



HAL
open science

Structure et dynamique du cytoplasme auto-organisé : exemple par la ségrégation du génome bactérien

Gabriel David

► **To cite this version:**

Gabriel David. Structure et dynamique du cytoplasme auto-organisé : exemple par la ségrégation du génome bactérien. Physique [physics]. Université Montpellier, 2019. Français. NNT : 2019MONT098 . tel-02470985v2

HAL Id: tel-02470985

<https://theses.hal.science/tel-02470985v2>

Submitted on 2 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Physique

École doctorale Information Structures Systèmes

Unité de recherche Laboratoire Charles Coulomb

Structure et dynamique du cytoplasme auto-organisé : exemple par la ségrégation du génome bactérien

Présentée par Gabriel DAVID

Le 26 Novembre 2019

Sous la direction de John PALMERI
et Frédéric GENIET

Devant le jury composé de

Andrea PARMEGGIANI, Professeur au L2C

Jean-François JOANNY, Professeur au Collège de France

Jean-Marc VICTOR, Directeur de recherche CNRS au LPTMC

Anne-Florence BITBOL, Chargée de recherche CNRS au LJP

Frédéric GENIET, Maître de conférence au L2C

John PALMERI, Directeur de recherche CNRS au L2C

Président du jury

Rapporteur

Rapporteur

Examineur

Co-encadrant

Directeur de thèse



UNIVERSITÉ
DE MONTPELLIER

Résumé

Le vivant apparaît organisé. Les bactéries utilisent des compartiments sans membrane pour confiner les réactions chimiques dans l'espace et le temps. Il existe un paradigme général de l'auto-organisation de l'espace intracellulaire qui distingue les auto-assemblages, structures moléculaires assemblées par des mécanismes passifs de transition de phase, et les structures dissipatives, engendrés par exemple par des processus de réaction-diffusion. Si les auto-assemblages correspondent à l'évolution vers l'équilibre thermodynamique, les structures dissipatives sont des manifestations d'une dépense d'énergie hors équilibre thermodynamique. Nous illustrons ce paradigme en étudiant la ségrégation du matériel génétique chez les bactéries, en l'occurrence celle du plasmide F de *Escherichia coli* qui repose sur le système de partition ParABS. La ségrégation est une étape cruciale du cycle cellulaire bactérien puisqu'elle assure la transmission de l'information génétique dans les bactéries filles avant la division.

Le système ParABS est constitué d'une séquence de type centromérique *parS*; d'une protéine ParB capable de se lier à l'ADN, spécifiquement sur la séquence *parS* et non-spécifiquement ailleurs ; d'une protéine ATPase ParA capable de se lier à l'ADN. Les interactions entre protéines ParB sur l'ADN et l'adsorption spécifique sur la séquence *parS* mènent à la formation d'un foyer tridimensionnel appelé complexe ParBS localisé autour de la séquence *parS*. Les interactions entre protéines ParA et ParB mènent au positionnement de ce complexe au centre de la cellule. Après réplication, deux complexes ParBS existent et sont ségrégés par l'action des protéines ParA aux positions 1/4 et 3/4 de l'espace intracellulaire.

Nous cherchons d'abord à expliquer la formation des complexes ParBS par un mécanisme passif de séparation de phase entre états de haute et basse densité en protéines ParB dans l'espace. Nous construisons deux modèles de physique statistique à l'aide d'outils empruntés à la physique des transitions de phase. Notre première approche permet une étude préliminaire simple mais riche en résultats tandis que notre deuxième approche définit rigoureusement tous les éléments du système biologique constitué de l'ADN-polymère et des protéines ParB en interaction et nous permet de formuler un véritable critère d'existence des transitions de phase du premier ordre qui est vérifié par l'ADN.

Nous parvenons à tracer les diagrammes de phase de cette transition. Ces deux modèles nous permettent d'argumenter que le régime thermodynamique physiologique de ce système biologique est un régime de coexistence métastable en protéines ParB sur l'ADN. La séquence *parS* joue le rôle d'un défaut ou d'une graine de nucléation. Nous utilisons une troisième approche pour expliciter le lien entre les distributions tridimensionnelle et sur l'ADN des protéines ParB autour de la séquence *parS*.

Nous cherchons à expliquer les courbes de recouvrement de fluorescence issus d'expériences de photoblanchiment sur les complexes ParBS. Nous construisons une méthode de photoblanchiment *in silico*, c'est-à-dire que nous reproduisons ces courbes de recouvrement à partir d'une équation phénoménologique résolue numériquement. Nous développons un système d'équations qui décrivent l'évolution des protéines sur l'ADN à partir de l'approche physique statistique précédente pour produire un photoblanchiment *in silico* prenant en compte que les complexes ParBS sont issus d'une séparation de phase. Nous montrons qu'un pur système passif ne permet pas de faire des expériences de photoblanchiment à cause de la maturation d'Ostwald subie par les complexes. Nous corrigeons cette approche pour inclure les protéines ParA et leur cycle biochimique dans nos simulations. Nous montrons ainsi que les interactions entre les protéines ParA et ParB et l'hydrolyse de l'ATP permet la survie de plusieurs complexes ParBS grâce à un mécanisme d'inversion du murissement d'Ostwald. Cette approche fondamentale permet d'expliquer le positionnement des complexes ParBS durant la ségrégation.

Abstract

Living systems appear organized. Bacteria use membraneless compartments to confine chemical reactions in space and time. There is a general paradigm of intracellular space self-organization that distinguishes between self-assembly, molecular structures assembled by passive phase transition mechanisms, and dissipative structures, generated for example by reaction-diffusion processes. If self-assemblies correspond to the evolution towards thermodynamic equilibrium, dissipative structures are manifestations of an out of equilibrium energy cost. We illustrate this paradigm by studying the segregation of bacterial genome, in this case the F-plasmid segregation of *Escherichia coli*, based on the ParABS partition system. Segregation is a crucial step in the bacterial cell cycle since it ensures the transmission of genetic information in daughter bacteria before division.

The ParABS system consists of a *parS* centromeric sequence; a ParB protein which is able to bind to DNA, specifically on the *parS* sequence and not specifically elsewhere; and a ParA ATPase protein that can bind to DNA. Interactions between ParB proteins on DNA and specific adsorption on the *parS* sequence lead to the formation of a three-dimensional focus called the ParBS complex located around the *parS* sequence. Interactions between ParA and ParB proteins lead to the positioning of this complex at the center of the cell cytoplasm. After replication, two ParBS complexes exist and are segregated by the action of ParA proteins at positions 1/4 and 3/4 of the intracellular space.

We first seek to explain the formation of ParBS complexes by a passive phase separation mechanism between high- and low-density states of ParB proteins in space. We construct two statistical physics models using tools borrowed from the physics of phase transitions. Our first approach offers a first, simple but rich in results study while our second approach rigorously defines all the elements of the biological system consisting of the interacting DNA-polymer and ParB proteins and allows us to formulate a first-order phase transition existence criterion that is verified by the DNA. We can draw the phase diagrams of this transition. These two models allow us to argue that the physiological thermodynamic regime of this biological system is a regime of metastable coexistence in ParB proteins on DNA. The *parS* sequence plays

the role of a defect or nucleation seed. We use a third approach to explain the relationship between the three-dimensional and DNA distributions of ParB proteins around the *parS* sequence.

We try to explain the fluorescence recovery curves from photobleaching experiments on ParBS complexes. We construct an *in silico* photobleaching method, i.e. we reproduce these recovery curves from a phenomenological equation solved numerically. We then develop a system of equations that describe the evolution of proteins on DNA from the previous statistical physical approach to produce an *in silico* photobleaching taking into account that ParBS complexes are the result of phase separation. We show that a pure passive system does not allow photobleaching experiments because of the Ostwald maturation undergone by the complexes. We correct this approach by including ParA proteins and their biochemical cycle in our simulations. We show that the interactions between ParA and ParB proteins and the hydrolysis of ATP allows the survival of several ParBS complexes thanks to an inversion mechanism of Ostwald's ripening. This fundamental approach explains the positioning of ParBS complexes during segregation.

À Stéphane, mon frère

Remerciements

Je tiens à remercier en premier lieu Anne-Florence Bitbol du LPJ (Paris) et les rapporteurs Jean-Marc Victor du LPTMC (Paris) et Jean-François Joanny du Collège de France (Paris) pour avoir accepté de participer au jury de soutenance de thèse, pour leurs remarques bienveillantes et leurs conseils pertinents pour la conclusion de ce projet de recherche.

Je remercie les membres d'équipes extérieures au Laboratoire Charles Coulomb, sans lesquelles ce projet de thèse n'aurait pas débuté, c'est-à-dire les équipes de Marcelo Nollmann, avec Antoine Le Gall, Diego Cattoni et Baptiste Guilhas, du Centre de Biochimie Structurale de Montpellier et de Jean-Yves Bouet du Laboratoire de Microbiologie et Génétique Moléculaires de Toulouse : merci pour leur patience vis à vis des questions, souvent répétitives, du physicien. Je remercie également nos collaborateurs à l'étranger, Ned Wingreen, qui m'a permis de visiter son laboratoire à l'Université de Princeton et de discuter avec ses étudiants, Chenyi Fei, Benjamin Weiner, et collaborateurs Pierre Ronceray, Marianne Bauer et Amir Erez, ainsi que Chase Broedersz de l'Université Louis-et-Maximilien.

Je souhaite ensuite remercier chaleureusement mes maîtres du L2C ! Merci d'abord à John Palmeri, qui m'a donné avec confiance (ou folie) ce projet de thèse et qui, par sa direction, m'a transmis presque toutes les compétences que je possède aujourd'hui. Merci à Andrea Parmeggiani pour sa passion (ou propagande) et ses encouragements réguliers sans lesquels un jeune étudiant n'aurait jamais pris successivement les voies de la physique du vivant et de la recherche il y a quelques années. Merci à Frédéric Geniet pour son génie physique et synthétique et pour m'avoir montré qu'il y a une place pour chacun dans l'étude de la nature. Merci à Jérôme Dorniac pour son génie calculatoire et pour m'avoir montré qu'un calcul analytique n'est jamais assez moche pour ne rien en tirer. Merci à Jean-Charles Walter pour son expertise dans les systèmes biologiques, son humour et sa patience (ou accoutumance) avec Dominique que je remercie également. Merci à Nils-Ole Walliser pour sa rigueur et son dévouement, ainsi que pour les différents échanges que nous avons pu avoir. Merci à eux deux de m'avoir soutenu durant la mission aux États-Unis. Merci à Vladimir Lorman, dont je garde le souvenir d'une expertise physique rigoureuse et d'une grande bienveillance à l'égard des jeunes étudiants. Merci

donc à eux, pour leur patience et leur enseignement ! Je tiens à remercier Gilbert Moultaqa et Miguel Manna pour les nombreuses discussions de couloir très enrichissantes, et Michele Frigerio pour les nombreux conseils qu'il m'a prodigués. Merci également à Carole Chevalier, ainsi qu'aux autres doctorants du laboratoire Valerio Sorichetti, Chahine Abbas, Rupert Coy, Rodrigo Calderon, Anthony Gourdin pour les discussions.

Merci à ma mère, merci à mon père, ainsi qu'à mes frères, à ma soeur et à Sophanith. Merci à eux de m'avoir encouragé à poursuivre mes études et à m'engager dans un milieu qui m'était étranger. Je tiens à remercier Thierry, Yassine, Lucas et Paul, mes amis qui ont été avec moi aux moments les plus durs. Merci à Rick, Bob et Terry pour les propos parfois intéressants et pertinents qu'ils peuvent tenir.

« Il semble qu'il y ait deux "mécanismes" différents qui permettent la production d'événements ordonnés, le "mécanisme statistique" qui produit de "l'ordre à partir du désordre", et la méthode nouvelle, créatrice "d'ordre à partir d'ordre". Pour l'esprit sans préjugés, le second principe paraît beaucoup plus simple, beaucoup plus plausible. Il l'est certainement. C'est pourquoi les physiciens étaient si fiers d'avoir découvert l'autre, le principe de "l'ordre à partir du désordre", qui est réellement suivi par la nature et qui seul permet la compréhension de la suite des événements naturels, et en premier lieu de leur irréversibilité. [...] Nous ne devons pas, par conséquent, être découragés par la difficulté d'interpréter la vie par les lois ordinaires de la physique. Car c'est précisément là ce qui est à prévoir, compte tenu des connaissances que nous avons acquises au sujet de la structure de la matière vivante. Nous devons donc nous attendre à trouver un nouveau type de loi physique se manifestant en elle. Devrons-nous l'appeler une loi non-physique, pour ne pas dire super-physique ? Non, je ne le pense pas. »

Erwin Schrödinger,
Qu'est-ce que la vie ? De la physique à la biologie,
Cambridge University Press (1944).

