins six événements (en comptant l'association et la dissociation des deux complexes impliqués[60]) dont deux sont énergie-dépendants. Pris isolément, chaque site d'une queue d'histone se comporte donc comme un oscillateur stochastique relativement peu cohérent puisqu'il s'agit d'un cycle orienté de six transitions (son spectre correspondant au mieux à un hexagone de valeurs propres). Si l'on considère sur un promoteur que ces réactions sont totalement indépendantes les unes des autres, le système correspond alors à un ensemble d'oscillateurs stochastiques non-couplés. On peut montrer, en se basant sur le principe de convolution des spectres illustré à la figure III.15 (page 147), que la cohérence d'un tel système est simplement la même que celle de son oscillateur le plus cohérent. Cependant, dans ce type de systèmes, un couplage, même faible, entre les différents oscillateurs stochastiques permet de les synchroniser et ce, d'autant plus facilement qu'ils sont nombreux (Lindner *et al.*, 2004). Ainsi, même si l'on ne considère qu'un seul nucléosome, les très nombreuses modifications possibles des queues d'histones semblent pouvoir conduire à une activité sensiblement cohérente sans nécessiter un code particulièrement structuré. En effet, un couplage entre les différents oscillateurs peut être réalisé très simplement sans faire appel à des interactions combinatoires mais, par exemple, par une simple exclusion stérique entre complexes réalisant les modifications de la chromatine.

Ainsi, un comportement oscillant cohérent pourrait être apparu relativement spontanément et sans faire appel à des mécanismes complexes. Même si nous n'allons pas jusqu'à proposer que le comportement périodique observé aujourd'hui sur les promoteurs eucaryotes ne reflète que ce mécanisme, il pourrait toutefois représenter un étape importante dans leur histoire évolutive. En effet, à partir d'un tel mécanisme simple, un compor-

[60]Par exemple, s'il s'agit d'un acétylation, les six étapes sont : l'association d'une HAT, l'acétylation du résidu, la dissociation de l'HAT, l'association d'une HDAC, la désacétylation du résidu et enfin la dissociation de l'HDAC.

tement périodique davantage cohérent peut être obtenu soit en augmentant le nombre d'oscillateurs (c'est-à-dire le nombre de sites permettant des modifications covalentes sur les queues des histones), soit en structurant les interactions entre molécules en y introduisant progressivement des aspects combinatoires. Un tel scénario pourrait alors expliquer l'apparition d'un code et il ne serait pas exclu, dans ce cas, que le code actuel garde d'importantes traces de cette histoire évolutive et conserve notamment un fort degré d'aspécificité.

Conclusion et perspectives

Bien que clairement établi expérimentalement, le caractère stochastique de l'expression des gènes pose de nombreuses questions. Qu'il s'agisse de son origine moléculaire ou de son rôle dans le fonctionnement des systèmes vivants, l'étude de ce phénomène constitue aujourd'hui un sujet de recherche actif et très pluridisciplinaire. Dans ce cadre, la modélisation est devenue un outil essentiel, permettant à la fois d'interpréter des données expérimentales, de formuler des hypothèses théoriques et de proposer des expériences humides afin de les tester. Cependant, dans ce domaine, les travaux de modélisation se sont essentiellement concentrés sur les effets liés aux événements de synthèse/dégradation des transcrits et des protéines et au basculement du promoteur entre deux états, supposant implicitement ou explicitement une dynamique simple des différentes étapes du processus d'expression génique (diffusion des facteurs de transcription, association/dissociation de ces facteurs sur le promoteur, export des transcrits, ...). Or, si ces mécanismes simples peuvent constituer des sources importantes de stochasticité, ce ne sont clairement pas les seuls. En particulier, nous avons montré au chapitre I que la combinatoire des événements moléculaires sur le promoteur ou l'architecture nucléaire peuvent y contribuer significativement. C'est pourquoi nous avons proposé deux modèles permettant d'étudier, à des échelles spatio-temporelles très différentes, l'influence de la dynamique stochastique des complexes macromoléculaires sur le processus d'expression génique. Ces deux modèles ont permis de montrer que la structure de la stochasticité de l'expression d'un gène pouvait être influencée par des mécanismes à de nombreux niveaux : depuis les contraintes spatiales à l'échelle du noyau cellulaire, jusqu'à la dynamique de formation des complexes sur le promoteur d'un gène.

Afin d'étudier les phénomènes d'organisation spatiale au sein de l'espace nucléaire en relation avec le processus d'expression génique, nous avons proposé un premier modèle, basé sur les caractéristiques essentielles que sont la nature discrète des macromolécules, leur déplacement aléatoire dans l'espace et leurs interactions à la fois physique et biochimique. Bien que le logiciel de simulation que nous avons développé dans ce but ne soit pas encore exploitable et que l'approche elle-même comporte des limites évidentes, la démarche de conception du modèle – consistant à la fois en un travail bibliographique et une réflexion sur l'objet d'étude – nous a amené à formuler des hypothèses et à proposer des mécanismes potentiellement impliqués dans la génération d'une stochasticité structurée de l'expression des gènes. Certains de ces mécanismes ne peuvent toutefois pas être étudiés avec des outils de simulation tels que celui présenté ici et nécessitent des modèles basés sur des formalismes différents et prenant en compte plus spécifiquement certains aspects des interactions moléculaires. De fait, la régulation et l'activation transcriptionnelle au

niveau d'un promoteur mettent en jeu de très nombreuses molécules, interagissant selon des mécanismes combinatoires complexes dans lesquels intervient en particulier l'état conformationnel et covalent de la chromatine. Avec un modèle de simulation spatiale tel que celui décrit précédemment, un tel système ne peut être étudié qu'au prix de très fortes simplifications qui, selon toute vraisemblance, masqueraient une grande part du phénomène étudié. Nous avons ainsi adopté un approche sensiblement différente pour l'étude de la dynamique stochastique des complexes de régulation transcriptionnelle en développant un second modèle basé sur un formalisme permettant la représentation de promoteurs arbitrairement complexes et exploitable par des dérivations analytiques.

Comme nous avons pu le mettre en évidence dans le chapitre I en explorant différents champs de la littérature, un promoteur est un objet au comportement complexe, présentant de nombreuses facettes et pouvant être observé selon plusieurs points de vue. D'une part, il s'agit d'un système stochastique basculant de façon probabiliste entre différents états et pouvant constituer une source majeure de stochasticité dans l'expression d'un gène. Mais, d'autre part, un promoteur est un système présentant une dynamique multi-échelle caractérisée à la fois par des temps de résidence particulièrement courts de la plupart des molécules pouvant s'y associer et par un comportement lent et souvent périodique dans lequel les molécules ont des cinétiques de présence très caractéristiques. Un attrait majeur de notre modèle est qu'il permet de rapprocher dans un cadre commun ces caractéristiques (qui correspondent à des corpus de recherche différents) et les techniques expérimentales associées (offrant chacune un point de vue différent sur le même objet). Nous avons en effet dérivé des indicateurs reproduisant les mesures classiquement employées pour caractériser la stochasticité de l'expression génique (obtenues par des techniques de cytométrie et de vidéo-microscopie) et celles permettant d'observer la dynamique des complexes de régulation transcriptionnelle chez les eucaryotes à différentes échelles temporelles (des mesures de FRAP et de FRET, reflétant la mobilité des molécules, et des cinétiques de ChIP, décrivant la dynamique lente d'occupation du promoteur). En isolant la structure mathématique commune à ces différentes mesures (la décomposition modale du système), il est alors possible de comprendre leurs relations, de dresser certains parallèles entre les mesures expérimentales et théoriques (FRAP *vs* distribution des temps de résidence, cinétiques ChIP *vs* distribution de durée du cycle, ...) et de concilier des points de vue expérimentaux différents en montrant leur complémentarité plutôt que leur opposition (FRAP *vs* cinétiques ChIP).

La reproduction de périodicités globales lentes et régulières – telles qu'observées expérimentalement – sur la base d'une description locale des événements moléculaires discrets et stochastiques révèle à la fois des caractéristiques microscopiques et macroscopiques du système. Ce comportement périodique lent se base sur la progression cyclique de la chromatine du promoteur selon différents états méta-stables entre lesquels les transitions sont provoquées par des complexes macromoléculaires spécifiques. Nous avons ainsi pu étudier à la fois les aspects relatifs au nombre global d'étapes de cette progression cyclique et à la dynamique locale propre d'assemblage/désassemblage des complexes macromoléculaires sur le promoteur. L'interprétation de données expérimentales de cinétiques de ChIP nous a alors permis d'estimer une borne inférieure pour la taille du cycle et pour sa consommation énergétique. Ces estimations, obtenues pour des promoteurs relativement différents, semblent être cohérentes avec les caractéristiques structurelles des systèmes considérés

(promoteur minimaliste *in vitro* dans un cas et promoteur endogène sur cellules vivantes dans un autre). De plus, nous avons comparé différents modes d'interaction des molécules entre elles et sur la chromatine au regard de leur influence sur la dynamique globale lente. De façon remarquable, le mode d'interaction le plus plausible de ce point de vue est également celui permettant de reproduire une propriété relativement singulière, mise en évidence avec une technique de FRAP : la réduction significative de la mobilité des molécules sur la chromatine suite à une déplétion de la cellule en ATP. Ce modèle permet donc clairement d'étudier les relations entre les différentes échelles de temps de la dynamique du promoteur et en lien avec les données expérimentales.

Sur un plan plus fondamental, ce modèle nous a également permis de mettre en évidence un certain nombre de comportements inattendus. En particulier, nous avons vu qu'un promoteur permettait de générer une stochasticité structurée de l'expression d'un gène, aussi bien sur le plan de ses propriétés stationnaires que dynamiques. Notre étude a de plus révélé qu'un contrôle fin des caractéristiques dynamiques semblait nécessiter un certain degré de désordre dans la structure des interactions au sein du système. En effet, une activité périodique significativement régulière mais soumise à un contrôle précis et complexe ne semble pouvoir être obtenue que dans une situation intermédiaire entre un système très structuré (où les événements ont lieu dans un ordre précis) et un système totalement désordonné (où tout événement peut arriver à tout instant). Cette propriété n'est pas sans rappeler la notion de « frontière du chaos », bien connue en sciences de la complexité (Langton, 1986, 1991; Flake, 1998; Wolfram, 2002; Abel, 2009). Il a en effet été montré sur des modèles très abstraits que les conditions propices à l'émergence de comportements complexes correspondent généralement à des régimes de fonctionnement à la frontière entre ordre et désordre. Si un certain paramètre est capable d'influencer le comportement d'un système dynamique donné en provoquant, pour certaines plages de valeurs, des comportements simples et stéréotypés (attracteurs stables ou périodiques) et, pour d'autres, des comportements très irréguliers (attracteurs chaotiques), c'est alors pour des valeurs critiques de ce paramètre, à la frontière entre ces deux régimes, que l'on peut s'attendre à découvrir des comportements complexes. De façon remarquable, c'est à cette frontière que semblent se situer les promoteurs eucaryotes. En effet, la redondance fonctionnelle des complexes de modification de la chromatine au cours de la progression cyclique lente, ainsi que la forte mobilité des molécules et l'existence de nombreuses compositions alternatives pour les complexes constitue, dans notre modèle, les propriétés permettant un tel comportement complexe.

Alors que, depuis des années, la grande majorité des approches de modélisation en biologie des systèmes se focalisent sur les réseaux de régulation, cherchant les propriétés complexes pouvant émerger de l'interaction d'un ensemble de gènes et considérant souvent ces derniers comme des entités très simples et déterministes, nos travaux montrent que le gène, élément de base des réseaux de régulation, constitue en lui-même un système au comportement complexe, pouvant faire preuve d'activités dynamiques et stochastiques très variées.

Qu'il s'agisse des modèles en eux-mêmes ou des conclusions auxquelles ils nous ont conduit, nos travaux ouvrent de nombreuses pistes d'études. Bien entendu, les modèles sont toujours susceptibles d'extension ou d'amélioration. Ici, une perspective immédiate serait

bien sûr la finalisation du simulateur spatial (en particulier pour introduire un modèle de réactions). Par ailleurs, des utilisations conjointes des deux modèles présentés dans ce manuscrit peuvent être envisagées. On peut par exemple étudier, à l'aide du modèle spatial présenté au chapitre II, les caractéristiques de la distribution de durée d'export des transcrits hors du noyau (dont nous avons vu au chapitre III comment elle permettait de moduler significativement la transmission des fluctuation du promoteur) et ce, en particulier dans le cadre des mécanismes de rétention des transcrits observés dans le noyau. Une autre application commune de ces deux modèles pourrait consister à relâcher l'approximation markovienne du modèle de promoteur (qui revient à considérer un espace implicitement homogène et des molécules en grandes quantités) et de se baser sur des distributions de temps de réassociation décrites à l'aide du modèle spatial.