Table des matières

1	Introduction	14
1.1	Paradigme de l'auto-organisation	14
1.1.1	Auto-assemblage	15
1.1.2	Auto-organisation	16
1.2	Ségrégation de l'ADN et système ParABS	17
1.3	État de l'art	19
1.3.1	Distribution des protéines ParB...	20
1.3.2	Echange entre deux complexes ParBS	24
1.3.3	Positionnement des complexes ParBS	25
1.4	Plan de la thèse	26
2	Statique thermodynamique	29
2.1	Introduction	29
2.2	Rappels de physique statistique	30
2.2.1	Distribution canonique	30
2.2.2	Distribution grand canonique	31
2.3	Gaz sur réseau à courte portée	32
2.3.1	Gaz sur réseau à D dimensions	32
2.3.2	Gaz sur réseau et modèle d'Ising	34
2.3.3	Solution exacte pour $D = 1$	35
2.3.4	Champ moyen et approche variationnelle	36
2.3.5	Illustration des régimes thermodynamiques	43
2.3.6	Scénario métastable	46
2.3.7	Limite et succès du champ moyen	49
2.4	Polymère à 3D interagissant avec un bain de protéines	55
2.4.1	<i>Spreading</i> et <i>bridging</i>	55
2.4.2	Simplification du <i>spreading</i> et <i>bridging</i>	57
2.4.3	Critère d'existence pour les séparations de phase sur l'ADN	60
2.4.4	Diagrammes de phase	62
2.5	Lien entre séparation de phase 1D et 3D	73
2.5.1	<i>Looping</i> et <i>clustering</i>	74
2.5.2	Résultats	79

2.6	Conclusion	82
2.7	Annexe 1 : estimation de c_0 pour un polymère auto-évitant . . .	84
2.8	Annexe 2 : exposants critiques du modèle IRLG	86
3	Cinétique et dynamique	90
3.1	Introduction	90
3.2	Approche phénoménologique	91
3.2.1	Modèle de comptage des protéines	92
3.2.2	Extraire k_{in} et k_{out} des données de FRAP	93
3.2.3	Temps de résidence typiques à l'intérieur et à l'extérieur des complexes	94
3.3	Approche intermédiaire	96
3.3.1	Modèle de diffusion avec un potentiel spatial	96
3.3.2	Photoblanchiment <i>in silico</i>	97
3.4	Approche fondamentale	105
3.4.1	Gaz sur réseau continu à une et deux espèces	105
3.4.2	Photoblanchiment <i>in silico</i> sans protéine ParA	117
3.4.3	Equilibre des complexes ParBS	126
3.4.4	Photoblanchiment <i>in silico</i> avec protéine ParA	133
3.5	Lien entre cinétique et dynamique	136
3.5.1	Force de protéophorèse	137
3.5.2	Force dérivée de l'énergie libre	140
3.6	Conclusion	145
3.7	Annexe 1 : Logiciel COMSOL Multiphysics	147
4	Conclusion	150

Chapitre 1

Introduction

Le présent travail s'inscrit dans la problématique de l'auto-organisation du vivant, c'est-à-dire de l'organisation active de la matière vivante, sujet à la frontière de la biologie, de la physique statistique hors et à l'équilibre thermodynamique et de la physique de la matière molle. Les systèmes biologiques qui entrent de cette définition possèdent plusieurs composants chimiques distincts, dotés de propriétés propres, et qui interviennent dans des mécanismes complémentaires. Dans ce manuscrit, nous illustrons un paradigme simple de l'organisation active du vivant à l'aide d'un système biologique dont les éléments sont bien caractérisés.

Dans ce premier chapitre introductif, nous présentons d'abord les concepts utilisés pour décrire l'organisation de la matière vivante. Ces concepts généraux nous permettent d'introduire le système biologique qui a motivé notre étude. Nous dressons ensuite un état de l'art des travaux déjà effectués pour caractériser et modéliser ce système particulier. Nous détaillons pour finir les contributions apportées par ce manuscrit à sa compréhension.

1.1 Paradigme de l'auto-organisation

Le cytoplasme cellulaire est un environnement organisé [1]. Chez les cellules eucaryotes, le milieu intracellulaire est divisé en organites dotés de membranes phospholipidiques qui les séparent du reste du cytoplasme [2]. Ces compartiments sont spécialisés, c'est-à-dire qu'ils possèdent des fonctions biologiques particulières. De manière non exhaustive, ces organites rassemblent le noyau cellulaire qui isole le matériel génétique confiné en son sein et qui abrite les processus de la transcription et l'organisation de l'ADN ; l'appareil de Golgi qui gère le transport vésiculaire dans le cytoplasme ; les mitochondries qui synthétisent les molécules d'ATP. Cette organisation permet à la cellule eucaryote de localiser et de confiner les réactions chimiques dans l'espace intracellulaire. L'organisation intracellulaire peut ainsi être définie comme la gestion spatio-

temporelle des réactions chimiques à l'intérieur du cytoplasme. Les bactéries ne peuvent pas partager le même paradigme d'organisation : les organelles membranaires, tel que le noyau, n'existe pas dans ce domaine du vivant. Par exemple, le matériel génétique fluctue directement dans l'espace intracellulaire où il est confiné par des protéines liées à l'ADN.

La formation d'une organelle membranaire ne constitue pas la seule manière d'organiser le cytoplasme [3, 4]. Les bactéries, et les cellules eucaryotes, montrent des compartiments dits sans membranes c'est-à-dire des zones de confinement spatial d'entités chimiques comme les protéines [5]. Les cellules eucaryotes possèdent par exemple des centrosomes, centre de nucléation des microtubules et origine du transport intracellulaire, constitués d'une paire de centrioles escortés d'un nuage de matériel péricentriolaire. La classification des mécanismes de formation de ces zones de confinement ordonnées souffre de certaines ambiguïtés. Des définitions rigoureuses sont données par la thermodynamique. On distingue ainsi deux catégories de mécanismes en suivant la classification de [6] :

1. l'auto-assemblage d'entités chimiques comme les protéines, qui correspond à une maturation du système biologique vers son état d'équilibre thermodynamique. Une structure auto-assemblée est ainsi reliée à la minimisation d'une énergie libre ;
2. et l'auto-organisation, qui nécessite la consommation d'une énergie. Ces systèmes, dissipatifs, existent hors de l'équilibre thermodynamique.

Remarquablement, ces catégories correspondent aux deux méthodes de production d'ordre, respectivement l'ordre à partir de l'ordre et l'ordre à partir du désordre, définies par Erwin Schrödinger [7]. Une autre classification, qui nous semble plus pertinente dans ses termes, distingue dans l'auto-organisation du vivant les auto-assemblages, décrits par les transitions de phase, et les structures dissipatives, décrites par des instabilités dynamiques et des bifurcations [8].

1.1.1 Auto-assemblage

Il existe de nombreux exemples d'auto-assemblage par différents types de transitions de phase [9]. Leur description est largement empruntée à la physique des colloïdes [10]. L'auto-assemblage nécessite des entités chimiques qui possèdent des interactions attractives entre elles [11, 12, 13]. Par exemple, les tubulines des microtubules interagissent, engendrant ainsi un auto-assemblage dirigé dans l'espace [14]. Des travaux récents ont mis en exergue les séparations de phase liquide-liquide pour expliquer la formation des organelles sans membrane [15]. Certaines structures possèdent en effet des propriétés typiques des liquides : elles fusionnent, se déforment et se recomposent comme des liquides, prennent des formes sphériques indiquant la possible présence d'une

tension de surface, et le temps typique de réarrangement de leurs constituants est de l'ordre de quelques secondes. Les cas semblent être nombreux même s'ils sont rarement parfaitement caractérisés selon les critères mentionnés précédemment [16, 17]. L'exemple souvent retenu est la formation des granules P dans les cellules embryonnaires du ver *Caenorhabditis elegans* [18, 19]. Ces granules sont constitués d'ARN et de protéines qui se lient à l'ARN et vérifient toutes les propriétés mentionnées précédemment. Notons que ces séparations de phase nécessitent des structures polymériques pour s'initier, comme l'ARN dans le cas des granules P, des complexes de ribonucléoprotéines chez la bactérie *Caulobacter crescentus* [20, 21], ou l'ADN comme dans les cas des systèmes ParABS, traité dans ce manuscrit, ou PomXYZ [22, 23]. Aujourd'hui, le concept de séparation de phase liquide-liquide connaît une utilisation probablement abusive (voir par exemple [24]). Ces séparations de phase liquide-liquide, ou des gélations de protéines, autre type de transition reconstruite dans les systèmes biologiques [25], font intervenir des protéines intrinsèquement non structurées, c'est-à-dire des protéines qui possèdent un ou plusieurs motifs dont la structure tridimensionnelle est instable, ces motifs portant souvent les séquences d'interaction avec d'autres protéines [26]. La formation d'agrégats, solides, de protéines existent également [27]. Ils sont communs dans les causes de maladie neurodégénérative [28]. En particulier dans le cas de la maladie d'Alzheimer, deux transitions de phase distinctes pourraient avoir lieu : une séparation de phase liquide-liquide suivie d'une transition liquide-solide [29], propriété bien connue des systèmes de particules colloïdales [30]. Ces transitions de phase ont été modélisées très largement par des approches de type Flory [19, 31, 32].

1.1.2 Auto-organisation

Les composants d'un système d'auto-organisation cohérent et minimal de protéines, diffusent avec des temps typiques différenciés et peuvent réagir chimiquement pour initier un changement de conformation via une réaction d'hydrolyse [33]. L'auto-organisation concerne donc des systèmes actifs. Ces différents temps typiques entraînent la formation de motif. Les modèles de réaction-diffusion apparaissent comme le paradigme des systèmes auto-organisés [34, 35]. Un exemple fondamental de localisation d'une activité chimique spécifique concerne le système de positionnement de la nouvelle paroi cellulaire entre deux bactéries filles de l'espèce *Escherichia coli* à l'issue de la division cellulaire. Ce positionnement repose sur les trois protéines qui composent le système MinCDE [36]. Les composants de ce système montrent des oscillations pôle à pôle sur des échelles de temps distinctes, conséquence d'une dynamique de positionnement bistable [37]. Ces oscillations entraînent le positionnement d'un anneau de filaments de protéines FtsZ au milieu de la cellule. La protéine FtsZ recrute ensuite les protéines nécessaires à la formation d'une paroi et à

la scissiparité. Des expériences *in vivo* ont même pu mettre en évidence des motifs dynamiques spirales des protéines MinE et MinD [38], typiques des systèmes de réaction-diffusion. Le positionnement du chromosome principal de *E. coli* avec le système MukBEF a été modélisé par des équations de réaction-diffusion [39, 40]. Le positionnement des plasmides bactériens par le système ParABS est étudié dans ce manuscrit [41].

Ces classifications soulignent que l'auto-organisation de l'espace intracellulaire est un processus hors équilibre, cet équilibre existant préalablement et pouvant être décrit par un potentiel thermodynamique, grâce à la thermalisation des processus moléculaires. Dans ce manuscrit, nous illustrons ce paradigme général dans le cadre de la ségrégation du matériel génétique chez les bactéries.

1.2 Ségrégation de l'ADN et système ParABS

Les systèmes biologiques, et en l'occurrence le système qui nous intéresse, peuvent utiliser simultanément l'auto-organisation et l'auto-assemblage [8]. Notre cas d'étude concerne la ségrégation des plasmides chez les bactéries, et en particulier la ségrégation du plasmide F chez *E. coli*. L'ADN bactérien existe sous deux formes : le chromosome principal qui mesure plusieurs millions de paires de base, et les plasmides, bien plus courts, de l'ordre de quelques dizaines de milliers de paires de base. Ce matériel génétique est dit circulaire car les chromosomes et les plasmides bactériens forment des chaînes d'ADN fermées. Comme le chromosome principal, le plasmide peut porter des gènes essentiels au développement d'une bactérie mais il comporte généralement des gènes plutôt optionnel permettant à l'organisme de survivre dans des conditions particulières [42]. Par exemple, les plasmides jouent un rôle capital dans la transmission de la résistance aux antibiotiques [43]. La ségrégation est la migration du matériel génétique répliqué dans les deux sous-espaces du cytoplasme qui correspondent aux deux futures bactéries filles après la division cellulaire. Il s'agit donc d'une étape cruciale du cycle de vie bactérien puisqu'elle assure la transmission héréditaire du matériel génétique. Le chromosome principal et les plasmides possèdent des mécanismes distincts de ségrégation [44, 45]. Les plasmides à faible nombre de copies, dont fait partie le plasmide F, sont ségrégés par le système de partition dit de type I génériquement appelé système ParABS [46]. Ce système est largement utilisé puisque près de 70% des espèces de bactéries utilisent ce système ou un système analogue pour ségréger certains plasmides et leur chromosome principal, et sa séquence centromérique se retrouve même chez certaines archées [47]. Le système ParABS du plasmide F est constitué de trois éléments [48] :

1. Le premier élément est une séquence d'ADN de type centromérique

appelée *parS*. Cette séquence de 43 paires de base (pb) est répétée 12 fois autour de l'origine de réplication du plasmide F. Les deux autres éléments du système de partition sont des protéines codées par des gènes situés sur l'opéron *par*, à la suite de la séquence *parS*.

2. Le second élément est une protéine appelée ParB, exprimée à partir du gène *parB*. Les détails de sa structure sont importants. Les protéines ParB du plasmide F sont constituées de 323 acides aminés. A partir du domaine N terminal, les résidus 1 à 42 possèdent une séquence d'interaction avec la protéine motrice du système de partition, les résidus 174 à 201 un motif hélice-coude-hélice d'adsorption spécifique sur la séquence centromérique, les résidus 245 à 323 un domaine de dimérisation. Les protéines ParB dimérisent puis s'adsorbent de manière spécifique sur une séquence palindromique contenue dans la séquence *parS*. Les dimères de ParB peuvent aussi se fixer n'importe où sur l'ADN de manière non spécifique [49]. Un dimère de ParB possède une empreinte sur l'ADN de 16pb. Une protéine ParB possède un domaine d'oligomérisation du côté de la terminaison N, c'est-à-dire un domaine d'interaction avec une autre protéine ParB (résidus 36 à 72). Ce domaine d'oligomérisation se situe sur un motif replié mais flexible et instable concernant les résidus 1 à 156 : la protéine ParB entre ainsi dans la définition d'une protéine intrinsèquement non structurée. Ce motif favorise des interactions entre protéines ParB dans l'espace appelée interactions de *bridging*. Sur leur partie centrale (résidus 155 à 270), les protéines ParB montrent un second domaine de dimérisation responsable d'un second type d'interaction, le *spreading*, entre deux dimères de ParB adsorbés de manière adjacente sur l'ADN. Pour alléger les formulations, nous utiliserons le terme de protéine ParB pour désigner les dimères de protéines parB.
3. Le dernier élément est la protéine motrice du système de partition : il s'agit de la protéine ATPase ParA codée par le gène *parA*. Il existe deux formes de la protéine ParA : une forme ATP et une forme hydrolysée ADP. La forme ParA-ATP se lie partout sur l'ADN de manière non spécifique sous forme de dimères [50]. La protéine ParA peut se lier à une séquence qui précède l'opéron *par* pour réguler l'expression des gènes *parA* et *parB* [51]. La forme ATP liée à l'ADN interagit avec les protéines ParB. L'hydrolyse de l'ATP par la protéine ParA provoque un changement de sa conformation et entraîne sa libération de l'ADN dans le cytoplasme. L'activité d'hydrolyse par ParA est relativement faible mais elle est stimulée d'un facteur 40 lorsque la protéine ParA-ATP interagit avec une protéine ParB, et d'un facteur 120 en présence d'une séquence *parS* [52]. Une fois libérée, la protéine ParA est sous forme ADP et diffuse librement dans le cytoplasme bactérien. Plus précisément, le cycle biochimique de cette protéine ATPase possède deux branches [53]. La branche

mineure consiste en un détachement de l'ADN sans hydrolyse de l'ATP et la branche majeure en un détachement de l'ADN avec hydrolyse. Nous assimilerons dans ce manuscrit la forme ATP à la forme liée à l'ADN, et la forme ADP à la forme qui diffuse librement dans le cytoplasme. En conséquence de cette séparation entre forme liée et forme libre, les protéines ParA montrent deux coefficients de diffusion distincts [54] : une diffusion rapide pour la forme libre et une diffusion plus lente pour la forme liée à l'ADN. La formation d'oligomères ou de *patch* de protéines ParA montrent des interactions entre dimères de ParA-ATP [54]. Les protéines ParA du plasmide F montrent des oscillations pôle à pôle au sein du nucléoïde. La fonction biologique de ces oscillations n'a pas encore été déterminée.