Au delà de l'amélioration des modèles, il nous semble, à court terme, plus pertinent de rechercher un couplage étroit avec la biologie expérimentale. Ainsi, comme nous l'avons vu en perspective du chapitre III, on peut envisager des applications directes du modèle de promoteur à des problèmes très appliqués (comme l'analyse quantitative de données expérimentales de ChIP et de FRAP obtenues sur un même promoteur). Mais ce modèle permet également de s'intéresser à de nombreuses questions fondamentales. Il constitue par exemple un outil très adapté pour étudier le code des histones, permettant de représenter de façon générique toutes sortes d'influences réciproques entre la chromatine et les enzymes qui la modifient. Il ouvre en particulier des interrogations quant à la nature et à l'origine de ce code. L'idée d'un code stéréotypé et combinatoire (comme un automate obéissant à un ensemble de règles, tel qu'il a d'abord été perçu) est en effet actuellement sujet à débat dans la littérature. Comme nous l'avons suggéré, une partie de ce code (celui opérant lors de l'activation transcriptionnelle) pourrait avoir émergé d'un simple mécanisme de couplage entre des modifications stochastiques de différents résidus des queues d'histone par des interactions très simples et très peu combinatoires. Notre modèle pourrait ainsi apporter un éclairage nouveau sur certaines questions relatives à la structure de ce code, en relation avec son histoire évolutive. En effet, une caractéristique de ce modèle qui nous semble remarquable et mérite d'être soulignée, est sa capacité non seulement à poser de nombreuses questions fondamentales, mais à amener spontanément le modélisateur à replacer l'objet d'étude dans son contexte évolutif. Ainsi, lorsque nous avons cherché à caractériser les paramètres influençant la cohérence des oscillations d'une cinétique de ChIP, il nous est apparu que la cellule doit faire face à de nombreux compromis : entre la taille et la consommation énergétique du cycle, entre le nombre de molécules et la spécificité de leurs interactions, entre le degré d'organisation du système et la complexité de son comportement ... De plus, ces compromis indiquent souvent, pour les différents paramètres qu'ils mettent en jeu, des plages de valeurs hors desquelles il semble très peu probable de trouver le système, nous amenant alors à considérer les coûts adaptatifs que représentent chacun de ces paramètres (consommation énergétique, nombre de molécules à synthétiser, ...). Même si la question centrale de l'intérêt d'un comportement périodique reste encore essentiellement ouverte, nos résultats montrent qu'un tel comportement ne peut apparaître qu'au sein de structures complexes et faisant intervenir un très grand nombre de molécules, ce qui correspond de fait aux structures de régulation observées chez les eucaryotes et permettrait donc d'expliquer pourquoi l'évolution a conduit à une telle complexité.

Bibliographie

Abel, D. L. (2009). The capabilities of chaos and complexity. *Int. J. Mol. Sci.*, 10:247–91.

Abramowitz, M. et Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York, 9-ième édition.

Abyzov, A., Bjornson, R., Felipe, M. et Gerstein, M. (2010). RigidFinder: a fast and sensitive method to detect rigid blocks in large macromolecular complexes. *Proteins*, 78(2):309–324.

Ackers, G., Johnson, A. et Shea, M. (1982). Quantitative model for gene regulation by lambda phage repressor. *Proc. Natl. Acad. Sci. U.S.A.*, 79(2):1129–1133.

Adams, C. et Workman, J. (1995). Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol. Cell. Biol.*, 15(3):1405–1421.

Agresti, A., Scaffidi, P., Riva, A., Caiolfa, V. et Bianchi, M. (2005). GR and HMGB1 interact only within chromatin and influence each other's residence time. *Mol. Cell*, 18(1):109–121.

Allen, M. P. et Tildesley, D. J. (1991). *Computer Simulation of Liquids.* Oxford University Press, 1st édition.

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., Le Sceller, L., Loutellier, C., Michel, O., Molina, F., Monnier, C., Natowicz, R., Norris, V., Orange, N., Pollard, H., Raine, D., Ripoll, C., Rouviere-Yaniv, J., Saier, M., Soler, P., Tambourin, P., Thellier, M., Tracqui, P., Ussery, D., Vincent, J. C., Vannier, J. P., Wiggins, P. et Zemirline, A. (2002). Hyperstructures, genome analysis and I-cells. *Acta Biotheor.*, 50(4):357–373.

Amar, P., Bernot, G. et Norris, V. (2004). HSIM: a simulation programme to study large assemblies of proteins. *J. Biol. Phys. Chem.*, 4(2):79–84.

Amar, P., Legent, G., Thellier, M., Ripoll, C., Bernot, G., Nystrom, T., Saier, M. H. et Norris, V. (2008). A stochastic automaton shows how enzyme assemblies may contribute to metabolic efficiency. *BMC Syst. Biol.*, 2(27).

Andrews, S. S., Addy, N. J., Brent, R. et Arkin, A. P. (2010). Detailed simulations of cell biology with Smoldyn 2.1. *PLoS Comput. Biol.*, 6(3):e1000705.

Andrews, S. S. et Arkin, A. P. (2006). Simulating cell biology. *Curr. Biol.*, 16(14): R523–527.

Andrews, S. S. et Bray, D. (2004). Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Phys. Biol.*, 1(3):137–151.

Arya, G. et Schlick, T. (2006). Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proc. Natl. Acad. Sci. U.S.A.*, 103(44):16236–41.

Austin, D., Allen, M., McCollum, J., Dar, R., Wilgus, J., Sayler, G., Samatova, N., Cox, C. et Simpson, M. (2006). Gene network shaping of inherent noise spectra. *Nature*, 439(7076):608–611.

Ballet, P., Zemirline, A. et Marcé, L. (2004). The BioDyn language and simulator. application to an immune response and E. coli and phage interaction. *J. Biol. Phys. Chem.*, 4(2):93–101.

Bancaud, A., Huet, S., Daigle, N., Mozziconacci, J., Beaudouin, J. et Ellenberg, J. (2009). Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. *EMBO J.*, 28(24):3785–3798.

Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y. et Barkai, N. (2006). Noise in protein expression scales with natural protein abundance. *Nat. Genet.*, 38(6):636–643.

Beard, D. A. et Schlick, T. (2000). Inertial stochastic dynamics. I. Long-time-step methods for Langevin dynamics. *J. Chem. Phys.*, 12(17):7313–21.

Becker, M., Baumann, C., John, S., Walker, D., Vigneron, M., McNally, J. et Hager, G. (2002). Dynamic behavior of transcription factors on a natural promoter in living cells. *EMBO Rep.*, 3(12):1188–1194.

Becskei, A., Kaufmann, B. et van Oudenaarden, A. (2005). Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nat. Genet.*, 37(9):937–944.

Becskei, A. et Serrano, L. (2000). Engineering stability in gene networks by autoregulation. *Nature*, 405(6786):590–593.

Bedau, M. A. (2003). Artificial life: organization, adaptation and complexity from the bottom up. *Trends Cogn. Sci.*, 7(11):505–512.

Bedau, M. A., McCaskill, J. S., Packard, N. H., Rasmussen, S., Adami, C., Green, D. G., Ikegami, T., Kaneko, K. et Ray, T. S. (2000). Open problems in artificial life. *Artif. Life*, 6(4):363–376.

Belle, A., Tanay, A., Bitincka, L., Shamir, R. et O'Shea, E. K. (2006). Quantification of protein half-lives in the budding yeast proteome. *Proc. Natl. Acad. Sci. U.S.A.*, 103(35):13004–13009.

Benecke, A. (2006). Chromatin code, local non-equilibrium dynamics, and the emergence of transcription regulatory programs. *Eur. Phys. J. E Soft Matter*, 19(3):353–366.

Berg, H. C. (1993). *Random Walks in Biology.* Princeton University Press, 2nd édition.

Berg, J. M., Tymoczko, J. L. et Stryer, L. (2002). *Biochemistry.* W. H. Freeman and Company, New York, 5th édition.

Bernaschi, M., Castiglione, F., Ferranti, A., Gavrila, C., Tinti, M. et Cesareni, G. (2007). ProtNet: a tool for stochastic simulations of protein interaction networks dynamics. *BMC Bioinformatics*, 8 Suppl 1:S4.

Bernini, A., Spiga, O., Ciutti, A., Chiellini, S., Menciassi, N., Venditti, V. et Niccolai, N. (2004). On the dynamics of water molecules at the protein solute interfaces. *Homeopathy*, 93:199–202.

Bernstein, J. A., Khodursky, A. B., Lin, P. H., Lin-Chao, S. et Cohen, S. N. (2002). Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. U.S.A.*, 99(15):9697–9702.

Berry, H. (2002). Monte carlo simulations of enzyme reactions in two dimensions: fractal kinetics and spatial segregation. *Biophys. J.*, 83(4):1891–1901.

Beslon, G. (2008). *Apprivoiser la vie ; modélisation individu-centrée de systèmes biologiques complexes.* Habilitation à Diriger des Recherches, INSA-Lyon.

Bintu, L., Buchler, N., Garcia, H., Gerland, U., Hwa, T., Kondev, J. et Phillips, R. (2005a). Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.*, 15(2):116–124.

Bintu, L., Buchler, N., Garcia, H., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T. et Phillips, R. (2005b). Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.*, 15(2):125–135.

Blake, W. J., Balázsi, G., Kohanski, M. A., Isaacs, F. J., Murphy, K. F., Kuang, Y., Cantor, C. R., Walt, D. R. et Collins, J. J. (2006). Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell*, 24(6):853–865.

Blake, W. J., Kærn, M., Cantor, C. R. et J., C. J. (2003). Noise in eukaryotic gene expression. *Nature*, 422(6932):633–7.

Boeger, H., Griesenbeck, J. et Kornberg, R. D. (2008). Nucleosome retention and the stochastic nature of promoter chromatin remodeling for transcription. *Cell*, 133(4):716–726.

Boettiger, A. N. et Levine, M. (2009). Synchronous and stochastic patterns of gene activation in the Drosophila embryo. *Science*, 325(5939):471–3.

Bork, P. et Serrano, L. (2005). Towards cellular systems in 4D. *Cell*, 121:507–509.

Bosisio, D., Marazzi, I., Agresti, A., Shimizu, N., Bianchi, M. et Natoli, G. (2006). A hyper-dynamic equilibrium between promoter-bound and nucleoplasmic dimers controls NF-kappaB-dependent gene activity. *EMBO J.*, 25(4):798–810.

Branco, M. R. et Pombo, A. (2006). Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.*, 4(5):e138.

Brock, A., Chang, H. et Huang, S. (2009). Non-genetic heterogeneity–a mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.*, 10(5):336–342.

Brooks, S. P. J. (2004). Enzymes: The basis of catalysis. *In* Storey, K. B., éditeur : *Functional Metabolism: Regulation and Adaptation*, chapitre 2, pages 25–53. Wiley, Hoboken, New Jersey.

Browning, D. et Busby, S. (2004). The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.*, 2(1):57–65.

Buchete, N.-V., Straub, J. E. et Thirumalai, D. (2004). Orientation-dependent coarse-grained potentials derived by statistical analysis of molecular structural databases. *Polymer*, 45(2):597–608.

Burton, K. (1955). The free energy change associated with the hydrolysis of the thiol ester bond of acetyl coenzyme A. *Biochem. J.*, 59(1):44–46.

Çağatay, T., Turcotte, M., Elowitz, M. B., Garcia-Ojalvo, J. et Süel, G. M. (2009). Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell*, 139(3):512–522.

Cai, L., Friedman, N. et Xie, X. (2006). Stochastic protein expression in individual cells at the single molecule level. *Nature*, 440(7082):358–362.

Cairns, B. R. (2009). The logic of chromatin architecture and remodelling at promoters. *Nature*, 461(7261):193–198.

Carrero, G., Hendzel, M. J. et de Vries, G. (2006). Modelling the compartmentalization of splicing factors. *J. Theor. Biol.*, 239(3):298–312.

Chabrier-Rivier, N., Fages, F. et Soliman, S. (2005). The biochemical abstract machine biocham. *In Computational Methods in Systems Biology*, volume 3082, pages 172–91. Springer.

Chang, H., Hemberg, M., Barahona, M., Ingber, D. et Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*, 453(7194):544–547.

Chen, J. C. et Kim, A. S. (2004). Brownian Dynamics, Molecular Dynamics, and Monte Carlo modeling of colloidal systems. *Adv. Colloid Interface Sci.*, 112(1-3):159–173.

Chevereau, G., Palmeira, L., Thermes, C., Arneodo, A. et Vaillant, C. (2009). Thermodynamics of intragenic nucleosome ordering. *Phys. Rev. Lett.*, 103(18):188103.

Chu, M. T. (1998). Inverse eigenvalue problems. *SIAM Rev.*, 40(1):1–39.

Chubb, J., Trcek, T., Shenoy, S. et Singer, R. (2006). Transcriptional pulsing of a developmental gene. *Curr. Biol.*, 16(10):1018–1025.

Chung, F. R. K. (1997). *Spectral Graphs Theory.* American Mathematical Society, Providence, Rhode Island.

Colman-Lerner, A., Gordon, A., Serra, E., Chin, T., Resnekov, O., Endy, D., Pesce, C. G. et Brent, R. (2005). Regulated cell-to-cell variation in a cell-fate decision system. *Nature*, 437(7059):699–706.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz Jr., K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. et Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117(9):5179–5197.

Coulon, A., Beslon, G., Chatelain, F., Fuchs, A., Gandrillon, O., Gineste, M., Kupiec, J.-J., Mejia-Perez, C. et Paldi, A. (2009). Mécanismes moléculaires et fonction biologique de la variabilité de l'expression génique à l'échelle de la cellule unique : une approche systémique. *In Le Hasard au cœur de la cellule*, chapitre 3, pages 59–78. Syllepse.

Coulon, A., Beslon, G. et Gandrillon, O. (2008). Large multi-protein structures modeling and simulation: the need for mesoscopic models. *In* Thompson, J., éditeur : *Functional Proteomics*, volume 484 de *Methods in Molecular Biology*, chapitre 32, pages 537–558. Humana press.

Coulon, A., Gandrillon, O. et Beslon, G. (2010). On the spontaneous stochastic dynamics of a single gene: complexity of the molecular interplay at the promoter. *BMC Syst Biol*, 4:2.

Coulon, A., Soula, H., Mazet, O., Gandrillon, O. et Beslon, G. (2007). Modélisation cellulaire pour l'émergence de structures multiprotéiques auto-organisées. *Tech. Sci. Info.*, 26:123–148.

Cox, C. D., McCollum, J. M., Allen, M. S., Dar, R. D. et Simpson, M. L. (2008). Using noise to probe and characterize gene circuits. *Proc. Natl. Acad. Sci. U.S.A.*, 105(31):10809–10814.

Cox, C. D., McCollum, J. M., Austin, D. W., Allen, M. S., Dar, R. D. et Simpson, M. L. (2006). Frequency domain analysis of noise in simple gene circuits. *Chaos*, 16(2):026102.

Cranck, J. (1975). *The Mathematics of Diffusion.* Oxford University Press, 2nd édition.

Cremer, T. et Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.*, 2(4):292–301.

Cremer, T., Cremer, M., Dietzel, S., Müller, S., Solovei, I. et Fakan, S. (2006). Chromosome territories–a functional nuclear landscape. *Curr. Opin. Cell Biol.*, 18(3): 307–316.

Cvetković, D., Doob, M. et Sachs, H. (1979). *Spectra of Graphs – Theory and Application.* Accademic Press, New York.

Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S. M., Phair, R. D. et Singer, R. H. (2007). In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.*, 14(9):796–806.

Darzacq, X. et Singer, R. H. (2008). The dynamic range of transcription. *Mol. Cell*, 30(5):545–546.

Darzacq, X., Singer, R. H. et Shav-Tal, Y. (2005). Dynamics of transcription and mRNA export. *Curr. Opin. Cell Biol.*, 17(3):332–339.

Darzacq, X., Yao, J., Larson, D. R., Causse, S. Z., Bosanac, L., de Turris, V., Ruda, V. M., Lionnet, T., Zenklusen, D., Guglielmi, B., Tjian, R. et Singer, R. H. (2009). Imaging transcription in living cells. *Annu. Rev. Biophys.*, 38:173–196.

Degenhardt, T., Rybakova, K. N., Tomaszewska, A., Moné, M. J., Westerhoff, H. V., Bruggeman, F. J. et Carlberg, C. (2009). Population-level transcription cycles derive from stochastic timing of single-cell transcription. *Cell*, 138(3):489–501.

Delarue, M. et Dumas, P. (2004). On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc. Natl. Acad. Sci. U.S.A.*, 101(18):6957–6962.

Dion, M. F., Altschuler, S. J., Wu, L. F. et Rando, O. J. (2005). Genomic characterization reveals a simple histone H4 acetylation code. *Proc. Natl. Acad. Sci. U.S.A.*, 102(15):5501–5506.

Dittrich, P., Ziegler, J. et Banzhaf, W. (2001). Artificial chemistries–a review. *Artif. Life*, 7(3):225–275.

Dobrzyński, M. et Bruggeman, F. J. (2009). Elongation dynamics shape bursty transcription and translation. *Proc. Natl. Acad. Sci. U.S.A.*, 106(15):2583–2588.

Dodd, I., Shearwin, K., Perkins, A., Burr, T., Hochschild, A. et Egan, J. (2004). Cooperativity in long-range gene regulation by the lambda CI repressor. *Genes Dev.*, 18(3):344–354.

Doruker, P., Jernigan, R. L. et Bahar, I. (2002). Dynamics of large proteins through hierarchical levels of coarse-grained structures. *J. Comput. Chem.*, 23(1):119–127.

Dublanche, Y., Michalodimitrakis, K., Kümmerer, N., Foglierini, M. et Serrano, L. (2006). Noise in transcription negative feedback loops: simulation and experimental analysis. *Mol. Syst. Biol.*, 2(41).

Dundr, M., Hoffmann-Rohrer, U., Hu, Q., Grummt, I., Rothblum, L., Phair, R. et Misteli, T. (2002). A kinetic framework for a mammalian RNA polymerase in vivo. *Science*, 298(5598):1623–1626.

Dundr, M., Misteli, T. et Olson, M. O. (2000). The dynamics of postmitotic reassembly of the nucleolus. *J. Cell Biol.*, 150(3):433–446.

Dunlop, M., Cox, R., Levine, J., Murray, R. et Elowitz, M. (2008). Regulatory activity revealed by dynamic correlations in gene expression noise. *Nat. Genet.*, 40(12): 1493–1498.

Echeverría, C., Tucci, K. et Kapral, R. (2007). Diffusion and reaction in crowded environments. *J. Phys. Condens. Matter*, 19(6):65146.

Elbi, C., Walker, D. A., Romero, G., Sullivan, W. P., Toft, D. O., Hager, G. L. et DeFranco, D. B. (2004). Molecular chaperones function as steroid receptor nuclear mobility factors. *Proc. Natl. Acad. Sci. U.S.A.*, 101(9):2876–2881.

Elf, J., Li, G. W. et Xie, X. S. (2007). Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, 316(5828):1191–1194.

Ellis, R. J. (2001). Macromolecular crowding: obvious but underappreciated. *Trends Biochem. Sci.*, 26(10):597–604.

Elowitz, M., Levine, A., Siggia, E. et Swain, P. (2002). Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186.

Eskiw, C. H., Dellaire, G., Mymryk, J. S. et Bazett-Jones, D. P. (2003). Size, position and dynamic behavior of PML nuclear bodies following cell stress as a paradigm for supramolecular trafficking and assembly. *J. Cell. Sci.*, 116(21):4455–4466.

Ferrara, P., Apostolakis, J. et Caflisch, A. (2002). Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins*, 46(1):24–33.

Fischle, W., Wang, Y. et Allis, C. (2003). Histone and chromatin cross-talk. *Curr. Opin. Cell Biol.*, 15(2):172–183.

Flake, G. W. (1998). *The Computational Beauty of Nature*. MIT Press, 1st édition.

Fletcher, T. M., Xiao, N., Mautino, G., Baumann, C. T., Wolford, R., Warren, B. S. et Hager, G. L. (2002). ATP-dependent mobilization of the glucocorticoid receptor during chromatin remodeling. *Mol. Cell. Biol.*, 22(10):3255–3263.

Ford, E. et Thanos, D. (2009). Time's up: bursting out of transcription. *Cell*, 138(3): 430–432.

Fraser, H. B., Hirsh, A. E., Giaever, G., Kumm, J. et Eisen, M. B. (2004). Noise minimization in eukaryotic gene expression. *PLoS Biol.*, 2(6):e137.

Fraser, P. et Bickmore, W. (2007). Nuclear organization of the genome and the potential for gene regulation. *Nature*, 447(7143):413–417.

Frenkel, D. et Smit, B. (2002). *Understanding Molecular Simulation.* Academic Press, 2nd édition.

Friedman, N., Cai, L. et Xie, X. S. (2006). Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett.*, 97(16): 168302.

Füchslin, R. M., Maeke, T. et McCaskill, J. S. (2009). Spatially resolved simulations of membrane reactions and dynamics: multipolar reaction DPD. *Eur. Phys. J. E Soft Matter*, 29(4):431–448.

Galkin, O., Chen, K., Nagel, R. L., Hirsch, R. E. et Vekilov, P. G. (2002). Liquid-liquid separation in solutions of normal and sickle cell hemoglobin. *Proc. Natl. Acad. Sci. U.S.A.*, 99(13):8479–8483.

Gillespie, D. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361.

Golding, I. et Cox, E. (2006). Eukaryotic transcription: what does it mean for a gene to be 'on'? *Curr. Biol.*, 16(10):R371–373.

Golding, I., Paulsson, J., Zawilski, S. et Cox, E. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell*, 123(6):1025–1036.

Görisch, S. M., Wachsmuth, M., Ittrich, C., Bacher, C. P., Rippe, K. et Lichter, P. (2004). Nuclear body movement is determined by chromatin accessibility and dynamics. *Proc. Natl. Acad. Sci. U.S.A.*, 101(36):13221–13226.

Gorski, S. A., Dundr, M. et Misteli, T. (2006). The road much traveled: trafficking in the cell nucleus. *Curr. Opin. Cell Biol.*, 18(3):284–290.

Gorski, S. A., Snyder, S. K., John, S., Grummt, I. et Misteli, T. (2008). Modulation of RNA polymerase assembly dynamics in transcriptional regulation. *Mol. Cell*, 30(4): 486–497.

Gowers, D. M., Wilson, G. G. et Halford, S. E. (2005). Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc. Natl. Acad. Sci. U.S.A.*, 102(44):15883–15888.

Grillo, I. (2003). Small-angle neutron scattering study of a world-wide known emulsion: Le Pastis. *Colloid Surface A*, 225:153–160.

Grima, R. et Schnell, S. (2006). A systematic investigation of the rate laws valid in intracellular environments. *Biophys. Chem.*, 124(1):1–10.

Groot, R. D. et Warren, P. B. (1997). Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation. *J. Chem. Phys.*, 107(11):4423–35.

Guigas, G. et Weiss, M. (2008). Sampling the cell with anomalous diffusion - the discovery of slowness. *Biophys. J.*, 94(1):90–94.

Hager, G., Elbi, C., Johnson, T., Voss, T., Nagaich, A., Schiltz, R., Qiu, Y. et John, S. (2006). Chromatin dynamics and the evolution of alternate promoter states. *Chromosome Res.*, 14(1):107–116.

Hager, G. L., Elbi, C. et Becker, M. (2002). Protein dynamics in the nuclear compartment. *Curr. Opin. Genet. Dev.*, 12(2):137–141.

Hager, G. L., McNally, J. G. et Misteli, T. (2009). Transcription dynamics. *Mol. Cell*, 35(6):741–753.

Hairer, E. (2005). Important aspects of geometric numerical integration. *J. Sci. Comput.*, 25(1/2):67–81.

Hairer, E., Lubich, C. et Wanner, G. (2003). Geometric numerical integration illustrated by the Störmer/Verlet method. *Acta Numerica*, 12:399–450.

Halford, S. et Marko, J. (2004). How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.*, 32(10):3040–3052.

Halford, S. E. (2009). An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem. Soc. Trans.*, 37:343–348.

Hancock, R. (2004). A role for macromolecular crowding effects in the assembly and function of compartments in the nucleus. *J. Struct. Biol.*, 146(3):281–290.

Handwerger, K. E., Cordero, J. A. et Gall, J. G. (2005). Cajal bodies, nucleoli, and speckles in the Xenopus oocyte nucleus have a low-density, sponge-like structure. *Mol. Biol. Cell*, 16(1):202–211.

Handwerger, K. E. et Gall, J. G. (2006). Subnuclear organelles: new insights into form and function. *Trends Cell Biol.*, 16(1):19–26.

Handwerger, K. E., Wu, Z., Murphy, C. et Gall, J. G. (2002). Heat shock induces mini-Cajal bodies in the Xenopus germinal vesicle. *J. Cell. Sci.*, 115(10):2011–2020.

Hänggi, P. (2002). Stochastic resonance in biology. How noise can enhance detection of weak signals and help improve biological information processing. *Chemphyschem*, 3(3):285–290.

Head-Gordon, T. et Brown, S. (2003). Minimalist models for protein folding and design. *Curr. Opin. Struct. Biol.*, 13(2):160–167.

Hearon, J. Z. (1953). The kinetics of linear systems with special reference to periodic reactions. *Bull. Math. Biol.*, 15:121–141.

Henikoff, S. (2005). Histone modifications: combinatorial complexity or cumulative simplicity? *Proc. Natl. Acad. Sci. U.S.A.*, 102(15):5308–5309.

Henikoff, S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. *Nat. Rev. Genet.*, 9(1):15–26.

Hobza, P., Kabeláč, M., Šponer, J., Mejzlík, P. et Vondraśěk, J. (1997). Performance of empirical potentials (AMBER, CFF95, CVFF, CHARMM, OPLS, POLTEV), semiempirical quantum chemical methods (AM1, MNDO/M, PM3), and ab initio Hartree-Fock method for interaction of DNA bases: Comparison with nonempirical beyond Hartree-Fock results. *J. Comput. Chem.*, 18(9):1136–1150.

Hołyst, R., Błażejczyk, M., Burdzy, K., Góralski, G. et Bocquet, L. (2000). Reduction of dimensionality in a diffusion search process and kinetics of gene expression. *Physica A*, 277(1-2):71–82.