La ségrégation du plasmide F peut être résumée succinctement. Les interactions entre la séquence *parS* et les protéines ParB entraînent la formation d'un auto-assemblage de protéines ParB de forme sphérique appelé complexe de partition ParBS [22]. Des expériences ont montré qu'une seule séquence *parS* est nécessaire à la formation des complexes ParBS [55] : nous parlerons par la suite de la séquence *parS* par simplicité. Il est intéressant de noter que l'hydrolyse de l'ATP n'est pas nécessaire à la formation des complexes [56]. Avant la répllication, il n'y a qu'un seul plasmide F et un seul complexe ParBS positionné au centre du nucléoïde, volume occupé par l'ADN au sein du cytoplasme. Historiquement, l'auto-assemblage des complexes ParBS a été décrit par le scénario biologique de *spreading* [57]. Ce scénario stipule qu'une protéine ParB se lie à la séquence *parS* et que le complexe se structure d'abord linéairement sur la molécule d'ADN grâce aux interactions entre protéines voisines sur la chaîne. Des travaux ont néanmoins montré que les interactions de *bridging* sont nécessaires à la formation de cet auto-assemblage [58]. Des simulations numériques ont montré la nécessité des interactions de *spreading*, néanmoins leur importance *in vivo* pourrait être remise en question [55, 59]. Après la répllication, les deux copies des plasmides se séparent, formant deux complexes ParBS et migrent vers les positions 1/4 et 3/4 du nucléoïde [54], qui correspondent aux futures centres des nouvelles bactéries. La présence de la protéine ParA et de l'hydrolyse de l'ATP sont nécessaires à la ségrégation et au positionnement du ou des complexes ParBS. Sans hydrolyse, il n'y a qu'un seul complexe ParBS qui sort du nucléoïde et ne se positionne pas [54].

1.3 État de l'art

Du point de vue biophysique, le système ParABS forme un système modèle riche. Son faible nombre d'éléments, leurs interactions et le nombre conséquent de données expérimentales disponibles encouragent le physicien sur la voie d'une modélisation physique. Dans ce manuscrit, nous nous concentrons sur

quatre résultats expérimentaux distincts : la distribution des protéines ParB dans l'espace intracellulaire et sur l'ADN, la récupération de la fluorescence des complexes ParBS et la distribution des complexes ParBS et des protéines ParA lors de la ségrégation. Ces travaux ont été réalisés en collaboration avec l'équipe de Marcelo Nollmann du Centre de Biochimie Structurale de Montpellier et l'équipe de Jean-Yves Bouet du Laboratoire de Microbiologie et Génétique Moléculaire de Toulouse. Les données expérimentales qui en découlent ont été obtenues grâce à des techniques de visualisation reposant sur des avancées technologiques importantes, entre autre l'analyse génomique à haut débit pour le séquençage ChIP et les microscopies à super-résolution comme la localisation photo-activée. Une modélisation ambitieuse est justifiée par la confluence de ces différentes techniques. Nous résumons de manière non exhaustive dans le tableau 1.1 les données expérimentales et les efforts théoriques existants pour les interpréter. Nous distinguons alors les scénarios biologiques et les modèles physiques qui en sont inspirés. Ces éléments sont classés suivant les trois principales problématiques explorées dans ce manuscrit : la distribution des protéines ParB à l'intérieur du cytoplasme, la cinétique des protéines ParB et ParA, la dynamique des séquences *parS* et des complexes de partition.

1.3.1 Distribution des protéines ParB...

... dans l'espace

La distribution spatiale des protéines ParB à l'intérieur du cytoplasme sont données par les résultats des expériences de SPT-PALM qui corréle deux techniques : le suivi de molécule unique (*single particle tracking* ou SPT) et la microscopie par localisation photo-activée (*photo-activated localization microscopy* ou PALM) [61, 62]. Cette dernière est une microscopie à champ large qui permet d'obtenir des images d'entités chimiques fluorescentes avec une résolution nanométrique. Le suivi de molécule unique permet d'obtenir des informations cinétiques sur les entités tracées. Dans le cas du système de partition, les protéines ParB sont marquées avec une protéine fluorescente verte (ou GFP pour *green fluorescent protein*). Ces expériences ont permis d'estimer à environ 600 le nombre moyen de protéines ParB dans l'espace intracellulaire d'une bactérie [22], et à environ 300 protéines par complexe. Elles montrent également qu'au moins 90% des protéines ParB existent dans les complexes ParBS en présence de la séquence *parS*. Le cytoplasme apparaît séparé entre une phase à haute densité en protéines ParB, brillante, et une phase à basse densité, très peu éclairée (voir figure 1.1). Les expérimentateurs sont capables de générer un mutant du plasmide F qui ne possèdent pas de séquence *parS* : les protéines ParB apparaissent alors distribuées de manière homogène sur le nucléoïde. Ces résultats indiquent que ces séquences d'ADN activent la formation des complexes ParBS. Dans la théorie des séparations de phase, la

Phénomène observable	Technique expérimentale	Scénario biologique	Modèle physique
ParB dans l'espace	SPT-PALM [22]	<i>Nucleation</i> et <i>caging</i> [22]	<ul style="list-style-type: none"> • <i>Spreading</i> et <i>Bridging</i> [55] • <i>Long Range Lattice Gas</i> [60] • Chapitre 2
ParB sur l'ADN	ChIP-Seq [56]	<i>Spreading</i> [57]	<ul style="list-style-type: none"> • <i>Stochastic binding</i> [22] • <i>Looping</i> et <i>Clustering</i> [59] • Chapitre 2
Photoblanchiment	FRAP [63]	∅	Chapitre 3
Positionnement	3D-SIM [54]	<i>Hitch-hiking</i> [54]	<ul style="list-style-type: none"> • Protéophorèse [41] • Chapitre 3

TABLE 1.1 – **Vue d'ensemble** non exhaustive des différentes contributions apportées à l'étude du système ParABS. La première colonne rassemble les phénomènes observés expérimentalement qui nous semblent pertinents pour comprendre la nature des complexes ParBS et le rôle joué par les protéines ParA dans la régulation et la ségrégation des complexes. La deuxième colonne précise les techniques expérimentales utilisées pour obtenir ces informations. La troisième colonne résume les scénarios biologiques correspondants qui permettent d'expliquer ces phénomènes. La quatrième colonne rassemble les modèles physiques respectifs et résume l'édifice théorique construit pour expliquer ces mêmes phénomènes à l'aide des outils de la physique statistique.

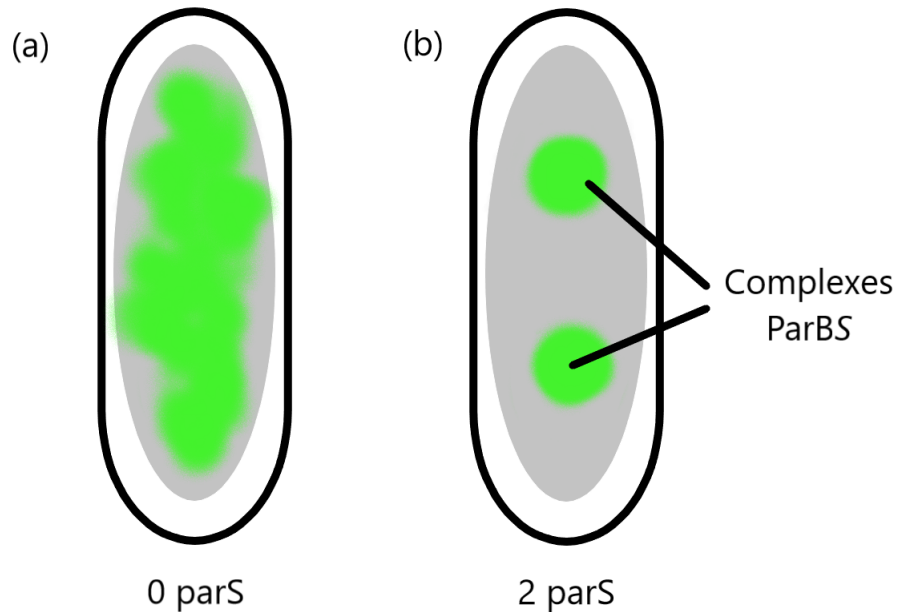


FIGURE 1.1 – **Schéma de la distribution spatiale des protéines ParB [22]** dans deux cas. Les bordures noires indiquent les frontières des bactéries. Le fond gris représente l'espace occupé par le nucléoïde. Les protéines ParB apparaissent en vert. Le cas (a) représente une bactérie mutante avec deux plasmides F sans séquence *parS*. Les protéines ParB apparaissent réparties de manière homogène sur le nucléoïde. Le cas (b) montre une cellule du type sauvage avec deux plasmides F possédant une séquence *parS*. Les protéines ParB forment alors deux complexes ParBS sur l'ADN autour des séquences *parS*.

séquence *parS* jouerait le rôle d'une graine de nucléation des complexes de protéines ParB. D'autres indices indiquent que les complexes ParBS résultent d'une séparation de phase au sens bien défini de la thermodynamique. La distribution de coefficient de diffusion des protéines ParB montre clairement deux comportements diffusifs distincts à l'intérieur et à l'extérieur des complexes ParBS [63]. De plus, les complexes ParBS connaissent des événements de fusion et se réarrangent rapidement. Des expériences *in vitro* de force-extension de l'ADN en présence de protéines ParB montrent que l'extension de l'ADN sans séquence *parS* à des concentrations élevées en protéines ParB suit la même courbe que l'extension du système avec une séquence *parS* et une concentration physiologique. En surexprimant les protéines ParB, le système générerait spontanément des complexes de protéines sans séquence *parS*, caractérisant un changement de régime de stabilité thermodynamique [64]. La formation des complexes ParBS a néanmoins été d'abord expliquée sans faire appel au concept de séparation de phase [22]. Le scénario biologique de *nucleation* et *caging* repose sur les interactions dynamiques entre les protéines ParB et l'ADN. La séquence *parS* capte une protéine ParB et les interactions entre protéines

mènent à une structure effective qui est le complexe *ParBS*. Les conditions d'existence des complexes *ParBS* ont été explorées numériquement avec une approche hamiltonienne nommée le *spreading* et *bridging* [55], pour des chaînes finies, loin de la limite thermodynamique. Ces simulations montrent ainsi ce qui pourrait être interprété comme le précurseur pour un système de taille finie d'une séparation de phase dans une limite d'interaction forte entre protéines sur l'ADN. Elles n'ont pas suffi à obtenir une vision d'ensemble du diagramme de phase d'un tel système.

... sur l'ADN

Les expériences de séquençage ChIP suivent deux étapes : l'immunoprécipitation de la molécule d'ADN liée aux protéines *ParB* afin d'isoler les morceaux d'ADN qui interagissent avec une protéine, et un séquençage à haut débit de ces pièces qui permet de reconnaître sur quelle séquence d'ADN les protéines étaient adsorbées [65]. Sur un grand nombre de bactéries, elles permettent de récupérer l'occupation moyenne des protéines *ParB* sur l'ADN en fonction de la coordonnée génomique à partir de la séquence *parS* (voir figure 1.2) [56]. Ces

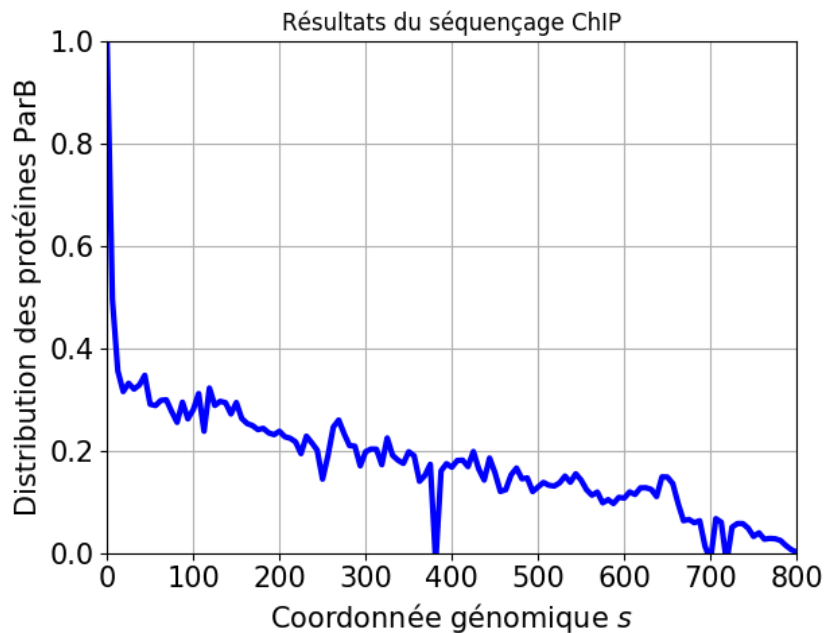


FIGURE 1.2 – **Résultats des expériences de séquençage ChIP [56]** : Ces données donnent la distribution moyenne des protéines *ParB* en fonction d'une coordonnée génomique définie à $s = 0$ pour la séquence *parS*. Cette distribution correspond à la structure tridimensionnelle du complexe *ParBS*.

données montrent donc la statistique moyenne unidimensionnelle des protéines *ParB* formant effectivement un complexe *ParBS* dans l'espace tridimensionnel.