Hornos, J., Schultz, D., Innocentini, G., Wang, J., Walczak, A., Onuchic, J. et Wolynes, P. (2005). Self-regulating gene: an exact solution. *Phys. Rev. E*, 72(5):051907.

Hornung, G. et Barkai, N. (2008). Noise propagation and signaling sensitivity in biological networks: a role for positive feedback. *PLoS Comput. Biol.*, 4(1):e8.

Innocentini, G. et Hornos, J. (2007). Modeling stochastic gene expression under repression. *J. Math. Biol.*, 55(3):413–431.

Istratov, A. A. et Vyvenko, O. F. (1999). Exponential analysis in physical phenomena. *Rev. Sci. Instr.*, 70(2):1233–57.

Jenuwein, T. et Allis, C. (2001). Translating the histone code. *Science*, 293(2):1074–80.

Jiang, C. et Pugh, B. F. (2009). Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.*, 10(3):161–172.

Johnson, T. A., Elbi, C., Parekh, B. S., Hager, G. L. et John, S. (2008). Chromatin remodeling complexes interact dynamically with a glucocorticoid receptor-regulated promoter. *Mol. Biol. Cell*, 19(8):3308–3322.

Kærn, M., Elston, T., Blake, W. et Collins, J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.*, 6(6):451–464.

Kang, Z., Pirskanen, A., Jänne, O. A. et Palvimo, J. J. (2002). Involvement of proteasome in the dynamic assembly of the androgen receptor transcription complex. *J. Biol. Chem.*, 277(50):48366–48371.

Kapral, R. (2008). Multiparticle Collision Dynamics: Simulation of Complex Systems on Mesoscales. *Adv. Chem. Phys.*, 140:89–146.

Karplus, M. et Kuriyan, J. (2005). Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.*, 102(19):6679–6685.

Karpova, T. S., Chen, T. Y., Sprague, B. L. et McNally, J. G. (2004). Dynamic interactions of a transcription factor with DNA are accelerated by a chromatin remodeller. *EMBO Rep.*, 5(11):1064–1070.

Karpova, T. S., Kim, M. J., Spriet, C., Nalley, K., Stasevich, T. J., Kherrouche, Z., Heliot, L. et McNally, J. G. (2008). Concurrent fast and slow cycling of a transcriptional activator at an endogenous promoter. *Science*, 319(5862):466–469.

Kassabov, S. R., Zhang, B., Persinger, J. et Bartholomew, B. (2003). SWI/SNF unwraps, slides, and rewraps the nucleosome. *Mol. Cell*, 11(2):391–403.

Kauffman, S. A. (1993). *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, 1sr édition.

Kaufman, M. (2002). *Principles of Thermodynamics*. Marcel Dekker Inc., New York.

Kaufmann, B. et van Oudenaarden, A. (2007). Stochastic gene expression: from single molecules to the proteome. *Curr. Opin. Genet. Dev.*, 17(2):107–112.

Kelbauskas, L., Woodbury, N. et Lohr, D. (2009). DNA sequence-dependent variation in nucleosome structure, stability, and dynamics detected by a FRET-based analysis. *Biochem. Cell Biol.*, 87(1):323–335.

Kepler, T. B. et Elston, T. C. (2001). Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophys. J.*, 81(6):3116–36.

Kim, H. D. et O'Shea, E. K. (2008). A quantitative model of transcription factor-activated gene expression. *Nat. Struct. Mol. Biol.*, 15(11):1192–8.

Kim, J.-I., Na, S. et Eom, K. (2009). Large Protein Dynamics Described by Hierarchical-Component Mode Synthesis. *J. Chem. Theor. Comput.*, 5(7):1931–9.

Kimura, H. (2005). Histone dynamics in living cells revealed by photobleaching. *DNA Repair (Amst.)*, 4(8):939–950.

Kitano, H. (2002). Systems biology: A brief overview. *Science*, 295(5560):1662–1664.

Ko, M. S. (1991). A stochastic model for gene induction. *J. Theor. Biol.*, 153(2):181–194.

Koga, N. et Takada, S. (2001). Roles of native topology and chain-length scaling in protein folding: a simulation study with a Gō-like model. *J. Mol. Biol.*, 313(1):171–180.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, 128(4):693–705.

Kryzhniy, V. V. (2004). High-resolution exponential analysis via regularized numerical inversion of Laplace transforms. *J. Compu. Phys*, 199(2):618–30.

Kundu, S., Sorensen, D. C. et Phillips, G. N. (2004). Automatic domain decomposition of proteins by a Gaussian Network Model. *Proteins*, 57(4):725–733.

Kupiec, J. J. (1997). A Darwinian theory for the origin of cellular differentiation. *Mol. Gen. Genet.*, 255(2):201–208.

Kurkcuoglu, O., Turgut, O. T., Cansu, S., Jernigan, R. L. et Doruker, P. (2009). Focused functional dynamics of supramolecules by use of a mixed-resolution elastic network model. *Biophys. J.*, 97(4):1178–1187.

Kuttler, C. (2006). Simulating bacterial transcription and translation in a stochastic π calculus. *In Trans. Comput. Syst. Biol. VI*, volume 4220, pages 113–49. Springer.

Lales, C., Parisey, N., Mazat, J.-P. et Beurton-Aimar, M. (2005). Simulation of mitochondrial metabolism using multi-agents system. *In Proc. MAS*BIOMED'05*, page 137.

Lamond, A. I. et Spector, D. L. (2003). Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.*, 4(8):605–612.

Lanctôt, C., Cheutin, T., Cremer, M., Cavalli, G. et Cremer, T. (2007). Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat. Rev. Genet.*, 8(2):104–115.

Langton, C. G. (1986). Studying artificial life with cellular automata. *Physica D*, 22(1-3):120–149.

Langton, C. G. (1991). Life at the edge of chaos. *In Artificial Life II*. Addison-Wesley.

Laradji, M. et Sunil Kumar, P. B. (2005). Domain growth, budding, and fission in phase-separating self-assembled fluid bilayers. *J. Chem. Phys.*, 123(22):224902.

Latham, J. A. et Dent, S. Y. (2007). Cross-regulation of histone modifications. *Nat. Struct. Mol. Biol.*, 14(11):1017–1024.

Lehman, A., Ellwood, K., Middleton, B. et Carey, M. (1998). Compensatory energetic relationships between upstream activators and the RNA polymerase II general transcription machinery. *J. Biol. Chem.*, 273(2):932–939.

Lemaire, V., Lee, C., Lei, J., Métivier, R. et Glass, L. (2006). Sequential recruitment and combinatorial assembling of multiprotein complexes in transcriptional activation. *Phys. Rev. Lett.*, 96(19):198102.

Lemerle, C., Di Ventura, B. et Serrano, L. (2005). Space as the final frontier in stochastic simulations of biological systems. *FEBS Lett.*, 579(8):1789–1794.

Lesne, A. (2006). The chromatin regulatory code: beyond a histone code. *Eur. Phys. J. E Soft Matter*, 19(3):375–377.

Lestas, I., Paulsson, J., Ross, N. et Vinnicombe, G. (2008). Noise in Gene Regulatory Networks. *IEEE Trans.*, 53(1):189–200.

Levsky, J. M., Shenoy, S. M., Pezo, R. C. et Singer, R. H. (2002). Single-cell gene expression profiling. *Science*, 297(5582):836–840.

Levy, Y., Caflisch, A., Onuchic, J. et Wolynes, P. (2004). The folding and dimerization of HIV-1 protease: evidence for a stable monomer from simulations. *J. Mol. Biol.*, 340(1):67–79.

Li, B., Carey, M. et Workman, J. (2007). The role of chromatin during transcription. *Cell*, 128(4):707–719.

Li, G., Levitus, M., Bustamante, C. et Widom, J. (2005). Rapid spontaneous accessibility of nucleosomal DNA. *Nat. Struct. Mol. Biol.*, 12(1):46–53.

Li, G.-W., Berg, O. G. et Elf, J. (2009). Effects of macromolecular crowding and DNA looping on gene regulation kinetics. *Nat. Phys.*, 5(4):294–7.

Lindner, B., García-Ojalvo, J., Neiman, A. et Schimansky-Geier, L. (2004). Effects of noise in excitable systems. *Phys. Rep.*, 392(6):321–424.

Lipniacki, T., Paszek, P., Marciniak-Czochra, A., Brasier, A. et Kimmel, M. (2006). Transcriptional stochasticity in gene expression. *J. Theor. Biol.*, 238(2):348–67.

Lu, P., Vogel, C., Wang, R., Yao, X. et Marcotte, E. M. (2007). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.*, 25(1):117–124.

Lyubartsev, A. P. (2005). Multiscale modeling of lipids and lipid bilayers. *Eur. Biophys. J.*, 35(1):53–61.

Lyubartsev, A. P. et Laaksonen, A. (2004). On the reduction of molecular degrees of freedom in computer simulations. *Lect. Notes Phys.*, 640:219–44.

MacKerell Jr., A. D., Wiórkiewicz-Kuczera, J. et Karplus, M. (1995). An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.*, 117(48):11946–11975.

Maheshri, N. et O'Shea, E. (2007). Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu. Rev. Biophys. Biomol. Struct.*, 36:413–434.

Mar, D. J., Chow, C. C., Gerstner, W., Adams, R. W. et Collins, J. J. (1999). Noise shaping in populations of coupled model neurons. *Proc. Natl. Acad. Sci. U.S.A.*, 96(18):10450–10455.

Martinez Arias, A. et Hayward, P. (2006). Filtering transcriptional noise during development: concepts and mechanisms. *Nat. Rev. Genet.*, 7(1):34–44.

Matera, A. G., Izaguire-Sierra, M., Praveen, K. et Rajendra, T. K. (2009). Nuclear bodies: random aggregates of sticky proteins or crucibles of macromolecular assembly? *Dev. Cell*, 17(5):639–647.

McDonnell, M. D. et Abbott, D. (2009). What is stochastic resonance? Definitions, misconceptions, debates, and its relevance to biology. *PLoS Comput. Biol.*, 5(5):e1000348.

McNally, J., Müller, W., Walker, D., Wolford, R. et Hager, G. (2000). The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science*, 287(5456):1262–1265.

Mellor, J. (2006). Dynamic nucleosomes and gene transcription. *Trends Genet.*, 22(6): 320–329.

Métivier, R., Gallais, R., Tiffoche, C., Le Péron, C., Jurkowska, R., Carmouche, R., Ibberson, D., Barath, P., Demay, F., Reid, G., Benes, V., Jeltsch, A., Gannon, F. et Salbert, G. (2008). Cyclical DNA methylation of a transcriptionally active promoter. *Nature*, 452(7183):45–50.

Métivier, R., Penot, G., Hubner, M., Reid, G., Brand, H., Kos, M. et Gannon, F. (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell*, 115(6):751–763.

Métivier, R., Reid, G. et Gannon, F. (2006). Transcription in four dimensions: nuclear receptor-directed initiation of gene expression. *EMBO Rep.*, 7(2):161–167.

Mettetal, J. et van Oudenaarden, A. (2007). Necessary noise. *Science*, 317(5837): 463–464.

Mettetal, J. T., Muzzey, D., Pedraza, J. M., Ozbudak, E. M. et van Oudenaarden, A. (2006). Predicting stochastic gene expression dynamics in single cells. *Proc. Natl. Acad. Sci. U.S.A.*, 103(19):7304–7309.

Miller, J. et Widom, J. (2003). Collaborative competition mechanism for gene activation in vivo. *Mol. Cell. Biol.*, 23(5):1623–1632.

Misteli, T. (2001). Protein dynamics: implications for nuclear architecture and gene expression. *Science*, 291(5505):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.

Mitarai, N., Dodd, I. B., Crooks, M. T. et Sneppen, K. (2008). The generation of promoter-mediated transcriptional noise in bacteria. *PLoS Comput. Biol.*, 4(7): e1000109.

Moler, C. et Van Loan, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review*, 45(1):3–49.

Mukherjee, A., Bhimalapuram, P. et Bagchi, B. (2005). Orientation-dependent potential of mean force for protein folding. *J. Chem. Phys.*, 123(1):014901–1–11.

Mukherjee, S., Nakanishi, H. et Fuchs, N. (1994). Markov chain analysis of random walks in disordered media. *Phys. Rev. E*, 49(6):5032–5045.

Murphy, K., Balázsi, G. et Collins, J. (2007). Combinatorial promoter design for engineering noisy gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, 104(31):12726–12731.

Nagaich, A., Walker, D., Wolford, R. et Hager, G. (2004). Rapid periodic binding and displacement of the glucocorticoid receptor during chromatin remodeling. *Mol. Cell*, 14(2):163–174.

Neildez-Nguyen, T., Parisot, A., Vignal, C., Rameau, P., Stockholm, D., Picot, J., Allo, V., Le Bec, C., Laplace, C. et Paldi, A. (2007). Epigenetic gene expression noise and phenotypic diversification of clonal cell populations. *Differentiation*, 76(1):33–40.

Nevozhay, D., Adams, R. M., Murphy, K. F., Josic, K. et Balázsi, G. (2009). Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, 106(13):5123–5128.