Cette distribution a été partiellement expliquée dans le cadre du scénario de *nucleation* et *caging*. Les protéines ParB se lient de manière spécifique à la séquence *parS* et de manière aléatoire sur le reste de la chaîne. Les interactions entre protéines ParB introduisent un biais qui aboutit à la décroissance de la distribution à partir de la séquence *parS* [22]. Le modèle de *stochastic binding* explique qu'à longue portée, la distribution des ParB décroît algébriquement. Dans la limite des interactions entre protéines fortes, le *spreading* et *bridging* montrent des profils triangulaires loin de la distribution expérimentale. Le modèle de *looping* et *clustering*, qui prend en compte l'entropie des boucles d'ADN, formées à partir d'un complexe de partition supposé existant, réussit à décrire complètement la distribution des protéines ParB autour des séquences *parS* [59]. Ce dernier modèle sera plus précisément décrit dans notre Chapitre 2.

1.3.2 Echange entre deux complexes ParBS

Dans les expériences de recouvrement de fluorescence après photoblanchiment, ou *fluorescence recovery after photobleaching* (FRAP) en anglais [66], l'observation se concentre sur une ou plusieurs espèces chimiques rendues détectables grâce à un radical fluorescent attaché à la molécule à l'intérieur du volume cytoplasmique. Grâce à une lumière excitatrice, on peut supprimer de manière irréversible la fluorescence d'une partie des protéines comprises dans une région définie de l'espace intracellulaire. Cette étape s'appelle le photoblanchiment. Une partie des molécules a conservé sa fluorescence et diffuse dans le cytoplasme. Un recouvrement de la fluorescence dans la région photoblanchie doit apparaître. La cinétique de ce recouvrement est riche en informations sur la diffusion des molécules, leur localisation ou la proportion de molécules liées à l'ADN par exemple.

Dans le cadre du système de partition ParABS, les protéines ParB sont rendues fluorescentes à l'aide de la protéine GFP, comme dans les expériences de SPT-PALM. Lorsque les foyers ParBS sont positionnés, on observe ainsi deux tâches lumineuses situées en moyenne aux positions 1/4 et 3/4 de la bactérie. Un seul complexe ParBS est photoblanchi. Les résultats de ces expériences sur une population de 146 bactéries sont exposés en figure 1.3 [63]. Les deux complexes échangent des protéines ParB via le cytoplasme, nous observons ainsi une récupération de la fluorescence dans le foyer photoblanchi et une perte de fluorescence dans le second. Nous constatons des temps typiques de récupération ou de perte de fluorescence de l'ordre de plusieurs minutes. Ces données sont cruciales pour comprendre la cinétique des protéines ParB à l'intérieur et à l'extérieur des complexes ParBS. Nous étudierons ces données dans le Chapitre 3.

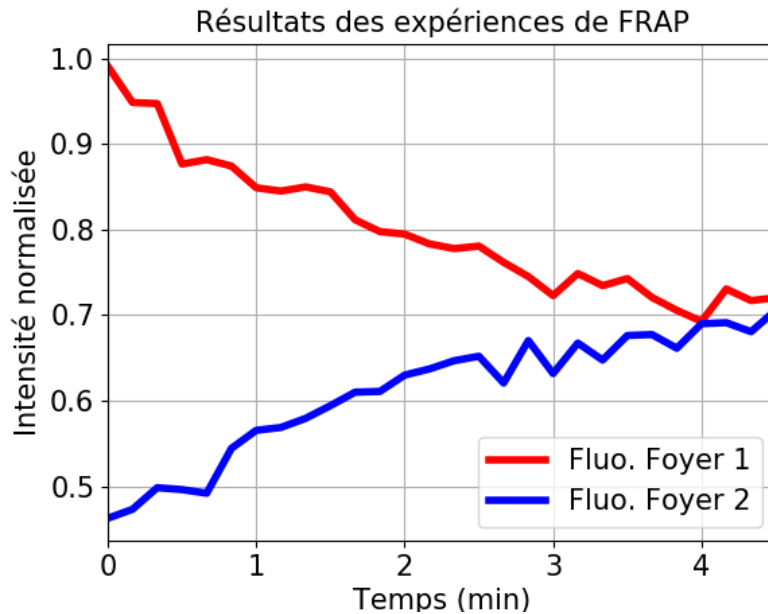


FIGURE 1.3 – Résultats des expériences de photoblanchiment sur les complexes ParBS [63] : Intensités lumineuses des complexes photoblanchi (en bleu) et non-photoblanchi (en rouge). Les intensités sont normalisées par l'intensité lumineuse du complexe non-photoblanchi à l'instant correspondant à la fin du photoblanchiment. On observe un transfert de fluorescence, et donc de protéines ParB, d'un complexe à un autre.

1.3.3 Positionnement des complexes ParBS

Des informations sur les positionnements collectifs moyens des protéines ParA et ParB peuvent être obtenus en utilisant la microscopie à illumination structurée à trois dimensions (3D-SIM pour *structured illumination microscopy*) [67]. Cette technique permet de capter l'émission fluorescente des protéines marquées dans l'espace cytoplasmique à trois dimensions et de dessiner leur distribution spatiale suivant les sections longitudinales et latérales d'une bactérie. Ces expériences permettent de montrer que les complexes ParBS se positionnent en périphérie du nucléoïde lorsque les protéines ParA sont déplétées [54]. De plus, ils sortent du nucléoïde pour les pôles de la bactérie lorsque l'hydrolyse de l'ATP par des protéines ParA mutantes n'est plus catalysée par les protéines ParB, c'est-à-dire lorsque l'hydrolyse par les protéines ParA est globalement faible dans le milieu intracellulaire. Les protéines ParA et l'hydrolyse de l'ATP sont ainsi nécessaires à une ségrégation fonctionnelle et donc au bon fonctionnement du système de partition. Ces expériences ont également pu montrer que les protéines ParA forment des oligomères colocalisés avec des régions à haute densité de l'ADN du chromosome bactérien. Ces oligomères semblent escorter les complexes ParBS. Cette observation est expliquée par le scénario de *hitch-hiking*. Les complexes ParBS provoquent l'hydrolyse des ATP

des protéines ParA sur l'ADN, ce qui entraîne le déplacement des complexes ParB*S*. Les protéines ParA sont libérées sous forme ADP dans le cytoplasme. Elle se lie à nouveau à l'ADN lorsque leur ATP est recyclé. Ce cycle se répète jusqu'au positionnement des complexes ParB*S*. Physiquement, ce positionnement des complexes ParB*S* aux positions 1/4 et 3/4 a été expliqué par un modèle de protéophorèse [41], c'est-à-dire à l'aide de la force en volume (contrairement à une force de chimiophorèse en surface) des protéines ParA sur le complexe ParB*S*. Les complexes ParB*S* remontent les gradients de concentration de protéines ParA par l'effet de cette force. Ce modèle explique que les complexes sont positionnés si et seulement les gradients des protéines ParA à gauche et à droite des complexes ParB*S* sont symétriques. Nous développons les résultats de ce modèle dans notre Chapitre 3.

1.4 Plan de la thèse

Deux phénomènes principaux émergent du système ParABS. Le premier est la formation de complexes ParB*S* tels qu'ils sont décrits par les expériences de SPT-PALM et de séquençage ChIP, sans consommation d'ATP, c'est-à-dire sans utilisation d'énergie. Le second est le rôle des protéines ParA et de l'hydrolyse dans le maintien d'une part et la ségrégation d'autre part des complexes ParB*S* tels que montrés dans les expériences de photoblanchiment et de 3D-SIM. C'est à l'aide des outils de la physique statistique que nous apporterons des éléments d'explication à ces deux phénomènes.

Nous distinguons ainsi trois types d'étude du système de partition :

- la statique thermodynamique, c'est-à-dire l'étude de l'auto-assemblage que forme un complexe ParB*S* et qui ne concerne que l'ADN, la séquence *parS* et les protéines ParB à l'équilibre thermodynamique ;
- la cinétique, c'est-à-dire l'étude des propriétés de diffusion des protéines ParA et ParB, la convergence vers l'état d'équilibre thermodynamique ou vers un état stationnaire hors équilibre, et notamment l'échange des protéines ParB entre deux complexes ParB*S* ;
- la dynamique ou l'étude du positionnement et de la ségrégation des complexes ParB*S* dans l'espace intracellulaire, processus hors équilibre.

Nous organisons notre propos en deux chapitres centraux. Le Chapitre 2 traite des conditions de formation des complexes ParB*S* dans l'hypothèse d'une séparation de phase des protéines ParB sur le nucléoïde. Dans cette étude, nous isolons conceptuellement les protéines ParB et la séquence *parS* pour étudier la statique du système de partition. Nous chercherons ainsi à caractériser les différents régimes accessibles au système constitué de l'ADN du plasmide F et des protéines ParB à l'équilibre thermodynamique. Ce chapitre contient deux approches. Une première approche en champ moyen nous permet de définir les quantités et les concepts pertinents à notre étude, et d'expliquer à l'aide du

modèle simple et analytique du gaz sur réseau avec interactions à courte portée les conditions d'existence d'un complexe ParBS à partir du diagramme de cette séparation de phase. Nous montrons que le régime physiologique des protéines ParB sur l'ADN est la coexistence métastable, au regard d'une distribution homogène, des états de haute et basse occupation sur la chaîne, seul régime pouvant expliquer les résultats des expériences de SPT-PALM. La deuxième est une approche où la statistique du polymère ADN est prise en compte et où nous distinguons les interactions de *spreading* sur la chaîne et de *bridging* dans l'espace. Nous montrons que ces interactions de *bridging* dans l'espace tridimensionnel correspondent à des interactions effectives à longue portée sur la chaîne unidimensionnelle formée par l'ADN. Nous montrons que ces interactions à longue portée dépendent algébriquement des exposants statistiques du polymère ADN. Cette dépendance nous permet de démontrer un critère général d'existence des séparations de phase unidimensionnelle de particules interagissant sur un polymère. Ce critère est rempli dans le cas de l'ADN. En utilisant une méthode variationnelle judicieusement testée sur le cas exact, dans la limite thermodynamique, du gaz sur réseau avec interactions à portée infinie, nous déterminons le diagramme de phase de cette séparation de phase sur l'ADN : sa température critique dépend de la statistique de l'ADN, de la portée des interactions de *bridging* et des intensités des interactions entre protéines ParB. Nous montrons à nouveau que le système biologique peut exister dans un régime de coexistence métastable.

Dans le Chapitre 3, nous développons d'abord un attirail théorique, couplé à plusieurs simulations numériques, pour interpréter les expériences de photoblanchiment. Ces expériences nous donnent des informations cruciales sur la cinétique des protéines ParB. Trois approches sont présentées. La première approche phénoménologique nous permet d'extraire des temps typiques d'existence des protéines ParB à l'intérieur et à l'extérieur des complexes. La deuxième, intermédiaire, nous permet de simuler *in silico* ces expériences de photoblanchiment. La dernière approche, basée sur un gaz sur réseau à plusieurs espèces traité en champ moyen dans la limite continu, nous oblige à intégrer les protéines ParA dans notre étude pour expliquer correctement les expériences de FRAP. En effet, si les complexes ParBS résultent d'une séparation de phase alors il faut déterminer le mécanisme qui permet au système biologique de ségréger deux complexes équilibrés aux positions 1/4 et 3/4 de l'espace intracellulaire. Cette dernière approche permet de proposer un scénario d'équilibrage des complexes ParBS par les protéines ParA et l'hydrolyse de l'ATP grâce à un mécanisme d'inversion de la maturation d'Ostwald, et de reproduire les résultats des expériences de photoblanchiment. Cette approche permet aussi d'étudier la ségrégation des complexes ParBS en ajoutant des équations dynamiques pour les séquences *parS*. Nous étudions une première approche proposée pour expliquer la ségrégation à partir d'un terme de force

dite de protéophorèse. Le gaz sur réseau à plusieurs espèces développée dans ce chapitre permet d'expliquer l'origine de cette force phénoménologique et d'expliquer l'équilibre des complexes ParBS grâce à un mécanisme d'inversion de la maturation d'Ostwald. Cette dernière approche permet aussi de produire une simulation numérique intégrant tous les éléments du système de partition et montrant un positionnement des complexes ParBS.

Ces différents travaux nous permettent d'esquisser en conclusion (Chapitre 4) une explication globale du système de partition ParABS et de la ségrégation du génome dans le cas du plasmide F d'*E. coli*.

Chapitre 2

Statique thermodynamique

2.1 Introduction

Dans ce premier chapitre, nous cherchons à expliquer les résultats des expériences de SPT-PALM et séquençage ChIP (voir chapitre Introduction et figures 1.1 et 1.2) sur la distribution des protéines ParB dans l'espace intracellulaire et sur l'ADN bactérien. Ces deux expériences apportent deux résultats majeurs. Les expériences de SPT-PALM montre que la séquence *parS* active la formation des complexes ParBS tridimensionnelle tandis que le séquençage ChIP nous donne accès à la distribution unidimensionnelle des protéines ParB autour de la séquence *parS*.

La formation des complexes ParBS s'inscrit dans la problématique générale de formation de structure tridimensionnelle essentielle à la vie cellulaire à partir du substrat unidimensionnel que forme la molécule d'ADN [68, 69] et des interactions entre particules et polymères [70]. Il est intéressant de noter que l'organisation de l'ADN, le repliement de la chromatine et les interactions entre domaines topologiques chez les cellules eucaryotes entrent aussi dans ces problématiques [71, 72, 73] et dans celle des transitions de phase [74, 75, 76]. Un motif commun à ces différents systèmes biologiques est la formation de boucle d'ADN par des interactions de *bridging* entre protéines. Trois types de modèles ont ainsi pu être étudiés, principalement à l'aide d'approche de type Flory et de simulations numériques :

1. des modèles de polymères dont les monomères interagissent directement entre eux [77, 78];
2. des modèles de polymères interagissant avec des particules qui peuvent se lier simultanément sur deux sites distincts et ainsi former un pont [78, 79];
3. des modèles de polymères interagissant avec des particules qui, une fois adsorbées, interagissent entre elles via des interactions de *spreading* et de *bridging* [32, 55, 80].

Notre système biologique fait partie de cette troisième catégorie. Malgré des études numériques poussées [55, 78, 81], les conditions d'existence et la nature des transitions de phase qui aboutissent à un ordre tridimensionnel observable demeure vague.