Newlands, S., Levitt, L. K., Robinson, C. S., Karpf, A. B., Hodgson, V. R., Wade, R. P. et Hardeman, E. C. (1998). Transcription occurs in pulses in muscle fibers. *Genes Dev.*, 12(17):2748–2758.

Newman, J., Ghaemmaghami, S., Ihmels, J., Breslow, D., Noble, M., DeRisi, J. et Weissman, J. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature*, 441(7095):840–846.

Nguyen, H. D. et Hall, C. K. (2004). Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc. Natl. Acad. Sci. U.S.A.*, 101(46):16180–5.

Nguyen, H. D. et Hall, C. K. (2006). Spontaneous fibril formation by polyalanines; discontinuous molecular dynamics simulations. *J. Am. Chem. Soc.*, 128(6):1890–1901.

Nielsen, S. O., Lopez, C. F., Srinivas, G. et Klein, M. L. (2004). Coarse grain models and the computer simulation of soft materials. *J. Phys. Condens. Matter*, 16:R481–R512.

Okabe, Y., Yagi, Y. et Sasai, M. (2007). Effects of the DNA state fluctuation on single-cell dynamics of self-regulating gene. *J. Chem. Phys.*, 127(10):105107.

Orsi, G. A., Couble, P. et Loppin, B. (2009). Epigenetic and replacement roles of histone variant H3.3 in reproduction and development. *Int. J. Dev. Biol.*, 53(2-3):231–43.

Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J. A., Lopes, S., Reik, W. et Fraser, P. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.*, 36(10):1065–1071.

Ozbudak, E., Thattain, M., Kurster, I., Grossman, A. et van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nat. Genet.*, 31(1):69–73.

Pal, S. K., Peon, J. et Zewail, A. H. (2002). Biological water at the protein surface: dynamical solvation probed directly with femtosecond resolution. *Proc. Natl. Acad. Sci. U.S.A.*, 99(4):1763–1768.

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*. McGraw Hill, New York, 3rd édition.

Paszek, P. (2007). Modeling stochasticity in gene regulation: characterization in the terms of the underlying distribution function. *Bull. Math. Biol.*, 69(5):1567–1601.

Paterlini, M. G. et Ferguson, D. M. (1998). Constant temperature simulations using the Langevin equation with velocity Verlet integration. *Chem. Phys.*, 236(1–3):243–52.

Paulsson, J. (2004). Summing up the noise in gene networks. *Nature*, 427(6973):415–8.

Paulsson, J. (2005a). Models of stochastic gene expression. *Phys. Life Rev.*, 2(2):157–75.

Paulsson, J. (2005b). Prime movers of noisy gene expression. *Nat. Genet.*, 37(9):925–6.

Peccoud, J. et Ycart, B. (1995). Markovian Modeling of Gene-Product Synthesis. *Theor. Pop. Biol.*, 48(2):222–234.

Pederson, T. (2001). Protein mobility within the nucleus–what are the right moves? *Cell*, 104(5):635–638.

Pedraza, J. et Paulsson, J. (2008). Effects of molecular memory and bursting on fluctuations in gene expression. *Science*, 319(5861):339–343.

Pedraza, J. et van Oudenaarden, A. (2005). Noise propagation in gene networks. *Science*, 307(5717):1965–1969.

Peitgen, H.-O., Jürgens, H. et Saupe, D. (2004). *Chaos and Fractals – New Frontiers of Science*. Springer, New York, 2nd édition.

Periole, X., Cavalli, M., Marrink, S.-J. et Ceruso, M. A. (2009). Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *J. Chem. Theor. Comput.*, 5(9):2531–43.

Perissi, V. et Rosenfeld, M. G. (2005). Controlling nuclear receptors: the circular logic of cofactor cycles. *Nat. Rev. Mol. Cell Biol.*, 6(7):542–554.

Perlmann, T., Eriksson, P. et Wrange, O. (1990). Quantitative analysis of the glucocorticoid receptor-DNA interaction at the mouse mammary tumor virus glucocorticoid response element. *J. Biol. Chem.*, 265(28):17222–17229.

Phair, R., Scaffidi, P., Elbi, C., Vecerová, J., Dey, A., Ozato, K., Brown, D., Hager, G., Bustin, M. et Misteli, T. (2004). Global nature of dynamic protein-chromatin interactions in vivo: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. *Mol. Cell. Biol.*, 24:6393–6402.

Phair, R. D. et Misteli, T. (2000). High mobility of proteins in the mammalian cell nucleus. *Nature*, 404(14):604–609.

Poirier, M. G., Oh, E., Tims, H. S. et Widom, J. (2009). Dynamics and function of compact nucleosome arrays. *Nat. Struct. Mol. Biol.*, 16(9):938–944.

Potestio, R., Pontiggia, F. et Micheletti, C. (2009). Coarse-grained description of protein internal dynamics: an optimal strategy for decomposing proteins in rigid subunits. *Biophys. J.*, 96(12):4993–5002.

Raj, A., Peskin, C., Tranchina, D., Vargas, D. et Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.*, 4(10):e309.

Raj, A. et van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–26.

Ramanathan, S. et Swain, P. S. (2005). Tracing the sources of cellular variation. *Dev. Cell*, 9(5):576–578.

Ramón y Cajal, S. (1903). Un sencillo método de coloración selectiva del retículo protoplásmico y sus efectos en los diversos órganos nerviosos de vertebrados e invertebrados. *Tra. Lab. Invest. Biol.*, 129-221.

Rao, C., Wolf, D. et Arkin, A. (2002). Control, exploitation and tolerance of intracellular noise. *Nature*, 420(6912):231–237.

Raser, J. et O'Shea, E. (2004). Control of stochasticity in eukaryotic gene expression. *Science*, 304(5678):1811–1814.

Raser, J. et O'Shea, E. (2005). Noise in gene expression: origins, consequences, and control. *Science*, 309(5743):2010–2013.

Raveh-Sadka, T., Levo, M. et Segal, E. (2009). Incorporating nucleosomes into thermodynamic models of transcription regulation. *Genome Res.*, 19(8):1480–1496.

Rayasam, G. V., Elbi, C., Walker, D. A., Wolford, R., Fletcher, T. M., Edwards, D. P. et Hager, G. L. (2005). Ligand-specific dynamics of the progesterone receptor in living cells and during chromatin remodeling in vitro. *Mol. Cell. Biol.*, 25(6):2406–2418.

Reid, G., Gallais, R. et Métivier, R. (2009). Marking time: the dynamic role of chromatin and covalent modification in transcription. *Int. J. Biochem. Cell Biol.*, 41(1): 155–163.

Ridgway, D., Broderick, G., Lopez-Campistrous, A., Ru'aini, M., Winter, P., Hamilton, M., Boulanger, P., Kovalenko, A. et Ellison, M. J. (2008). Coarse-grained molecular simulation of diffusion and reaction kinetics in a crowded virtual cytoplasm. *Biophys. J.*, 94(10):3748–3759.

Rigney, D. R. et Schieve, W. C. (1977). Stochastic model of linear, continuous protein synthesis in bacterial populations. *J. Theor. Biol.*, 69(4):761–766.

Rosenfeld, N., Young, J., Alon, U., Swain, P. et Elowitz, M. (2005). Gene regulation at the single-cell level. *Science*, 307(5717):1962–1965.

Rouquette, J., Genoud, C., Vazquez-Nin, G. H., Kraus, B., Cremer, T. et Fakan, S. (2009). Revealing the high-resolution three-dimensional network of chromatin and interchromatin space: a novel electron-microscopic approach to reconstructing nuclear architecture. *Chromosome Res.*, 17(6):801–810.

Roussel, M. R. et Zhu, R. (2006). Stochastic kinetics description of a simple transcription model. *Bull. Math. Biol.*, 68(7):1681–1713.

Rouvas-Nicolis, C. et Nicolis, G. (2007). Stochastic resonance. *Scholarpedia*, 2(11): 1474.

Saiz, L., Rubi, J. M. et Vilar, J. M. (2005). Inferring the in vivo looping properties of DNA. *Proc. Natl. Acad. Sci. U.S.A.*, 102(49):17642–17645.

Saiz, L. et Vilar, J. (2006a). Stochastic dynamics of macromolecular-assembly networks. *Mol. Syst. Biol.*, 2:2006.0024.

Saiz, L. et Vilar, J. M. (2006b). DNA looping: the consequences and its control. *Curr. Opin. Struct. Biol.*, 16(3):344–350.

Samoilov, M., Price, G. et Arkin, A. (2006). From fluctuations to phenotypes: the physiology of noise. *Sci. STKE*, 2006(366):re17.

Sánchez, A. et Kondev, J. (2008). Transcriptional control of noise in gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, 105(13):5081–5086.

Sauve, A. A., Wolberger, C., Schramm, V. L. et Boeke, J. D. (2006). The biochemistry of sirtuins. *Annu. Rev. Biochem.*, 75:435–465.

Saxton, M. J. (2001). Anomalous subdiffusion in fluorescence photobleaching recovery: a Monte Carlo study. *Biophys. J.*, 81(4):2226–2240.

Schaff, J., Fink, C. C., Slepchenko, B., Carson, J. H. et Loew, L. M. (1997). A general computational framework for modeling cellular structure and function. *Biophys. J.*, 73(3):1135–1146.

Schlick, T. et Perišić, O. (2009). Mesoscale simulations of two nucleosome-repeat length oligonucleosomes. *Phys. Chem. Chem. Phys.*, 11(45):10729–10737.

Schnakenberg, J. (1976). Network theory of microscopic and macroscopic behavior of master equation systems. *Rev. Mod. Phys.*, 48(4):571–585.

Scholl, D. et Nixon, B. T. (1996). Cooperative binding of DctD to the dctA upstream activation sequence of Rhizobium meliloti is enhanced in a constitutively active truncated mutant. *J. Biol. Chem.*, 271(42):26435–26442.

Scholten, E., Linden, E. et This, H. (2008). The life of an anise-flavored alcoholic beverage: does its stability cloud or confirm theory? *Langmuir*, 24(5):1701–1706.

Segal, E. et Widom, J. (2009). From DNA sequence to transcriptional behaviour: a quantitative approach. *Nat. Rev. Genet.*, 10(7):443–456.

Seksek, O., Biwersi, J. et Verkman, A. S. (1997). Translational diffusion of macromolecule-sized solutes in cytoplasm and nucleus. *J. Cell Biol.*, 138(1):131–142.

Setty, Y., Mayo, A., Surette, M. et Alon, U. (2003). Detailed map of a cis-regulatory input function. *Proc. Natl. Acad. Sci. U.S.A.*, 100(13):7702–7707.

Sexton, T., Schober, H., Fraser, P. et Gasser, S. M. (2007). Gene regulation through nuclear organization. *Nat. Struct. Mol. Biol.*, 14(11):1049–1055.

Shahrezaei, V., Ollivier, J. F. et Swain, P. S. (2008). Colored extrinsic fluctuations and stochastic gene expression. *Mol. Syst. Biol.*, 4:196.

Shahrezaei, V. et Swain, P. S. (2008a). Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, 105(45):17256–61.

Shahrezaei, V. et Swain, P. S. (2008b). The stochastic nature of biochemical networks. *Curr. Opin. Biotechnol.*, 19(4):369–374.

Shang, Y., Hu, X., DiRenzo, J., Lazar, M. A. et Brown, M. (2000). Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell*, 103(6):843–52.

Sharma, S., Ding, F. et Dokholyan, N. V. (2007). Multiscale modeling of nucleosome dynamics. *Biophys. J.*, 92(12):1457–1470.

Sharp, Z. D., Mancini, M. G., Hinojos, C. A., Dai, F., Berno, V., Szafran, A. T., Smith, K. P., Lele, T. P., Lele, T. T., Ingber, D. E. et Mancini, M. A. (2006). Estrogen-receptor-alpha exchange and chromatin dynamics are ligand- and domain-dependent. *J. Cell. Sci.*, 119(20):4101–4116.

Shav-Tal, Y., Darzacq, X., Shenoy, S. M., Fusco, D., Janicki, S. M., Spector, D. L. et Singer, R. H. (2004a). Dynamics of single mRNPs in nuclei of living cells. *Science*, 304(5678):1797–1800.

Shav-Tal, Y., Darzacq, X. et Singer, R. H. (2006). Gene expression within a dynamic nuclear landscape. *EMBO J.*, 25(15):3469–3479.

Shav-Tal, Y., Singer, R. H. et Darzacq, X. (2004b). Imaging gene expression in single living cells. *Nat. Rev. Mol. Cell Biol.*, 5(10):855–861.

Shelley, J. C., Shelley, M. Y., Reeder, R. C., Bandyopadhyay, S. et Klein, M. L. (2001a). A coarse grain model for phospholipid simulations. *J. Phys. Chem. B*, 105(16):4464–70.

Shelley, J. C., Shelley, M. Y., Reeder, R. C., Bandyopadhyay, S., Moore, P. B. et Klein, M. L. (2001b). Simulations of phospholipids using a coarse grain model. *J. Phys. Chem. B*, 105(40):9785–9792.

Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Liron, Y., Rosenfeld, N., Danon, T., Perzov, N. et Alon, U. (2006). Variability and memory of protein levels in human cells. *Nature*, 444(7119):643–646.