Nous proposons ainsi deux approches de physique statistique pour expliquer la formation des complexes ParBS comme le résultat d'une séparation de phase. La première approche se base sur un modèle très simplifié du système biologique constitué de l'ADN et des protéines ParB : traitée en champ moyen, elle sera l'occasion de définir les concepts et relations utilisés dans ce manuscrit. La deuxième approche forme un cadre théorique complet pour étudier les systèmes biologiques de la catégorie (3) de la précédente classification. Elle permet d'écrire un critère d'existence des transitions de phase du premier ordre dans ce type de système et d'obtenir plusieurs résultats analytiques dans une approche qui va au-delà du champ moyen. Dans ces deux développements, nous expliciterons le régime thermodynamique du système biologique. Nous montrons finalement, à l'aide des expériences de séquençage ChIP et des résultats du modèle de *looping* et *clustering*, comment cette séparation de phase de protéines sur l'ADN, espace unidimensionnel, aboutit à la formation d'une structure tridimensionnelle, le complexe ParBS.

Ces approches utilisent des concepts de la physique statistique : nous rappelons en premier lieu quelques relations essentielles à la compréhension de ces travaux.

2.2 Rappels de physique statistique

Nous terminons cette introduction en rappelant très succinctement quelques relations classiques de la physique statistique utilisées dans ce manuscrit [82]. La description d'un système physique dépend des échanges d'énergie et de matière avec son environnement : de ces conditions d'échanges découlent des distributions statistiques différentes.

2.2.1 Distribution canonique

On parle de distribution ou d'ensemble canonique lorsque le système considéré contient un nombre fixe d'unités élémentaires, comme des particules ou des spins, et peut échanger de l'énergie avec un thermostat. La température T d'équilibre est la température de ce thermostat. Il est possible d'associer au système un hamiltonien canonique \mathcal{H}_C donnant l'énergie macroscopique du système selon la configuration microscopique des éléments qui le compose. Dans un cas discret, choisi pour illustrer notre propos, la probabilité P_k que l'hamiltonien \mathcal{H}_C prenne la valeur d'énergie particulière E_k est donné par le

facteur de Boltzmann :

$$P_k \sim \exp\left(-\frac{E_k}{k_B T}\right) \quad (2.1)$$

à une constante de normalisation près, et où $k_B \approx 1,38 \times 10^{-23} \text{ J/K}$ est la constante de Boltzmann. La constante de normalisation vérifie que $\sum_k P_k = 1$ d'où :

$$P_k = \frac{1}{\mathcal{Z}_C} \exp\left(-\frac{E_k}{k_B T}\right) \quad (2.2)$$

où \mathcal{Z}_C définit la fonction de partition de l'ensemble canonique

$$\mathcal{Z}_C = \sum_k \exp[-\beta \mathcal{H}_C] \quad (2.3)$$

avec $\beta = (k_B T)^{-1}$. La fonction de partition prend ainsi en compte toutes les configurations microscopiques possibles du système. Une autre quantité importante est l'énergie libre de Helmholtz notée \mathcal{F} . Toutes les quantités thermodynamiques importantes du système, telles que l'entropie, la pression ou le potentiel chimique peuvent être dérivées de l'énergie libre \mathcal{F} . L'énergie libre est simplement reliée à la fonction de partition \mathcal{Z}_C par la relation :

$$\mathcal{F} = -k_B T \ln \mathcal{Z}_C \quad (2.4)$$

Une autre définition de l'énergie libre est donnée par la thermodynamique :

$$\mathcal{F} = U - TS \quad (2.5)$$

où U et S sont respectivement l'énergie interne et l'entropie du système. La pression P et le potentiel chimique μ dérivent de l'énergie libre de Helmholtz. Dans le cadre d'un gaz sur réseau, nous considérons M particules sur un réseau de taille N et de volume V . Chaque site possède un volume $V_0 = V/N$ et une occupation $\Phi = M/N$. Pour la pression, nous avons ainsi à température et nombre de particules constants :

$$P = -\left(\frac{\partial \mathcal{F}}{\partial V}\right)_{T,M} = -\frac{1}{V_0} \left(\frac{\partial \mathcal{F}}{\partial N}\right)_{T,M} \quad (2.6)$$

et dans le cas homogène pour le potentiel chimique à température et volume constants :

$$\mu = \left(\frac{\partial \mathcal{F}}{\partial M}\right)_{T,V} = \frac{1}{N} \left(\frac{\partial \mathcal{F}}{\partial \Phi}\right)_{T,V}. \quad (2.7)$$

2.2.2 Distribution grand canonique

On adopte la distribution grand canonique lorsque le système considéré est plongé dans un réservoir de particules avec lequel il peut échanger de l'énergie

et de la matière. Identiquement à l'ensemble canonique, la température d'équilibre T est celle du réservoir. On associe au système un hamiltonien dit grand canonique \mathcal{H}_{GC} comptant l'énergie du système et qui dépend d'un potentiel chimique μ . En effet, la présence d'une entité chimique dans le système étudié contribue à diminuer l'énergie du système d'un facteur d'énergie μ . Un micro-état k est ainsi caractérisé par une énergie E_k et une nombre de particules N_k . L'hamiltonien s'écrit en conséquence $\mathcal{H}_{GC} = E_k - \mu N_k$. La probabilité P_k qu'un micro-état k soit adopté par le système est donnée par le facteur de Boltzmann :

$$P_k = \frac{1}{\mathcal{Z}_{GC}} \exp[-\beta(E_k - \mu N_k)] \quad (2.8)$$

où \mathcal{Z}_{GC} est la fonction de partition grand canonique, définie telle que $\sum_k P_k = 1$, soit :

$$\mathcal{Z}_{GC} = \sum_k \exp[-\beta(E_k - \mu N_k)]. \quad (2.9)$$

Le nombre de particules dans le système modifie la probabilité du micro-état k par rapport au cas canonique. La fonction de partition grand canonique définit un potentiel thermodynamique appelé potentiel grand canonique Ω qui s'écrit :

$$\Omega = -k_B T \ln \mathcal{Z}_{GC} \quad (2.10)$$

Dans le cas d'un gaz sur réseau, l'occupation moyenne Φ sur le réseau est dérivée du potentiel grand canonique à température et volume constant :

$$\Phi = -\frac{1}{N} \left(\frac{\partial \Omega}{\partial \mu} \right)_{T,V} \quad (2.11)$$

L'énergie libre de Helmholtz \mathcal{F} et le grand potentiel Ω sont liés par une transformation de Legendre. On a en effet dans le cas homogène :

$$F = \Omega + M\mu \quad (2.12)$$

2.3 Gaz sur réseau à courte portée

Dans cette première approche, le cadre théorique d'interactions entre protéines et ADN se trouve très simplifié. Cette stratégie offre des résultats analytiques simples pour un nombre minimal de paramètres. Bien qu'elle pêche par de nombreuses limites à cause de l'approximation du champ moyen, nous verrons qu'elle n'est pas sans intérêt dans la limite basse densité.

2.3.1 Gaz sur réseau à D dimensions

Nous assimilons le volume V occupé par le nucléoïde, volume d'occupation de l'ADN, au volume décrit par le cytoplasme bactérien. Ce volume est divisé

en N volumes V_0 de telle sorte que chaque volume élémentaire V_0 contient exactement la longueur d'ADN nécessaire pour adsorber une protéine ParB. Nous formons ainsi un réseau d'ADN composé de N sites d'adsorption de protéines. Pour des arguments de simplicité, nous définissons ce réseau comme cubique simple avec un paramètre de maille ℓ_m et un nombre de coordination $q = 2D$ où D est la dimension du système. Nous considérons enfin que l'ADN *in vivo* est suffisamment compact pour qu'il n'existe aucun volume V_0 vide. Nous assimilons ainsi le réseau à l'ADN lui-même. Ce réseau est mis en contact avec un réservoir de particules ou protéines ParB. Nous adoptons ainsi une approche grand canonique. Chaque site i possède alors deux attributs :

1. une occupation Φ_i qui prend la valeur 1 si une protéine ParB est adsorbée sur ce site, et 0 sinon ;
2. une énergie d'adsorption spécifique ϵ_i qui permet de spécifier une séquence *parS* par exemple.

De fait, le modèle construit fait partie de la famille des modèles de gaz sur réseau [83]. Dans la configuration donnée, nous pouvons considérer que deux particules situées dans deux petits volumes proches dans l'espace interagissent comme deux particules sur deux sites voisins de la chaîne d'ADN : nous ne faisons pas de distinction entre les interactions de *spreading* et de *bridging*. Deux protéines adjacentes sur le réseau interagissent avec une énergie J^1 et chaque protéine adsorbée contribue à baisser l'énergie du système d'une quantité $\mu - \epsilon_i$ où μ est le potentiel chimique qui régit l'occupation moyenne des particules sur le réseau. L'hamiltonien grand canonique de ce gaz sur réseau avec interactions à courte portée (SRLG pour Short Range Lattice Gas) $\mathcal{H}_{\text{SRLG}}$ dépend de l'ensemble des Φ_i et s'écrit :

$$\mathcal{H}_{\text{SRLG}}(\{\Phi_i\}) = -\frac{J}{2} \sum_{\langle i,j \rangle} \Phi_i \Phi_j - \sum_{i=1}^N \Phi_i (\mu - \epsilon_i) \quad (2.13)$$

où $\langle i, j \rangle$ indique une somme sur tous les couples de sites i et j premiers voisins du réseau. Nous appelons M , le nombre moyen de particules sur le réseau

$$M = \sum_{i=1}^N \langle \Phi_i \rangle, \quad (2.14)$$

et Φ , la densité moyenne en protéine du système

$$\Phi = \frac{M}{N} = \frac{1}{N} \sum_{i=1}^N \langle \Phi_i \rangle \quad (2.15)$$

1. Dans ce modèle, nous ne cherchons pas à distinguer les interactions entre deux protéines proches le long de la chaîne de celles entre deux protéines proches dans l'espace.

où $\langle Y \rangle$ est la moyenne statistique de la grandeur Y sur l'ensemble définie par H_{SRLG} . Par ailleurs, grâce à la définition de notre réseau, la densité ρ_i de particules dans l'espace et l'occupation Φ_i locales sont simplement reliées par l'expression :

$$\rho_i = \frac{\Phi_i}{\sigma^D} \quad (2.16)$$

Nous distinguons deux situations : un cas inhomogène présenté jusqu'à présent et un cas homogène. Nous parlerons de cas homogène lorsque l'occupation d'un site ne dépend pas du site en question $\langle \Phi_i \rangle = \Phi$ et $\varepsilon_i = \varepsilon$ pour tout site i . Nous identifions ce cas homogène à la situation des plasmides F mutants sans séquence *parS* : en effet, dans ce cas, il n'y a pas l'énergie spécifique d'adsorption venant de *parS* et il ne reste que l'énergie d'adsorption non spécifique qui est non nulle [22, 55]. Dans les faits, nous prendrons quand même $\varepsilon = 0$ dans le cas homogène puisque cette quantité ne possède qu'un rôle d'ajustement du potentiel chimique dans cette étude de statique thermodynamique. Le cas inhomogène est identifié au cas sauvage, pour des plasmides F possédant effectivement une séquence *parS*.

Bien que ce modèle ne concerne qu'une seule espèce chimique, il s'agit en fait d'un modèle à deux états, 0 pour des sites vides ou 1 pour des sites occupés. Nous pouvons réécrire l'hamiltonien $\mathcal{H}_{\text{SRLG}}$ en utilisant la variable complémentaire h_i qui compatibilise les trous au lieu des sites occupés, auquel cas $h_i = 1$ si le site i est vide et 0 sinon, donc $\Phi_i + h_i = 1$. Il y a complémentarité entre un système de particules ou de trous : nous parlerons de symétrie particule-trou. Il y a invariabilité de la thermodynamique du système par le changement de variable $\Phi_i \rightarrow 1 - \Phi_i = h_i$: l'hamiltonien $\mathcal{H}_{\text{SRLG}}$ reste le même à une constante additive près qui ne modifie donc en rien le comportement thermodynamique du système. Nous verrons dans les paragraphes suivants les conséquences de cette symétrie dans notre description des séparations de phases.

2.3.2 Gaz sur réseau et modèle d'Ising

La formulation hamiltonienne précédente du gaz sur réseau est très semblable à la formulation du problème de Lenz et Ising sur le magnétisme proposée par Pauli [84, 85]. L'hamiltonien du modèle d'Ising peut s'écrire très généralement :

$$\mathcal{H}_I = -J_I \sum_{\langle i,j \rangle} \sigma_i \sigma_j - K \sum_{\langle i,j \rangle} (\sigma_i + \sigma_j) - \sum_{i=1}^N h_i \sigma_i \quad (2.17)$$

où σ_i prend les valeurs $+1$ (*spin up*) ou -1 (*spin down*). Le modèle d'Ising et le gaz sur réseau sont isomorphes et les relations entre les constantes d'interaction sont déduites de la transformation $\sigma_i = 2\Phi_i - 1$. On trouve alors $J_I = K = \frac{J}{4}$ et $h_i = \frac{\mu - \varepsilon_i}{2}$. Au-delà de ces aspects purement formels, ces deux modèles

présentent une différence d'approche : le modèle d'Ising résulte d'une approche canonique et possède par conséquent un nombre constant de spins, tandis que le gaz sur réseau tel que défini possède un nombre non-conservé de particules contrôlé par le potentiel chimique μ .

2.3.3 Solution exacte pour $D = 1$

Cette isomorphie permet d'appliquer les méthodes efficaces pour le modèle d'Ising au gaz sur réseau. Entre autre, la méthode des matrices de transfert permet de résoudre analytiquement notre modèle à une dimension ($D = 1$) [86]. Ce rappel sera l'occasion de définir les quantités que nous utiliserons dans ce chapitre. La fonction de partition du gaz sur réseau peut s'écrire dans le cas homogène :

$$\begin{aligned} \mathcal{Z}_{\text{SRLG}} &= \sum_{\{\Phi_i\}} \exp \left(\beta J \sum_{i=1}^N \Phi_i \Phi_{i+1} + \beta \frac{\mu}{2} \sum_{i=1}^N (\Phi_i + \Phi_{i+1}) \right) \\ &= \sum_{\{\Phi_i\}} \prod_{i=1}^N \langle \Phi_i | G | \Phi_{i+1} \rangle \end{aligned} \quad (2.18)$$

avec par définition $\langle \Phi_i | G | \Phi_{i+1} \rangle = \exp [\beta J \Phi_i \Phi_{i+1} + \beta \frac{\mu}{2} (\Phi_i + \Phi_{i+1})]$. La matrice G est une matrice 2x2 dont les éléments $\langle \Phi_i | G | \Phi_{i+1} \rangle$ dépendent de la représentation $\{|\Phi_i\rangle\}$ choisie. Nous faisons le choix d'une base orthonormée telle que :

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (2.19)$$

Nous avons donc d'une part

$$G = \begin{pmatrix} 1 & e^{\beta \frac{\mu}{2}} \\ e^{\beta \frac{\mu}{2}} & e^{\beta(J+\mu)} \end{pmatrix} \quad (2.20)$$

et d'autre part, du fait que $\sum_{\Phi=\pm 1} |\Phi\rangle\langle\Phi| = 1$, nous obtenons une expression simple de la fonction de partition d'un gaz sur réseau à une dimension :

$$\mathcal{Z}_{\text{SRLG}} = \sum_{\Phi=\pm 1} \langle \Phi | G^N | \Phi \rangle = \text{Tr } G^N = \lambda_+^N + \lambda_-^N \quad (2.21)$$

d'après la définition de la trace d'une matrice. La matrice G est diagonalisable et λ_+ et λ_- sont ses deux valeurs propres, définies comme :

$$\lambda_{\pm} = e^X \left(\cosh X \pm \sqrt{\sinh^2 X + B} \right) \quad (2.22)$$

avec $X = \beta(J + \mu)/2$ et $B = \exp(-\beta J)$. La limite thermodynamique est donnée par $N \rightarrow +\infty$. Dans cette limite, la fonction de partition est réduite

2.3. GAZ SUR RÉSEAU À COURTE PORTÉE

à $\mathcal{Z}_{\text{SRLG}} = \lambda_+^N$ puisque $\lambda_+ > \lambda_-$. Nous obtenons finalement l'expression du potentiel grand canonique dans la limite thermodynamique :

$$\Omega_{\text{SRLG}} = -Nk_B T \ln \lambda_+. \quad (2.23)$$

L'occupation moyenne Φ est définie par la dérivée

$$\Phi = -\frac{1}{N} \left(\frac{\partial \Omega_{\text{SRLG}}}{\partial \mu} \right)_{T,V} = \frac{1}{2} \left(1 + \frac{\sinh X}{\sqrt{\sinh^2 X + B}} \right). \quad (2.24)$$

Le potentiel chimique μ en fonction de Φ peut être déterminé en inversant la précédente expression :

$$\mu = 2k_B T \ln \left(\sqrt{1 - A^2 + BA^2} - A\sqrt{B} \right) - k_B T \ln (1 - A^2) - J. \quad (2.25)$$

avec $A = 2\Phi - 1$. La pression P du gaz sur réseau à une dimension vérifie l'égalité $\Omega_{\text{SRLG}} = -PV$, d'où :

$$P = \frac{k_B T}{V_0} \ln \lambda_+. \quad (2.26)$$

Un autre résultat utile est l'expression des corrélations à deux sites dans la limite thermodynamique [86] :

$$\langle \Phi_i \Phi_j \rangle = \Phi^2 + \Phi (1 - \Phi) e^{-|i-j|/\xi} = \Phi^2 + \Phi (1 - \Phi) r_{\text{LG}}^{|i-j|} \quad (2.27)$$

où $\xi = -1/\ln r_{\text{LG}}$ est la longueur de corrélation à deux sites du gaz sur réseau. Elle dépend du rapport des valeurs propres $r_{\text{LG}} = \lambda_-/\lambda_+$. Une simple analyse montre que quelque soit la valeur de la température T , les variations de la pression et du potentiel chimique sont monotones : comme le modèle d'Ising, un gaz sur réseau à une dimension ne présente pas de comportement de transition de phases. Landau a donné un argument physique très simple pour expliquer l'absence de transition ferromagnétique dans les systèmes unidimensionnels en utilisant le modèle d'Ising [87]. Pour les dimensions supérieures ou égales à $D = 1$, nous pouvons mener toute l'étude thermodynamique et déterminer analytiquement, mais approximativement, toutes les grandeurs d'intérêt en utilisant la théorie du champ moyen.

2.3.4 Champ moyen et approche variationnelle

La théorie du champ moyen désigne un ensemble de méthodes permettant d'obtenir des approximations simples des grandeurs physiques d'un problème [88, 89]. Une approximation en champ moyen consiste toujours à négliger les corrélations entre les quantités physiques considérées, par exemple

nos occupations locales Φ_i . Dans les faits, cette décorrélation est soit définie comme point de départ, comme dans l'approximation de Bragg-Williams, soit elle est une conséquence de principes variationnels ou d'approches permettant de confondre la distribution moyenne avec la distribution la plus probable des grandeurs du problème. Dans cette partie, nous retrouvons les expressions des quantités thermodynamiques d'un gaz sur réseau dans l'approximation du champ moyen dans le cas homogène en utilisant une approche variationnelle [90]. Cette méthode de calcul consiste en l'introduction dans l'hamiltonien du gaz sur réseau d'un ou de plusieurs champ fictifs, ou variationnels, de telle sorte qu'on puisse écrire :

$$\mathcal{H}_{\text{SRLG}} = \mathcal{H}_0 + \Delta\mathcal{H} \quad (2.28)$$

où \mathcal{H}_0 joue le rôle d'un hamiltonien de référence. La fonction de partition $\mathcal{Z}_{\text{SRLG}}$ correspondante devient :

$$\mathcal{Z}_{\text{SRLG}} = \mathcal{Z}_0 \times \frac{\sum_{\{\Phi_i\}} e^{-\beta(\mathcal{H}_0 + \Delta\mathcal{H})}}{\mathcal{Z}_0} = \mathcal{Z}_0 \langle \exp(-\beta\Delta\mathcal{H}) \rangle_0 \quad (2.29)$$

où $\langle \rangle_0$ désigne une moyenne dans l'ensemble statistique défini par \mathcal{H}_0 . Le terme $\langle \exp(\Delta\mathcal{H}) \rangle_0$ n'est en principe pas calculable analytiquement. On peut néanmoins majorer $\mathcal{Z}_{\text{SRLG}}$ en utilisant la propriété de convexité de la fonction exponentielle. D'où l'inégalité :

$$\mathcal{Z}_{\text{SRLG}} \leq \mathcal{Z}_0 \exp(-\beta\langle\Delta\mathcal{H}\rangle) \quad (2.30)$$

de laquelle est déduite une majoration du potentiel grand canonique Ω_{SRLG} est :

$$\Omega_{\text{SRLG}} \leq \Omega_V = \Omega_0 + \langle\Delta\mathcal{H}\rangle_0 \quad (2.31)$$

aussi connue sous le nom d'inégalité de Bogoliubov. Le potentiel grand canonique variationnel Ω_V offre une borne supérieure du potentiel grand canonique exact Ω_{SRLG} en fonction des paramètres variationnels introduits initialement. Dans l'espace des paramètres variationnels, la fonction Ω_V possède un minimum Ω_V^* . Il faut donc joindre à l'inégalité de Bogoliubov des équations d'optimisation de Ω_V . Le nombre de ces équations est donné par le nombre de paramètres variationnels introduits initialement. Nous retiendrons la valeur optimisée comme approximation du potentiel exact :

$$\Omega_{\text{SRLG}} \approx \Omega_V^* \quad (2.32)$$

Toute l'astuce de cette méthode consiste à choisir un \mathcal{H}_0 bien connu analytiquement. Le gaz sur réseau défini en première partie possède deux paramètres : le

2.3. GAZ SUR RÉSEAU À COURTE PORTÉE

couplage entre premiers voisins J et le potentiel chimique. Nous pouvons donc définir très généralement \mathcal{H}_0 dans le cas homogène comme

$$\mathcal{H}_0 = -\frac{J_0}{2} \sum_{\langle i,j \rangle} \Phi_i \Phi_j - \mu_0 \sum_{i=1}^N \Phi_i \quad (2.33)$$

où J_0 et μ_0 sont nos paramètres variationnels. Ainsi, il y existe trois types de calculs variationnels possibles :

1. On pose dans le cas le plus simple $J_0 = 0$. Le calcul variationnel devient un calcul en champ moyen puisque les corrélations, comprises dans le terme en J , n'entre pas en compte dans l'hamiltonien de référence \mathcal{H}_0 .
2. On pose $J_0 = J$. Ce cas est ici superflu pour un gaz sur réseau à courte portée. L'hamiltonien de référence \mathcal{H}_0 serait alors identique à celui du gaz sur réseau et ainsi plus complexe que le résidu $\Delta\mathcal{H}$. Il deviendrait utile dans le cadre d'un modèle plus complexe (voir section suivante).
3. On laisse J_0 tel quel. En plus des remarques du cas 2, cette définition permettrait de mieux tenir compte des corrélations entre premiers voisins, modifiées par un éventuel nouveau terme.

Pour obtenir les résultats dans l'approximation du champ moyen, nous avons donc

$$\mathcal{H}_0 = -\mu_0 \sum_{i=1}^N \Phi_i \quad (2.34)$$

et

$$\Delta\mathcal{H} = -J \sum_{\langle i,j \rangle} \Phi_i \Phi_j - (\mu - \mu_0) \sum_{i=1}^N \Phi_i. \quad (2.35)$$

Le potentiel d'essai Ω_V s'écrit

$$\Omega_V = -k_B T \ln \mathcal{Z}_0 - J \sum_{\langle i,j \rangle} \langle \Phi_i \Phi_j \rangle_0 - (\mu - \mu_0) \sum_{i=1}^N \langle \Phi_i \rangle_0 \quad (2.36)$$

avec \mathcal{Z}_0 facilement calculable. Nous avons en effet :

$$\mathcal{Z}_0 = \sum_{\{\Phi_i\}} \exp \left(\beta \mu_0 \sum_{i=1}^N \Phi_i \right) = [1 + \exp(\beta \mu_0)]^N. \quad (2.37)$$

L'occupation moyenne $\langle \Phi_i \rangle_0$ se calcule d'une manière similaire :

$$\langle \Phi_i \rangle_0 \equiv \Phi_0 = \frac{\exp(\beta \mu_0)}{1 + \exp(\beta \mu_0)}. \quad (2.38)$$

et par la même méthode, conformément à la théorie du champ moyen, nous pouvons montrer que dans l'ensemble statistique défini par \mathcal{H}_0 , les sites sont décorrélés :

$$\langle \Phi_i \Phi_j \rangle_0 = \Phi_0^2. \quad (2.39)$$

L'équation d'optimisation est dans ce cas :

$$\left(\frac{\partial \Omega_V}{\partial \mu_0} \right)_{\mu_0 = \mu_0^*} = 0. \quad (2.40)$$

La valeur optimisée μ_0^* est donnée par

$$\mu_0^* = \mu + Jq\Phi_0 \quad (2.41)$$

et en montrant simplement que l'occupation moyenne Φ , par sa définition,

$$\Phi \equiv -\frac{1}{N} \frac{\partial \Omega_{\text{SRLG}}}{\partial \mu} \approx -\frac{1}{N} \frac{\partial \Omega_V^*}{\partial \mu} = \Phi_0 \quad (2.42)$$

nous pouvons écrire l'expression du potentiel grand canonique Ω_{SRLG} dans notre approximation entièrement fonction de Φ :

$$\Omega_{\text{SRLG}} \approx \Omega_V^* = Nk_B T \ln(1 - \Phi) + \frac{1}{2} N J q \Phi^2 \quad (2.43)$$

Le potentiel chimique μ réel, approximé par le potentiel chimique μ_{MF} du champ moyen se déduit des expressions précédentes :

$$\mu \approx \mu_{\text{MF}} = k_B T \ln \left(\frac{\Phi}{1 - \Phi} \right) - Jq\Phi \quad (2.44)$$

et la pression réelle est donnée par le potentiel grand canonique puisque $\Omega_{\text{SRLG}} = -PV$, d'où

$$P \approx P_{\text{MF}} = -\frac{1}{V_0} \left[\frac{Jq}{2} \Phi^2 + k_B T \ln(1 - \Phi) \right]. \quad (2.45)$$

où P_{MF} est l'approximation en champ moyen de la pression. Connaissant le potentiel grand canonique et le potentiel chimique, nous pouvons calculer l'énergie libre de Helmholtz du système. Elle est donnée dans le cas homogène par la transformation de Legendre

$$\mathcal{F}_{\text{SRLG}} = \Omega_{\text{SRLG}} + N\mu\Phi. \quad (2.46)$$

Au-delà de cette transformation mathématique, il faut garder à l'esprit que passer du potentiel grand canonique à l'énergie libre de Helmholtz correspond à un changement d'ensemble au sens de la physique statistique. L'énergie $\mathcal{F}_{\text{SRLG}}$ résulte d'une approche canonique pour laquelle le nombre de particules est

2.3. GAZ SUR RÉSEAU À COURTE PORTÉE

conservé : il n'y a plus d'échange avec un réservoir médié par le potentiel chimique μ . Ainsi,

$$\mathcal{F}_{\text{SRLG}} \approx -\frac{1}{2}N J q \Phi^2 + N k_B T [\Phi \ln \Phi + (1 - \Phi) \ln (1 - \Phi)] \quad (2.47)$$

Cette énergie libre peut aussi s'écrire $\mathcal{F}_{\text{SRLG}} = U_{\text{MF}} + T S_{\text{MF}}$ avec :

$$U_{\text{MF}} = -\frac{1}{2}N J q \Phi^2 \quad (2.48)$$

est l'énergie interne du gaz sur réseau évaluée dans l'approximation du champ moyen, et

$$S_{\text{MF}} = N k_B [\Phi \ln \Phi + (1 - \Phi) \ln (1 - \Phi)] \quad (2.49)$$

l'entropie associée au gaz sur réseau. Le terme d'énergie interne U_{MF} représente l'énergie totale moyenne d'interaction entre particules sur le réseau : il y a $Nq/2$ interactions au total, d'intensité J chacune. L'entropie S_{MF} est une entropie de mélange entre les sites occupés et vides : en conséquence, ce terme est laissé tel quel par le changement de variable $\Phi_i \rightarrow 1 - \Phi_i$.