Simpson, M., Cox, C. et Sayler, G. (2003). Frequency domain analysis of noise in autoregulated gene circuits. *Proc. Natl. Acad. Sci. U.S.A.*, 100(8):4551–4556.

Simpson, M., Cox, C. et Sayler, G. (2004). Frequency domain chemical Langevin analysis of stochasticity in gene transcriptional regulation. *J. Theor. Biol.*, 229(3):383–94.

Snijder, B., Sacher, R., Rämö, P., Damm, E. M., Liberali, P. et Pelkmans, L. (2009). Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature*, 461(7263):520–523.

Soula, H., Robardet, C., Perrin, F., Gripon, S., Beslon, G. et Gandrillon, O. (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.

Spector, D. L. (2004). Stopping for FISH and chips along the chromatin fiber superhighway. *Mol. Cell*, 15(6):844–846.

Spector, D. L. (2006). SnapShot: Cellular bodies. *Cell*, 127(6):1071.

Sprague, B. L. et McNally, J. G. (2005). FRAP analysis of binding: proper and fitting. *Trends Cell Biol.*, 15(2):84–91.

Sprague, B. L., Pego, R. L., Stavreva, D. A. et McNally, J. G. (2004). Analysis of binding reactions by fluorescence recovery after photobleaching. *Biophys. J.*, 86(6): 3473–3495.

Stavreva, D. A., Müller, W. G., Hager, G. L., Smith, C. L. et McNally, J. G. (2004). Rapid glucocorticoid receptor exchange at a promoter is coupled to transcription and regulated by chaperones and proteasomes. *Mol. Cell. Biol.*, 24(7):2682–2697.

Steinbach, P. J. (2005). Introduction to Macromolecular Simulation. Biophysical Society's Biophysics Textbook Online (BTOL).

Stenoien, D. L., Patel, K., Mancini, M. G., Dutertre, M., Smith, C. L., O'Malley, B. W. et Mancini, M. A. (2001). FRAP reveals that mobility of oestrogen receptor-alpha is ligand- and proteasome-dependent. *Nat. Cell Biol.*, 3(1):15–23.

Süel, G., Kulkarni, R., Dworkin, J., Garcia-Ojalvo, J. et Elowitz, M. (2007). Tunability and noise dependence in differentiation dynamics. *Science*, 315(5819):1716–9.

Summers, D. et Scott, J. M. (1988). Systems of first-order chemical reactions. *Math. Comput. Modelling*, 10(12):901–9.

Sun, L., Yang, G., Zaidi, M. et Iqbal, J. (2008). TNF-induced gene expression oscillates in time. *Biochem. Biophys. Res. Commun.*, 371(4):900–905.

Sutherland, H. et Bickmore, W. A. (2009). Transcription factories: gene expression in unions? *Nat. Rev. Genet.*, 10(7):457–466.

Swain, P. (2004). Efficient attenuation of stochasticity in gene expression through post-transcriptional control. *J. Mol. Biol.*, 344(20):965–976.

Swain, P. S., Elowitz, M. B. et Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, 99(20):12795–800.

Swedlow, J. R. et Lamond, A. I. (2001). Nuclear dynamics: where genes are and how they got there. *Genome Biol.*, 2:REVIEWS0002.

Takagi, F., Koga, N. et Takada, S. (2003). How protein thermodynamics and folding mechanisms are altered by the chaperoning cage: molecular simulations. *Proc. Natl. Acad. Sci. U.S.A.*, 100(20):11367–11372.

Takahashi, K., Arjunan, S. N. V. et Tomita, M. (2005). Space in systems biology of signaling pathways – towards intracellular molecular crowding in silico. *FEBS Lett.*, 579(8):1783–1788.

Talbert, P. B. et Henikoff, S. (2006). Spreading of silent chromatin: inaction at a distance. *Nat. Rev. Genet.*, 7(10):793–803.

Tanaka, M. (1996). Modulation of promoter occupancy by cooperative DNA binding and activation-domain function is a major determinant of transcriptional regulation by activators in vivo. *Proc. Natl. Acad. Sci. U.S.A.*, 93(9):4311–4315.

Tao, Y., Zheng, X. et Sun, Y. (2007). Effect of feedback regulation on stochastic gene expression. *J. Theor. Biol.*, 247(4):827–836.

Thattai, M. et van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A.*, 98(15):8614–9.

Thomas, M. C. et Chiang, C. M. (2006). The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.*, 41(3):105–178.

Tipler, P. A. et Mosca, G. (1997). *Physics for scientists and engineers*. W. H. Freeman, 5th édition.

Tkačik, G., Gregor, T. et Bialek, W. (2008). The role of input noise in transcriptional regulation. *PLoS ONE*, 3(7):e2774.

Tkačik, G. et Bialek, W. (2009). Diffusion, dimensionality, and noise in transcriptional regulation. *Phys. Rev. E*, 79(5):051901.

Tozzini, V. (2005). Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.*, 15(2): 144–150.

Tozzini, V. (2010). Multiscale modeling of proteins. *Acc. Chem. Res.*, 43(2):220–230.

Truong, K. et Ikura, M. (2001). The use of FRET imaging microscopy to detect protein-protein interactions and protein conformational changes in vivo. *Curr. Opin. Struct. Biol.*, 11(5):573–578.

Tsui, V. et Case, D. A. (2000). Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers*, 56(4):275–291.

Tănase-Nicola, S., Warren, P. et ten Wolde, P. (2006). Signal detection, modularity, and the correlation between extrinsic and intrinsic noise in biochemical networks. *Phys. Rev. Lett.*, 97(6):068102.

Turing, A. M. (1952). The chemical basis of morphogenesis. *Phil. Trans. Royal Soc. Lond. B*, 237(641):37–72.

Turner, T. E., Schnell, S. et Burrage, K. (2004). Stochastic approaches for modelling in vivo reactions. *Comput. Biol. Chem.*, 28(3):165–78.

Tüzel, E., Pan, G., Ihle, T. et Kroll, D. M. (2007). Mesoscopic model for the fluctuating hydrodynamics of binary and ternary mixtures. *Europhys. Lett.*, 80(4): 40010.

Väisänen, S., Dunlop, T. W., Sinkkonen, L., Frank, C. et Carlberg, C. (2005). Spatio-temporal activation of chromatin on the human CYP24 gene promoter in the presence of 1alpha,25-Dihydroxyvitamin D3. *J. Mol. Biol.*, 350(1):65–77.

van Kampen, N. (2007). *Stochastic processes in physics and chemistry.* North-Holland, New York, 3rd édition.

van Werven, F. J., van Teeffelen, H. A., Holstege, F. C. et Timmers, H. T. (2009). Distinct promoter dynamics of the basal transcription factor TBP across the yeast genome. *Nat. Struct. Mol. Biol.*, 16(10):1043–1048.

van Zon, J., Morelli, M., Tanase-Nicola, S. et ten Wolde, P. (2006). Diffusion of transcription factors can drastically enhance the noise in gene expression. *Biophys. J.*, 91(12):4350–4367.

van Zon, J. et ten Wolde, P. (2005a). Green's-function reaction dynamics: a particle-based approach for simulating biochemical networks in time and space. *J. Chem. Phys.*, 123(23):234910.

van Zon, J. et ten Wolde, P. (2005b). Simulating biochemical networks at the particle level and in time and space: Green's function reaction dynamics. *Phys. Rev. Lett.*, 94(12):128103.

Vattulainen, I. et Karttunen, M. (2005). Modeling of Biologically Motivated Soft Matter Systems. *In* Rieth, M. et Schommers, W., éditeurs : *Handbook of Theoretical and Computational Nanotechnology*, chapitre 29. American Scientific Publishers.

Veening, J., Smits, W. et Kuipers, O. (2008). Bistability, Epigenetics, and Bet-Hedging in Bacteria. *Annu. Rev. Microbiol.*, 62:193–210.

Venturoli, M., Sperotto, M. M., Kranenburg, M. et Smit, B. (2006). Mesoscopic models of biological membranes. *Phys. Rep.*, 437(1–2):1–54.

Verdone, L., Caserta, M. et Di Mauro, E. (2005). Role of histone acetylation in the control of gene expression. *Biochem. Cell Biol.*, 83(3):344–353.

Vilar, J. et Saiz, L. (2005). DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr. Opin. Genet. Dev.*, 15(5):136–144.

Volfson, D., Marciniak, J., Blake, W., Ostroff, N., Tsimring, L. et Hasty, J. (2006). Origins of extrinsic variability in eukaryotic gene expression. *Nature*, 439(7078): 861–864.

Voliotis, M., Cohen, N., Molina-ParÃs, C. et Liverpool, T. B. (2008). Fluctuations, pauses, and backtracking in DNA transcription. *Biophys. J.*, 94(2):334–348.

Voss, T. C., Schiltz, R. L., Sung, M. H., Johnson, T. A., John, S. et Hager, G. L. (2009). Combinatorial probabilistic chromatin interactions produce transcriptional heterogeneity. *J. Cell. Sci.*, 122(3):345–356.

Wagoner, J. et Baker, N. A. (2004). Solvation forces on biomolecular structures: a comparison of explicit solvent and Poisson-Boltzmann models. *J. Comput. Chem.*, 25(13):1623–1629.

Walczak, A. M., Mugler, A. et Wiggins, C. H. (2009). A stochastic spectral analysis of transcriptional regulatory cascades. *Proc. Natl. Acad. Sci. U.S.A.*, 106(16):6529–34.

Wang, Y., Guo, L., Golding, I., Cox, E. C. et Ong, N. P. (2009). Quantitative transcription factor binding kinetics at the single-molecule level. *Biophys. J.*, 96(2):609–20.

Wang, Y., Liu, C. L., Storey, J. D., Tibshirani, R. J., Herschlag, D. et Brown, P. O. (2002). Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. U.S.A.*, 99(9):5860–5865.

Wang, Z., Zang, C., Rosenfeld, J., Schones, D., Barski, A., Cuddapah, S., Cui, K., Roh, T., Peng, W., Zhang, M. et Zhao, K. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.*, 40(7):897–903.

Warmflash, A. et Dinner, A. (2008). Signatures of combinatorial regulation in intrinsic biological noise. *Proc. Natl. Acad. Sci. U.S.A.*, 105(45):17262–17267.

Warren, P., Tanase-Nicola, S. et ten Wolde, P. (2006). Exact results for noise power spectra in linear biochemical reaction networks. *J. Chem. Phys.*, 125(14):144904.

WATERBORG, J. (2001). Dynamics of histone acetylation in Saccharomyces cerevisiae. *Biochemistry*, 40(8):2599–2605.

WATERBORG, J. H. (2002). Dynamics of histone acetylation in vivo. A function for acetylation turnover? *Biochem. Cell Biol.*, 80(3):363–378.

WEINBERGER, L. S., BURNETT, J. C., TOETTCHER, J. E., ARKIN, A. P. et SCHAFFER, D. V. (2005). Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell*, 122(2):169–182.

WIESMEIJER, K., KROUWELS, I. M., TANKE, H. J. et DIRKS, R. W. (2008). Chromatin movement visualized with photoactivable GFP-labeled histone H4. *Differentiation*, 76(1):83–90.

WILKINSON, D. J. (2009). Stochastic modelling for quantitative description of heterogeneous biological systems. *Nat. Rev. Genet.*, 10(2):122–133.

WOLFRAM, S. (2002). *A new kind of science*. Wolfram Media inc.

WRAY, G. A., HAHN, M. W., ABOUHEIF, E., BALHOFF, J. P., PIZER, M., ROCKMAN, M. V., ROMANO, L. A. et WRAY, G. A. (2003). The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.*, 20(9):1377–1419.

WUNDERLICH, Z. et MIRNY, L. A. (2008). Spatial effects on the speed and reliability of protein-DNA search. *Nucleic Acids Res.*, 36(11):3570–3578.

XU, N., CHEN, C. Y. et SHYU, A. B. (1997). Modulation of the fate of cytoplasmic mRNA by AU-rich elements: key sequence features controlling mRNA deadenylation and decay. *Mol. Cell. Biol.*, 17(8):4611–4621.

YU, J., REN, X., LAO, K. et XIE, X. (2006). Probing gene expression in live cells, one protein molecule at a time. *Science*, 311(5767):1600–3.

ZACCAI, G. (2000). How soft is a protein? A protein dynamics force constant measured by neutron scattering. *Science*, 288(5471):1604–1607.

ZENKLUSEN, D., LARSON, D. R. et SINGER, R. H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.*, 15(12):1263–1271.

ZOFALL, M., PERSINGER, J., KASSABOV, S. R. et BARTHOLOMEW, B. (2006). Chromatin remodeling by ISW2 and SWI/SNF requires DNA translocation inside the nucleosome. *Nat. Struct. Mol. Biol.*, 13(4):339–346.

Annexes

A Décomposition propre de $\tilde{\mathbf{M}}$ et réduction de l'expression du FRAP

Les valeurs propres et vecteurs propres à droite et à gauche des matrices $\mathbf{M}$ et $\check{\mathbf{M}}$ (cf équation (III.37)) sont notés respectivement $(\boldsymbol{\lambda}, \boldsymbol{\Lambda}, \boldsymbol{\Lambda}^{-1})$ et $(\check{\boldsymbol{\lambda}}, \check{\boldsymbol{\Lambda}}, \check{\boldsymbol{\Lambda}}^{-1})$. Et $\Lambda_i^{\overline{\mathcal{U}}}$ et $\Lambda_i^{\mathcal{U}}$ (resp. $\Lambda_i^{-1\overline{\mathcal{U}}}$ et $\Lambda_i^{-1\mathcal{U}}$) désignent les éléments de Λ_i (resp. Λ_i^{-1}) correspondant aux états respectivement hors de $\mathcal{U}$ et dans $\mathcal{U}$.

À partir de la structure en blocs de la matrice $\tilde{\mathbf{M}}$ (équation (III.38)), on remarque que les vecteurs $[\Lambda_i^{\overline{\mathcal{U}}}, \Lambda_i^{\mathcal{U}}, \mathbf{0}]$ et $[\Lambda_i^{-1\overline{\mathcal{U}}} | \Lambda_i^{-1\mathcal{U}} | \Lambda_i^{-1\mathcal{U}}]$ (les blocs font la même taille que ceux de $\tilde{\mathbf{M}}$) sont des vecteurs propres à droites et à gauche de $\tilde{\mathbf{M}}$, associés aux valeurs propres λ_i (les valeurs propres de $\mathbf{M}$). De même, on remarque que les vecteurs $[\mathbf{0}, -\check{\Lambda}_i, \check{\Lambda}_i]$ et $[\mathbf{0}|\mathbf{0}|\check{\Lambda}_i^{-1}]$ sont des vecteurs propres à droites et à gauche de $\tilde{\mathbf{M}}$, associés aux valeurs propres $\check{\lambda}_i$ (les valeurs propres de $\check{\mathbf{M}}$).

Dans l'expression du FRAP (III.39), l'exponentielle matricielle $e^{\tau\tilde{\mathbf{M}}}$ est projetée à droite par $\tilde{\Lambda}_0^\circ = [\Lambda_0^{\overline{\mathcal{U}}}, \mathbf{0}, \Lambda_0^{\mathcal{U}}]$ et à gauche par $\boldsymbol{\delta}_{f\bullet}{}^{\mathrm{T}} = [\mathbf{0}|\mathbf{1}|\mathbf{0}]$. Sa décomposition en modes propres est

$$\mathrm{FRAP}_f(\tau) = \sum_i e^{\tau\check{\lambda}_i} \left[\boldsymbol{\delta}_{f\bullet}{}^{\mathrm{T}} \tilde{\Lambda}_i \tilde{\Lambda}_i^{-1} \tilde{\Lambda}_0^\circ\right] \tag{A.1}$$

Pour toutes les valeurs propres de $\tilde{\mathbf{M}}$ correspondant à celles de $\mathbf{M}$, la projection $\tilde{\Lambda}_i^{-1}\tilde{\Lambda}_0^\circ$ se réécrit $[\Lambda_i^{-1\overline{\mathcal{U}}} | \Lambda_i^{-1\mathcal{U}} | \Lambda_i^{-1\mathcal{U}}]\ [\Lambda_0^{\overline{\mathcal{U}}}, \mathbf{0}, \Lambda_0^{\mathcal{U}}] = [\Lambda_i^{-1\overline{\mathcal{U}}} | \Lambda_i^{-1\mathcal{U}}]\ [\Lambda_0^{\overline{\mathcal{U}}}, \Lambda_0^{\mathcal{U}}] = \Lambda_i^{-1}\Lambda_0$ et vaut donc 1 pour $i = 0$ et 0 sinon. Le cas $i = 0$ est donc le seul mode propre provenant de $\mathbf{M}$ qui soit non nul et vaut la constante $\mathbf{1}\Lambda_0^{\mathcal{U}}$.

De la même façon, pour toutes les valeurs propres de $\tilde{\mathbf{M}}$ correspondant à celles de $\check{\mathbf{M}}$, la projection $\tilde{\Lambda}_i^{-1}\tilde{\Lambda}_0^\circ$ se réécrit $[\mathbf{0}|\mathbf{0}|\check{\Lambda}_i^{-1}]\ [\Lambda_0^{\overline{\mathcal{U}}}, \mathbf{0}, \Lambda_0^{\mathcal{U}}] = \check{\Lambda}_i^{-1}\Lambda_0^{\mathcal{U}}$ et la projection $\boldsymbol{\delta}_{f\bullet}{}^{\mathrm{T}}\tilde{\Lambda}_i$ se réécrit $[\mathbf{0}|\mathbf{1}|\mathbf{0}]\ [\mathbf{0}, -\check{\Lambda}_i, \check{\Lambda}_i] = -\mathbf{1}\check{\Lambda}_i$.

On obtient ainsi

$$\mathrm{FRAP}_f(\tau) = \mathbf{1}\Lambda_0^{\mathcal{U}} - \sum_i e^{\tau\check{\lambda}_i} \left[\mathbf{1}\ \check{\Lambda}_i \check{\Lambda}_i^{-1} \Lambda_0^{\mathcal{U}}\right] \tag{A.2}$$

$$= \mathbf{1}\Lambda_0^{\mathcal{U}} - \mathbf{1}e^{\tau\check{\mathbf{M}}}\Lambda_0^{\mathcal{U}} \tag{A.3}$$

B Paramétrage (chapitre III)

B.1 Paramètres des différents systèmes

Promoteur	Figures	Défini par	Cadre simplifié (sec. 2.5)	Optimisation (annexe B.2)			Énergie-dép.
				Matrices	Structure prédéfinie	α	
Pr0	III.2	Tab. B.2	Oui	———	Non	———	Non
PrEuc1	III.3c, III.4, III.6, III.7, III.9, III.10, III.11, III.12a-c, III.16, III.17	Fig. III.4a1-3	Non	**G** et **E**	Oui	0	Oui
PrMin	III.18, III.19a1-4	Tab. B.3	Oui	———	Non	———	Non
PrProc	III.20	Tab. B.4	Oui	———	Non	———	Non
PrEuc2	III.21, III.22	Tab. B.5	Oui	**M**	Non	0.1	Oui
PrEuc3a	III.12d1-e1, III.29b-f	Fig. III.29b	Non	**G** et **E**	Oui	0	Oui
PrEuc3b	III.12d2-e2, III.30a1-f1	Fig. III.30b1	Non	**E**	Oui	0	Oui
PrEuc3c	III.12d3-4,e3-4, III.30a2-f2	Fig. III.30b2	Non	**G** et **E**	Oui	0	Oui

Tab. B.1: Récapitulatif des différents promoteurs présentés dans le chapitre III et de leur définition.

Paramètre	N	K_A^{d}	K_B^{d}	$[A]$	$[B]$	k_A^{off}	k_B^{off}
Valeur	3	10	10	3	3	0.1	0.2
Unités		nM	nM	nM	nM	s^{-1}	s^{-1}
Paramètre	k_C^{on}	k_C^{off}	ΔG_{AB}	ΔG_{AC}	ΔG_{BC}	γ	$\tilde{\gamma}$
Valeur	1/30	1/30	-2	2	2	1/15	1/30
Unités	min^{-1}	min^{-1}	kcal/mol	kcal/mol	kcal/mol	min^{-1}	min^{-1}
Paramètre	$\tilde{\rho}$		ρ				
Valeur	$200\,\tilde{\gamma}$		[2.5, 3, 3, 4, .1, .1, .1, .1]				
Unités	min^{-1}		min^{-1}				

Tab. B.2: Paramètres pour le système représenté sur la figure III.2. Il s'agit d'un exemple-jouet très simple servant uniquement à des fins d'illustration.

Paramètre	N	K_A^{d}	K_B^{d}	$[A]$	$[B]$	$1/k_A^{\mathrm{off}}$	$1/k_B^{\mathrm{off}}$
Valeur	2	0.5	5	[10^{-2},10^3]	5	30	60
Unités		nM	nM	nM	nM	s	s

Paramètre	ΔE	γ	$\tilde{\gamma}$	$\tilde{\rho}$	ρ
Valeur	2.5	1/5	1/20	$100\,\tilde{\gamma}$	[0.04, 0.04, 0.4, 0.4]
Unités	kcal/mol	min^{-1}	min^{-1}	min^{-1}	s^{-1}

$$\mathbf{k}^0 = \begin{array}{c} \\ A \\ B \end{array} \begin{array}{c} \begin{array}{cccc} \varnothing & A & B & AB \end{array} \\ \begin{bmatrix} a & \bar{a} & a & \bar{a} \\ b & bc & \bar{b} & \bar{b}c \end{bmatrix} \end{array} \qquad \text{avec } a = \frac{k_A^{\mathrm{off}}}{K_A^{\mathrm{d}}} \;,\; \bar{a} = k_A^{\mathrm{off}} \;,\; b = \frac{k_B^{\mathrm{off}}}{K_B^{\mathrm{d}}} \;,\; \bar{b} = k_B^{\mathrm{off}} \text{ et } c = e^{-\Delta E/RT}$$

Tab. B.3: Paramètres pour le système représenté sur la figure III.18, montrant une augmentation de la variabilité de l'expression lors de l'augmentation de la concentration du FT A. Ce système étant défini dans le cadre simplifié de la section 2.5, les valeurs non-nulles de la matrice $\mathbf{M}^0$ (de dimension $N \times N$) sont définies à partir de la matrice $\mathbf{k}^0$ (de dimensions $n \times N$) par $M^0_{s \ominus f,s} = k^0_{f,s}, \forall f \in \mathcal{F}, \forall s \in 2^{\mathcal{F}}$.

Paramètre	N	K_A^{d}	K_C^{d}	$[A]$	$[C]$	$1/k_A^{\mathrm{off}}$	$1/k_C^{\mathrm{off}}$
Valeur	4	20	1	[0.01,100]	[0.01,1000]	20	60
Unités		nM	nM	nM	nM	s	s

Paramètre	k^{close}	ΔG_{loop}	ΔG_{A-A}	ΔG_{A-C}	$\Delta G_{\mathrm{loop}-A}$	$\Delta E_{\mathrm{loop}-C}$	γ
Valeur	1	9	-2	1.5	-5.5	2.5	1/5
Unités	s^{-1}	kcal/mol	kcal/mol	kcal/mol	kcal/mol	kcal/mol	min^{-1}

Paramètre	ρ
Valeur	[5, 5, 10, 10, 14, 14, 14, 14, 1, 1, 3, 3, 9, 9, 9, 9]
Unités	min^{-1}

$$\mathbf{k}^0 = \begin{array}{c} \\ A_1 \\ A_2 \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccccccccccccccc} \phi & A_1 & A_2 & A_1A_2 & C & A_1C & A_2C & A_1A_2C & D & A_1D & A_2D & A_1A_2D & CD & A_1CD & A_2CD & A_1A_2CD \end{array} \\ \begin{bmatrix} a & \bar{a} & a & \bar{a}g & a & \bar{a} & a & \bar{a}g & a & \bar{a} & a & \bar{a}g & a & \bar{a} & a & \bar{a}g \\ a & a & \bar{a} & \bar{a}g & a & a & \bar{a}h & \bar{a}gh & a & a & \bar{a} & \bar{a}g & a & a & \bar{a}h & \bar{a}gh \\ c & c & c & c & \bar{c} & \bar{c} & \bar{c}h & \bar{c}h & cp & cp & cp & cp & \bar{c}p & \bar{c}p & \bar{c}hp & \bar{c}hp \\ d & d & d & d & d & d & d & d & \bar{d} & \bar{d}l & \bar{d}l & \bar{d}l^2 & \bar{d} & \bar{d}l & \bar{d}l & \bar{d}l^2 \end{bmatrix} \end{array}$$

$$\text{avec } a = \frac{k_A^{\mathrm{off}}}{K_A^{\mathrm{d}}} \;,\; \bar{a} = k_A^{\mathrm{off}} \;,\; c = \frac{k_C^{\mathrm{off}}}{K_C^{\mathrm{d}}} \;,\; \bar{c} = k_C^{\mathrm{off}} \;,\; d = k^{\mathrm{close}} \;,\; \bar{d} = k^{\mathrm{close}} e^{\Delta G_{\mathrm{loop}}/RT} \;,$$
$$g = e^{\Delta G_{A-A}/RT} \;,\; h = e^{\Delta G_{A-C}/RT} \;,\; l = e^{\Delta G_{\mathrm{loop}-A}/RT} \text{ et } p = e^{-\Delta E_{\mathrm{loop}-C}/RT}$$

Tab. B.4: Paramètres pour le système représenté sur la figure III.20. Il s'agit d'un promoteur présentant des caractéristiques typiques des promoteurs procaryotes et illustrant la plasticité des propriétés stochastiques de l'expression génique que permet la dynamique du promoteur. Ce système étant défini dans le cadre simplifié de la section 2.5, les valeurs non-nulles de la matrice $\mathbf{M}^0$ (de dimension $N \times N$) sont définies à partir de la matrice $\mathbf{k}^0$ (de dimensions $n \times N$) par $M^0_{s \ominus f,s} = k^0_{f,s}, \forall f \in \mathcal{F}, \forall s \in 2^{\mathcal{F}}$.