De manière traditionnelle, on étudie ces fonctions d'état en fonction de la température. Pour le système biologique considéré, la température ne semble pas être un paramètre pertinent puisque sa valeur doit avoisiner celle de la température ambiante $T_r = 300$ K. Néanmoins, l'intensité des interactions entre premiers voisins J n'est pas un paramètre connu. En plus de l'occupation moyenne Φ , il peut donc être intéressant d'utiliser le paramètre adimensionnel $\tilde{J} = \beta J$, au lieu de la température, pour mener notre étude thermodynamique. Par suite, on définit le potentiel chimique $\tilde{\mu}_{\text{MF}} = \beta \mu_{\text{MF}}$ et la pression $\tilde{P}_{\text{MF}} = \beta V_0 P_{\text{MF}}$ normalisés. La température est considérée à $T = T_r = 300$. Nous donnerons par la suite les résultats en température en juxtaposition et entre parenthèses avec ceux que nous considérons comme davantage pertinent.

Le paramètre $\tilde{J} = J/(k_B T)$ joue le rôle inverse à la température dans la théorie des transitions de phase du premier ordre. En effet, il existe une valeur critique \tilde{J}_c telle que pour tout $\tilde{J} > \tilde{J}_c$, la pression \tilde{P}_{MF} présente des variations non physiques caractérisée par :

$$\left(\frac{\partial \tilde{P}_{\text{MF}}}{\partial \Phi} \right)_{\Phi \in]\Phi_-, \Phi_+[} < 0. \quad (2.50)$$

Les valeurs Φ_- et Φ_+ sont définies par l'annulation de la dérivée précédente ou, autrement dit, par la divergence de la compressibilité isotherme χ_T :

$$\chi_T = -\frac{1}{V} \left(\frac{\partial V}{\partial P_{\text{MF}}} \right)_T. \quad (2.51)$$

La valeur critique \tilde{J}_c est obtenue lorsque Φ_- et Φ_+ sont racine double :

$$\tilde{J}_c^{\text{MF}} = \frac{4}{q} \left(T_c^{\text{MF}} = \frac{Jq}{4k_B} \right) \quad (2.52)$$

et l'occupation critique correspondante est :

$$\Phi_c = \frac{1}{2}. \quad (2.53)$$

La phénoménologie de notre gaz sur réseau est identique à celle d'un gaz réel de van der Waals. Le comportement de la pression (voir figure 2.1) permet de définir des états de haute et basse occupation en particules sur le réseau, c'est-à-dire dans notre cas en protéines sur l'ADN. Il existe des valeurs de

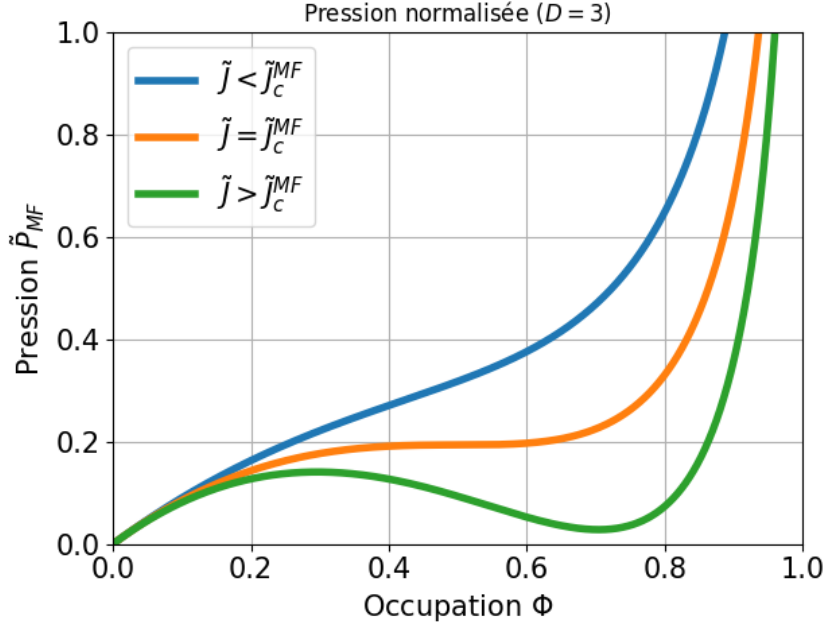


FIGURE 2.1 – **Pression en fonction de l'occupation** : la pression ne possède pas les mêmes variations analytiques selon les valeurs de \tilde{J} . Elle possède des variations non-physiques (pression décroissante pour occupation croissante) lorsque $\tilde{J} > \tilde{J}_c^{\text{MF}}$ qui définissent un régime de coexistence instable entre état de basse (branche droite) et haute (branche gauche) occupation.

Φ telles que l'état d'équilibre thermodynamique du système correspond à un volume divisé en domaines de haute occupation Φ_h et des domaines de basse occupation Φ_b . Cette séparation est appelée régime de coexistence. Les valeurs du couple (Φ_b, Φ_h) sont contraintes par le choix de \tilde{J} et la collection de ces couples définit la courbe de coexistence des états de haute et basse occupation. Une définition rigoureuse de la courbe de coexistence est donnée par l'équilibre

thermodynamique :

$$\begin{cases} \tilde{P}_{\text{MF}}(\Phi_b) = \tilde{P}_{\text{MF}}(\Phi_h) \\ \tilde{\mu}_{\text{MF}}(\Phi_b) = \tilde{\mu}_{\text{MF}}(\Phi_h) \end{cases} \quad (2.54)$$

La pression et le potentiel chimique ont les mêmes valeurs dans les phases de haute et basse occupations. Ce système d'équations est directement résoluble avec la méthode numérique et des paramètres initiaux appropriés : il faut en effet veiller à ce que le schéma numérique ne tombe pas dans la solution triviale homogène $\Phi_h = \Phi_b$. Il est néanmoins possible de déterminer une expression analytique pour la courbe de coexistence. La symétrie particule-trou du gaz sur réseau implique que les courbes du plan (Φ, \tilde{J}) admettent l'axe $\Phi = \Phi_c = 1/2$ comme axe de symétrie. Ainsi, pour la courbe de coexistence, la symétrie particule-trou implique que :

$$\Phi_h + \Phi_b = 1 \quad (2.55)$$

Cette dernière équation permet de réduire le système précédent à une seule équation dont la solution est l'expression analytique de la courbe de coexistence en champ moyen :

$$\tilde{J}_{\text{coex}}^{\text{MF}} = \frac{1}{q(\Phi - \Phi_c)} \ln \left(\frac{\Phi}{1 - \Phi} \right) \quad \left(T_{\text{coex}}^{\text{MF}} = \frac{Jq(\Phi - \Phi_c)}{k_B \ln \left(\frac{\Phi}{1 - \Phi} \right)} \right). \quad (2.56)$$

Ce résultat semble être nouveau. Cette équation donne la valeur du couplage de transition $\tilde{J}_{\text{coex}}^{\text{MF}}$. La collection des (Φ_-, Φ_+) forme la ligne spinodale, définie par la divergence de la compressibilité isotherme :

$$\tilde{J}_{\text{spino}}^{\text{MF}} = \frac{1}{q\Phi(1 - \Phi)} \quad (T_{\text{spino}}^{\text{MF}} = Jq\Phi(1 - \Phi)). \quad (2.57)$$

Cette dernière équation donne la valeur du couplage $\tilde{J}_{\text{spino}}^{\text{MF}}$ qui borne le régime instable. Ce couplage et le couplage de transition $\tilde{J}_{\text{coex}}^{\text{MF}}$ diminue quand q augmente comme le couplage critique \tilde{J}_c^{MF} . La courbe de coexistence et la ligne spinodale sont représentées dans le plan (Φ, \tilde{J}) en figure 2.2. Ce diagramme de phase permet de distinguer plusieurs territoires :

1. Pour $\tilde{J} < \tilde{J}_c^{\text{MF}}$, on parle généralement d'un état fluide pour lequel la distinction entre basse et haute occupation n'est pas pertinente. L'équilibre thermodynamique correspond à un système homogène en particules.
2. Pour $\tilde{J} > \tilde{J}_c^{\text{MF}}$:
 - Sous la courbe de coexistence ($\tilde{J} < \tilde{J}_{\text{coex}}^{\text{MF}}$) et bien que l'équilibre thermodynamique soit aussi un système homogène, on parle d'état basse occupation si $\Phi < \Phi_b(\tilde{J})$ et d'état haute occupation si $\Phi > \Phi_h(\tilde{J})$.

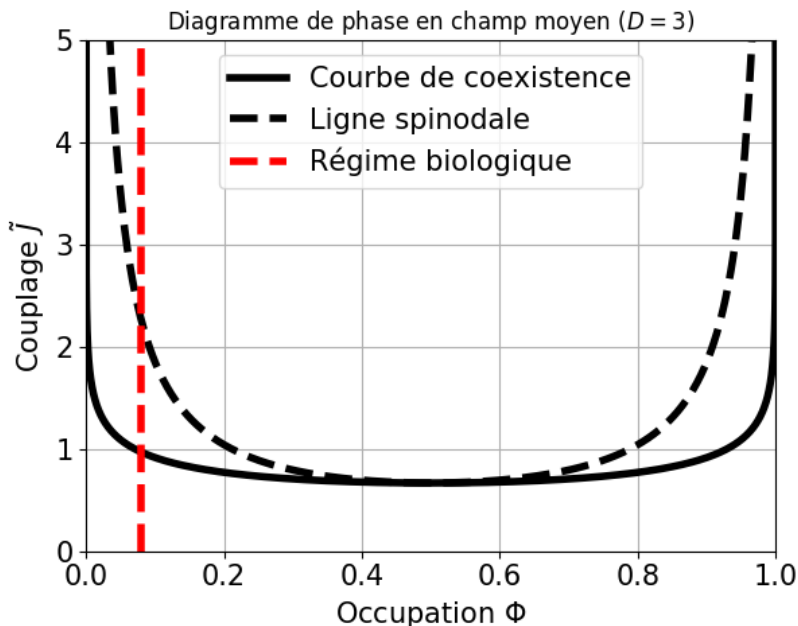


FIGURE 2.2 – **Diagramme de phase pour $D = 3$** : la courbe de coexistence est en trait plein noir et la ligne spinodale en trait discontinu noir. Ce diagramme permet de cartographier les différents régimes thermodynamiques d’un système supportant des transitions du premier ordre. La ligne rouge indique une valeur d’occupation compatible avec une occupation biologique majorée à $\Phi_{\text{bio}} = 0,08$ (voir la partie Scénario métastable).

- Au-dessus de la ligne spinodale ($\tilde{J} > \tilde{J}_{\text{spino}}^{\text{MF}}$), l’état homogène est instable thermodynamiquement et la séparation de phase a lieu spontanément. Il s’agit de la région de coexistence instable.
- Entre la ligne spinodale et la courbe de coexistence ($\tilde{J}_{\text{coex}}^{\text{MF}} < \tilde{J} < \tilde{J}_{\text{spino}}^{\text{MF}}$), le système se trouve dans la région de coexistence métastable ou métastabilité. Alors que l’état séparé est stable, l’état homogène est métastable : dans ce cas, le système peut évoluer très lentement vers l’équilibre thermodynamique. Néanmoins, un apport énergétique comme une énergie d’adsorption spécifique peut accélérer cette cinétique.

Nous illustrons ces différentes régions en produisant des profils d’occupation sur l’ADN dans la prochaine sous-section.

2.3.5 Illustration des régimes thermodynamiques

Nous sommes en mesure d’illustrer ces différents régimes thermodynamiques en utilisant une approche en champ moyen pour recalculer notre énergie libre 2.47 dans le cas inhomogène. Cette approche repose sur le principe de l’équilibre thermodynamique locale : nous allons considérer que nos sites de

2.3. GAZ SUR RÉSEAU À COURTE PORTÉE

longueur l_m correspondent à des régions à l'équilibre thermodynamique et où les quantités telles que la pression et le potentiel chimique sont localement bien définies. En adoptant un formalisme vectoriel, l'hamiltonien $\mathcal{H}_{\text{SRLG}}$ s'écrit :

$$\mathcal{H} = -\frac{1}{2}\Phi^T J \Phi - \Phi^T \Psi \quad (2.58)$$

où J est vue comme une matrice symétrique $N \times N$ dont l'élément ij vaut J si les sites i et j sont voisins sur le réseau. Les vecteurs Φ et $\psi = \mu - \epsilon$ ont pour i -ème éléments respectifs Φ_i et $\mu - \epsilon_i$. Nous écrivons la fonction de partition du gaz sur réseau inhomogène en introduisant un champ fictif $\Theta - \Psi$, d'élément $\Theta_i - \Psi_i$, sur le réseau à l'aide d'une transformation de Hubbard-Stratonovich qui permet de simplifier le calcul de la fonction de partition en intégrant sur Φ . La fonction de partition devient alors :

$$\begin{aligned} \mathcal{Z}_{\text{SRLG}} = C \int_{-\infty}^{+\infty} \prod_{i=1}^N d\Theta_i \exp \left[-\frac{\beta}{2} (\Theta - \Psi)^T J^{-1} (\Theta - \Psi) \right] \\ \times \sum_{\{\Phi_i\}} \exp(\beta \Theta^T \Phi) \end{aligned} \quad (2.59)$$

avec $C = \sqrt{\frac{\beta^N}{(2\pi)^N \det J}}$, une constante. La quantité $\det J$ dépend uniquement du réseau. En calculant la partie des N sommes, la fonction de partition peut s'écrire :

$$\mathcal{Z}_{\text{SRLG}} = C \int \prod_{i=1}^N d\Theta_i \exp[-\beta S(\Theta, \Psi)] \quad (2.60)$$

avec

$$S(\Theta, \Psi) = \frac{1}{2} (\Theta - \Psi)^T J (\Theta - \Psi) - \frac{1}{\beta} \sum_{i=1}^N \ln[1 + \exp(\beta \Theta_i)] \quad (2.61)$$