Paramètre	N	$[A]$,$[B]$,$[C]$,$[D]$,$[E]$,$[F]$ et $[G]$	t_{slow}	t_{fast}	α
Valeur	7	3	40	40	0.1
Unités		nM	min	s	

Paramètre	γ	$\tilde{\gamma}$	$\tilde{\rho}$	ρ
Valeur	1/10	1/25	$1000\,\tilde{\gamma}$	$6\,\delta_A\,\delta_B\,(1+\delta_C)$
Unités	min^{-1}	min^{-1}	min^{-1}	min^{-1}

Tab. B.5: Paramètres pour le système représenté sur les figures III.21 et III.22, correspondant à un promoteur eucaryote très simplifié et illustrant la complexité potentielle du contrôle de la dynamique du promoteur par la concentration des FT. Ce système s'inscrit dans le cadre simplifié présenté à la section 2.5, chapitre III, et le graphe d'états du promoteur n'est pas structuré *a priori* par des contraintes telles que celles de la figure III.4, page 114. Toutes les combinaisons de facteurs sont donc possibles. Les valeurs de la matrice $\mathbf{M}$ sont obtenues par l'algorithme d'optimisation présenté en annexe B.2 en utilisant les paramètres présentés dans ce tableau. δ_f représente le vecteur à N éléments dont le s-ième élément est 1 si $f \in s$ et 0 sinon.

B.2 Optimisation paramétrique pour une activité périodique

Dans certains cas, les éléments de la matrice $\mathbf{M}$ sont optimisés à l'aide d'un algorithme de descente de gradient par Monte Carlo afin d'obtenir un comportement périodique significativement cohérent (voir par exemple la figure III.22).

Après une initiation aléatoire, les étapes suivantes sont répétées itérativement : la matrice $\mathbf{M}$ est randomisée en ajoutant des variables aléatoires indépendantes (lois normales centrées d'écart type 0.3 kcal/mol) soit (i) aux éléments de $\mathbf{G}$ et de $\mathbf{E}$ (en maintenant inchangées les différences $E_{s',s} - E_{s,s'}$), soit (ii) aux éléments de $\mathbf{E}$ uniquement, soit (iii) directement aux valeurs de $\mathbf{M}$ (en modifiant chaque énergie de réaction $E_{s',s} - G_s$ indépendamment). Les deux premiers cas ne modifient pas la consommation énergétique des réactions, contrairement au troisième. La nouvelle matrice $\mathbf{M}$ ainsi obtenue est acceptée et utilisée pour l'itération suivante si et seulement si elle améliore le critère d'optimisation suivant :

$$\frac{\big\langle \kappa(\lambda_i) \big\rangle \quad \kappa(\lambda_1)}{\left\langle \left[\frac{t_{\text{fast}}}{t_{\text{slow}}} \frac{\text{Re}(-\lambda_i)}{|\text{Im}(\lambda_1)|/2\pi} - 1 \right]^2 \right\rangle^{\alpha}}$$

où $\lambda_0, \lambda_1, \ldots, \lambda_{2^N-1}$ sont les valeurs propres de $\mathbf{M}$ triées par ordre croissant de leur module, où les moyennes $\langle\cdot\rangle$ sont calculées sur l'ensemble des valeurs propres $i \in [0; N-1]$ et où $\kappa(\cdot)$ représente le facteur de cohérence (équation (III.17)).

Le premier terme au numérateur favorise la cohérence des valeurs propres, le second assure que le mode le plus lent est également le plus cohérent. Le critère au dénominateur (dont l'importance est déterminée par α et qui n'est effectif que si $\alpha \neq 0$) conduit les vitesses de relaxation des différents modes à être en moyenne $t_{\text{fast}}/t_{\text{slow}}$ fois plus rapides que la fréquence d'oscillation du mode le plus lent. Dans l'unique cas où ce critère est utilisé (figure III.21 et table B.5), on constate que les modes apériodiques sont répartis sur des temps de relaxation très variable dont les plus rapides vont jusqu'à 6 s. Contrairement au cas de la figure III.4 où le graphe de transitions est globalement pré-structuré en cycle, cette propriété n'apparaît pas spontanément si le critère n'est pas utilisé ($\alpha = 0$).

FOLIO ADMINISTRATIF

THÈSE SOUTENUE DEVANT L'INSTITUT NATIONAL DES SCIENCES APPLIQUÉES DE LYON

NOM : COULON DATE de SOUTENANCE : 1er juillet 2010

Prénoms : Antoine

TITRE : Stochasticité de l'expression génique et régulation transcriptionnelle – Modélisation de la dynamique spatiale et temporelle des structures multiprotéiques

NATURE : Doctorat Numéro d'ordre : 2010-ISAL-0045

École doctorale : Informatique et Mathématiques (InfoMaths)

Spécialité : Informatique

Cote B.I.U. - Lyon : T 50/210/19 / et bis CLASSE :

RÉSUMÉ :
La nature stochastique de l'expression génique est maintenant clairement établie expérimentalement et apparaît comme une composante à part entière de la dynamique cellulaire. Une source importante de cette variabilité est liée au caractère dynamique des diverses structures multiprotéiques impliquées dans le processus d'expression génique. Nous étudions ici, par la modélisation, comment les interactions entre des molécules au comportement individuel probabiliste sont susceptibles de faire naître des dynamiques globales pouvant influencer l'expression génique. Nous nous concentrons plus particulièrement sur deux aspects du processus d'expression : d'une part, son caractère spatialisé au sein d'un noyau cellulaire structuré et dynamique et, d'autre part, la combinatoire des événements moléculaires stochastiques au niveau du promoteur d'un gène.

Pour l'étude des phénomènes d'organisation mésoscopique au sein du noyau cellulaire, nous proposons un modèle de simulation "4D" (intégrant l'espace et le temps). Il emprunte différentes techniques aux formalismes des échelles inférieures (moléculaires) et supérieures (cellulaires), en gardant les aspects essentiels à notre étude (individualité de certaines molécules, exclusion stérique, interactions électromagnétiques, réactions chimiques ...).

Afin d'étudier spécifiquement la dynamique stochastique de la régulation transcriptionnelle, nous proposons un second modèle décrivant les événements d'association/dissociation et de modification de la chromatine en se basant sur l'affinité coopérative/compétitive des molécules et leur potentielle activité enzymatique ou de remodelage. Par des techniques analytiques et computationnelles, nous caractérisons alors l'activité du promoteur à l'aide d'outils de théorie du signal, mais aussi en reproduisant les mesures obtenues par diverses techniques expérimentales (cinétique de ChIP, FRAP, FRET, cytométrie de flux ...). L'analyse de ce modèle démontre que l'activité spontanée du promoteur peut être complexe et structurée, présentant en particulier des dynamiques multi-échelles similaires à celles observées expérimentalement (turnover rapide des molécules, comportements cycliques lents, hétérogénéités transcriptionnelles ...). Nous montrons enfin comment la confrontation de mesures expérimentales de diverses natures peut renseigner sur la structure du système sous-jacent. Ce modèle apparaît alors comme un cadre théorique général pour l'étude de la dynamique des promoteurs et pour l'interprétation intégrée de données expérimentales.

MOTS-CLÉS : Biologie computationnelle, biophysique, bioinformatique, bruit, fluctuations, corps nucléaires, promoteur, chromatine, code des histones, épigénétique, chaîne de Markov, équation maîtresse, valeurs propres.

Laboratoires de recherche : Laboratoire d'InfoRmatique en Image et Systèmes d'information
UMR CNRS 5205 – INSA de Lyon – Bâtiment Blaise Pascal
20, avenue Albert Einstein, 69621 Villeurbanne Cedex, France

Centre de Génétique Moléculaire et Cellulaire
UMR CNRS 5534 – Université Lyon 1 – Bâtiment Gregor Mendel
16 rue Raphaël Dubois, 69622 Villeurbanne Cedex, France

Directeurs de thèse :	Guillaume Beslon,	Olivier Gandrillon
Composition du jury :	Alain Arnéodo	Directeur de recherche, CNRS, ENS Lyon (président du jury)
	Hugues Berry	Chargé de recherche HDR, INRIA (examinateur)
	Guillaume Beslon	Professeur, INSA de Lyon (directeur de thèse)
	Xavier Darzacq	Chargé de recherche HDR, CNRS, ENS (rapporteur)
	Hidde de Jong	Directeur de recherche, INRIA (rapporteur)
	Olivier Gandrillon	Directeur de recherche, CNRS, Univ. de Lyon (directeur de thèse)
	Raphaël Métivier	Chargé de recherche, CNRS, Université Rennes 1 (examinateur)

Stochasticité de l'expression génique et régulation transcriptionnelle
Modélisation de la dynamique spatiale et temporelle des structures multiprotéiques

Résumé : La nature stochastique de l'expression génique est maintenant clairement établie expérimentalement et apparaît comme une composante à part entière de la dynamique cellulaire. Une source importante de cette variabilité est liée au caractère dynamique des diverses structures multiprotéiques impliquées dans le processus d'expression génique. Nous étudions ici, par la modélisation, comment les interactions entre des molécules au comportement individuel probabiliste sont susceptibles de faire naître des dynamiques globales pouvant influencer l'expression génique. Nous nous concentrons plus particulièrement sur deux aspects du processus d'expression : d'une part, son caractère spatialisé au sein d'un noyau cellulaire structuré et dynamique et, d'autre part, la combinatoire des événements moléculaires stochastiques au niveau du promoteur d'un gène.

Pour l'étude des phénomènes d'organisation mésoscopique au sein du noyau cellulaire, nous proposons un modèle de simulation "4D" (intégrant l'espace et le temps). Il emprunte différentes techniques aux formalismes des échelles inférieures (moléculaires) et supérieures (cellulaires), en gardant les aspects essentiels à notre étude (individualité de certaines molécules, exclusion stérique, interactions électromagnétiques, réactions chimiques ...).

Afin d'étudier spécifiquement la dynamique stochastique de la régulation transcriptionnelle, nous proposons un second modèle décrivant les événements d'association/dissociation et de modification de la chromatine en se basant sur l'affinité coopérative/compétitive des molécules et leur potentielle activité enzymatique ou de remodelage. Par des techniques analytiques et computationnelles, nous caractérisons alors l'activité du promoteur à l'aide d'outils de théorie du signal, mais aussi en reproduisant les mesures obtenues par diverses techniques expérimentales (cinétique de ChIP, FRAP, FRET, cytométrie de flux ...). L'analyse de ce modèle démontre que l'activité spontanée du promoteur peut être complexe et structurée, présentant en particulier des dynamiques multi-échelles similaires à celles observées expérimentalement (turnover rapide des molécules, comportements cycliques lents, hétérogénéités transcriptionnelles ...). Nous montrons enfin comment la confrontation de mesures expérimentales de diverses natures peut renseigner sur la structure du système sous-jacent. Ce modèle apparaît alors comme un cadre théorique général pour l'étude de la dynamique des promoteurs et pour l'interprétation intégrée de données expérimentales.

Mots-clés : Biologie computationnelle, biophysique, bioinformatique, bruit, fluctuations, corps nucléaires, promoteur, chromatine, code des histones, épigénétique, chaîne de Markov, équation maîtresse, valeurs propres.

Stochasticity of gene expression and transcriptional regulation
Modeling the spatial and temporal dynamics of multiprotein structures

Abstract: The stochastic nature of gene expression is now clearly established experimentally and appears as an essential part of the cellular dynamics. An important source of this variability comes from the dynamics of the different multiprotein structures involved in the process of gene expression. Using a modeling approach, we study here how the interactions between molecules which individual behavior is probabilistic can result in global dynamics and can influence the expression level. We focus more precisely on two aspects of the gene expression process: first, its spatial context within a structured and dynamic nuclear landscape, and, second, the combinatorial stochastic molecular events occurring on the promoter of a gene.

For the study of mesoscopic organization phenomena in the cell nucleus, we propose a "4D" simulation model (considering space and time). It is built upon different techniques taken from higher and lower scale formalisms (ie. molecular simulations and cellular models), while keeping the essential aspects for our purpose (individuality of some molecules, steric volume exclusion, electromagnetic interactions, chemical reactions, ...).

For investigating the stochastic dynamics of transcriptional regulation, we propose a second model describing the events of association/dissociation and chromatin modification, based on the cooperative/competitive affinity of molecules and their potential enzymatic or remodeling activity. Using analytical and computational techniques, we describe the activity of the promoter with tools from signal theory, but also by reproducing the measures of various experimental techniques (ChIP kinetics, FRAP, FRET, flow cytometry, ...). The analysis of the model reveals that the spontaneous activity of a promoter can be highly complex and demonstrate a multi-scale dynamics similar to what is observed experimentally (rapid turnover of molecules, slow cyclical behavior, transcriptional heterogeneity, ...). Finally, we show that confronting experimental data of various nature can reveal the structure of the underlying system. This model appears as a general theoretical framework for the investigation of promoter dynamics and the interpretation of experimental data.

Keywords: Computational biology, biophysics, bioinformatics, noise, fluctuations, nuclear bodies, promoter, chromatin, histone code, epigenetics, stochastic process, Markov chain, master equation, eigenvalues.