Nous utilisons une méthode de point col pour obtenir l'approximation en champ moyen de $\mathcal{Z}_{\text{SRLG}}$. Au premier ordre, on obtient effectivement :

$$\mathcal{Z}_{\text{SRLG}} = \exp(-\beta \Omega_{\text{SRLG}}) \approx \exp(-\beta S(\Theta^*, \Psi)) \quad (2.62)$$

d'où le potentiel grand canonique $\Omega_{\text{SRLG}} \approx S(\Theta^*, \Psi)$ dans le cas inhomogène avec Θ^* la valeur du paramètre Θ qui optimise l'intégrale (2.60), c'est-à-dire :

$$\left(\frac{\partial S}{\partial \Theta} \right)_{\Theta=\Theta^*} = 0. \quad (2.63)$$

Nous trouvons alors que $\Theta^* = \Psi + J \mathbf{v}$ où \mathbf{v} est le vecteur d'élément :

$$v_i = \frac{\exp(\beta \Theta_i^*)}{1 + \exp(\beta \Theta_i^*)}. \quad (2.64)$$

Nous remarquons que le vecteur d'occupation locale moyenne Φ_{ih} dans le champ moyen inhomogène est :

$$\Phi_{\text{ih}} = -\frac{\partial \Omega_{\text{SRLG}}}{\partial \Psi} \approx \frac{\partial S(\Theta^*)}{\partial \Psi} = \mathbf{v}. \quad (2.65)$$

Ainsi, le potentiel grand canonique s'écrit dans le cas inhomogène :

$$\Omega_{\text{SRLG}} \approx \frac{J}{2} \sum_{\langle i,j \rangle} \langle \Phi_i \rangle \langle \Phi_j \rangle + k_B T \sum_{i=1}^N \ln(1 - \langle \Phi_i \rangle) \quad (2.66)$$

tandis que l'énergie libre de Helmholtz inhomogène, lié au potentiel grand canonique par la transformation de Legendre $\mathcal{F}_{\text{SRLG}} = \Omega_{\text{SRLG}} + \Phi_{\text{ih}}^T \boldsymbol{\mu}$ est :

$$\begin{aligned} \mathcal{F}_{\text{SRLG}} \approx & -\frac{J}{2} \sum_{\langle i,j \rangle} \langle \Phi_i \rangle \langle \Phi_j \rangle + \sum_{i=1}^N \langle \Phi_i \rangle \epsilon_i \\ & + k_B T \sum_{i=1}^N [\langle \Phi_i \rangle \ln \langle \Phi_i \rangle + (1 - \langle \Phi_i \rangle) \ln(1 - \langle \Phi_i \rangle)] \end{aligned} \quad (2.67)$$

Le cas homogène est une limite du cas inhomogène avec $\langle \Phi_i \rangle \approx \Phi$ et $\epsilon_i = \epsilon$. L'énergie libre précédente peut être minimisée à l'aide de l'algorithme adéquat² pour un ensemble d'énergie spécifique $\{\epsilon_i\}$ afin de tracer des profils d'occupation pour $D = 1$. Ces profils nous permettent d'illustrer les différents régimes thermodynamiques du diagramme de phase des états de haute et basse occupation issu du champ moyen. Nous reproduisons une expérience de trempe pour différentes valeurs de Φ avec et sans séquence *parS* dans un espace à une dimension contenant $N = 100$. La séquence est modélisée par un $\epsilon_i = -5k_B T$ au centre du réseau. Nous choisissons $\tilde{J} = 3$ comme couplage entre premiers voisins et trois valeurs de Φ positionnant successivement le système dans

1. un régime fluide, au-dessus de la courbe de coexistence (voir figures 2.3 et 2.4) : la distribution d'équilibre est une distribution homogène sans séquence *parS* et montre un maximum très local autour du site doté d'une forte énergie d'adsorption avec une séquence *parS*;
2. un régime de coexistence métastable, sous la courbe de coexistence mais au-dessus de la ligne spinodale (voir figure 2.5 et 2.6) : la distribution d'équilibre est homogène dans le cas sans séquence *parS* tandis que nous pouvons observer une phase haute occupation localisée autour de la séquence *parS*;

2. En l'occurrence, un algorithme Python basé sur les méthodes de minimisation de fonction mathématique de la librairie `scipy.optimize`³. Nous avons opté pour la méthode `minimize` prenant en entrée une fonctionnelle dont les variables peuvent être contraintes entre deux valeurs : ici, 0 et 1 pour les occupations locales. Dans les faits, nous choisissons une valeur de potentiel chimique impliquant une valeur d'occupation moyenne sur le réseau et donc un régime thermodynamique.

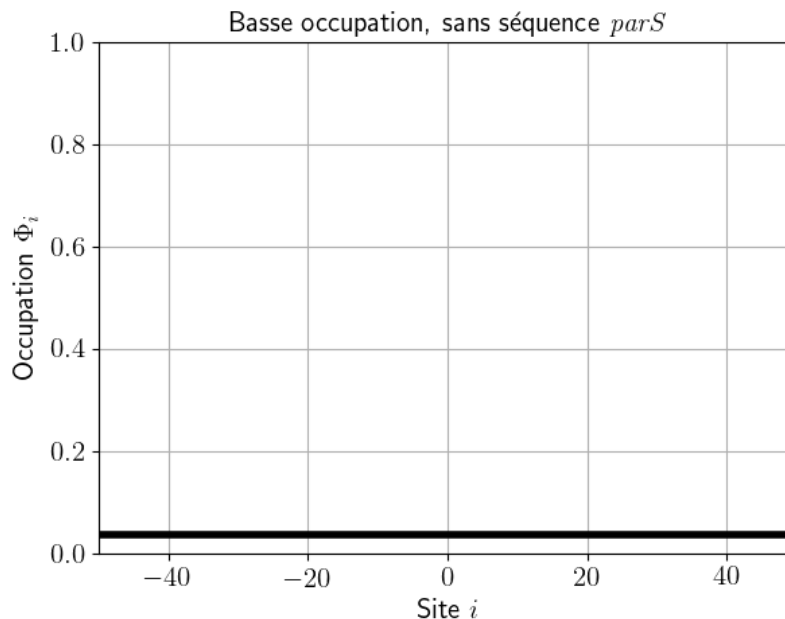


FIGURE 2.3 – Profil d’occupation dans le régime de basse occupation pour $D = 1$: sans séquence *parS*, la distribution d’équilibre est homogène en moyenne.

3. et dans un régime de coexistence instable, sous la ligne spinodale (voir figure 2.7 et 2.8) : la distribution d’équilibre montre toujours une phase haute occupation localisée autour de la séquence *parS* ou au centre du système. Ce régime aboutit à la formation spontanée de foyer.

Ces phases en haute occupation localisées représentent les complexes ParBS. Avec ces résultats préliminaires, nous sommes en mesure de déterminer le régime thermodynamique du système biologique constitué de l’ADN et des protéines ParB.

2.3.6 Scénario métastable

Le système de partition *parABS* est utilisé lors de la ségrégation du plasmide F chez *Escherichia Coli*. Le plasmide F mesure environ $L = 60 \text{ kb}$ tandis que les protéines ParB possèdent une empreinte de 16 bp sur l’ADN [22]. Le nombre de site d’adsorption disponible défini par notre approche est donc $N \approx 3750$. Il y a environ $M = 300$ protéines ParB par complexe de partition [22]. En supposant que toutes les protéines ParB se trouvent sur le plasmide F, nous obtenons une occupation moyenne $\Phi_m = M/N \approx 0,08$. Nous devons apporter deux observations à cette valeur d’occupation :

1. Lors d’un cycle cellulaire, le plasmide F est répliqué : le nombre de site disponible N est donc doublé. Les protéines ParB sont synthétisées et le

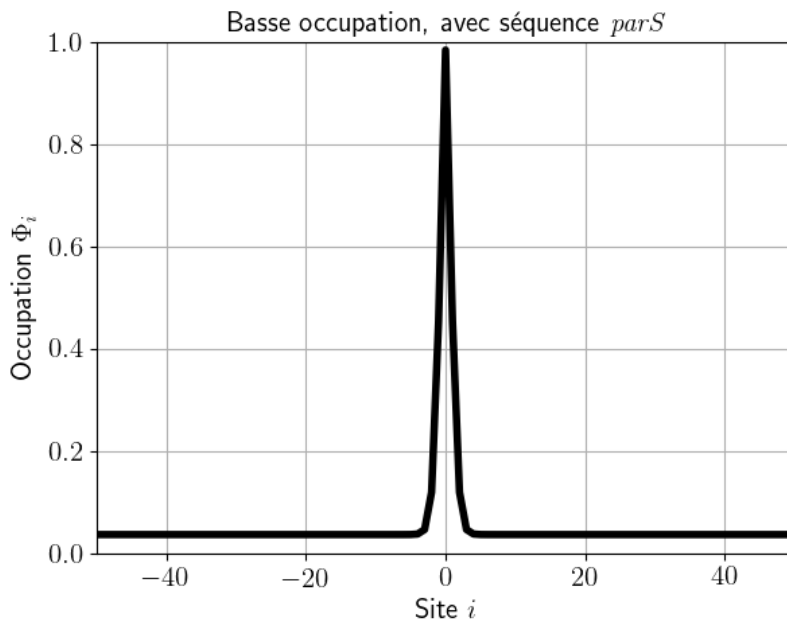


FIGURE 2.4 – **Profil d'occupation dans le régime de basse occupation pour $D = 1$** : avec une séquence *parS* au centre du réseau, la distribution des particules est quasiment homogène si ce n'est un maximum local autour de la séquence *parS*.

nombre de protéines M doublent aussi. Nous pouvons donc considérer Φ_{bio} comme constante pendant le processus de réplication.

2. Le chromosome principal joue le rôle d'un ADN compétiteur : le chromosome principal d'*E. Coli*, de longueur $L_{\text{cp}} = 4,6$ Mb bien supérieur à celle du plasmide F, doit priver le plasmide d'un nombre de protéines, difficile à estimer. En effet, si nous pouvons qualifier le *bridging* d'interaction plutôt neutre⁴ puisque deux protéines peuvent former un lien via ce type d'interaction qu'elles soient sur l'ADN ou un plasmide, ce n'est pas le cas des interactions de *spreading* qui doivent avoir lieu entre deux protéines voisines le long de la chaîne. Ce type d'interaction possède une longueur de corrélation. Les séquences *parS* faisant partie du plasmide F, il y a un biais en faveur de l'occupation sur ce plasmide.

Ainsi, nous considérons que Φ_{m} est au pire des cas une bonne majoration de l'occupation des protéines ParB sur le plasmide F. Nous connaissons les coordonnées ($\Phi_{\text{bio}}, T_{\text{bio}} = 300$ K) du système biologique. Ces coordonnées définissent les coordonnées du *régime biologique* relatif au système ParABS : il est représenté par une ligne rouge dans le diagramme de phase en figure 2.2.

4. En toute rigueur, l'interaction de *bridging* n'est pas totalement neutre puisque deux sites en interaction via un pont de *bridging* subissent des contraintes fortes dues à la chaîne elle-même. Il y a une distance minimale entre deux sites le long de la chaîne telles que ces deux sites peuvent interagir (voir la partie *Spreading* et *bridging*).

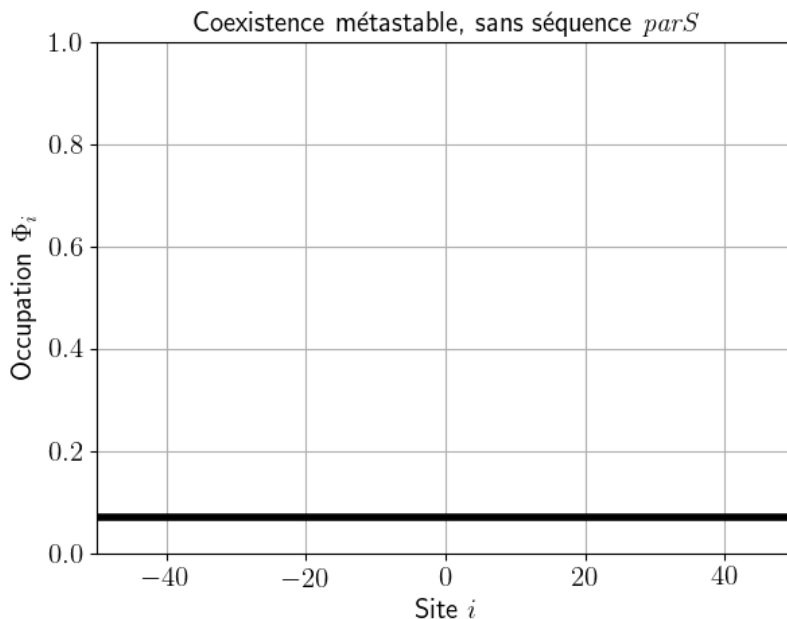


FIGURE 2.5 – **Profil d'occupation dans le régime de coexistence métastable pour $D = 1$:** sans séquence *parS*, la distribution d'équilibre est homogène en moyenne.

Pour des valeurs raisonnables de J de quelques $k_B T$, le régime biologique correspond au régime de coexistence métastable en basse occupation. Ce régime offre l'opportunité d'expliquer les résultats des expériences de SPT-PALM (voir Introduction). Sans séquence *parS*, les expériences montrent une distribution homogène en protéines ParB sur le nucléoïde tandis qu'avec des séquences *parS*, ces protéines forment des complexes ParBS : c'est une propriété bien connue de la coexistence métastable (voir sous-section précédente et figure 2.4). Dans ce que nous qualifions de *scénario métastable*, la séquence *parS* apporte l'énergie nécessaire pour dépasser la barrière de nucléation et permettre au système de transiter de la distribution homogène métastable à la coexistence stable de phase pour laquelle les foyers en haute occupation/densité sont localisés autour des défauts/séquence *parS*.

Cette phénoménologie de séparation de phase et ce scénario métastable possède une conséquence immédiate : le système biologique doit avoir le même comportement que les transitions de type liquide-vapeur. Notamment, la sur-expression ou sous-expression des protéines ParB doit favoriser ou inhiber la formation de complexes ParBS selon la position dans le diagramme de phase. En effet, un système sans séquence *parS* mais avec une concentration/occupation en ParB suffisamment élevée, due à la sur-expression, sera dans le régime de coexistence instable et pourra former des foyers de ParB spontanément. Au contraire, un système avec séquence *parS* mais trop peu de protéines ParB traverse la courbe de coexistence pour n'exister que dans un

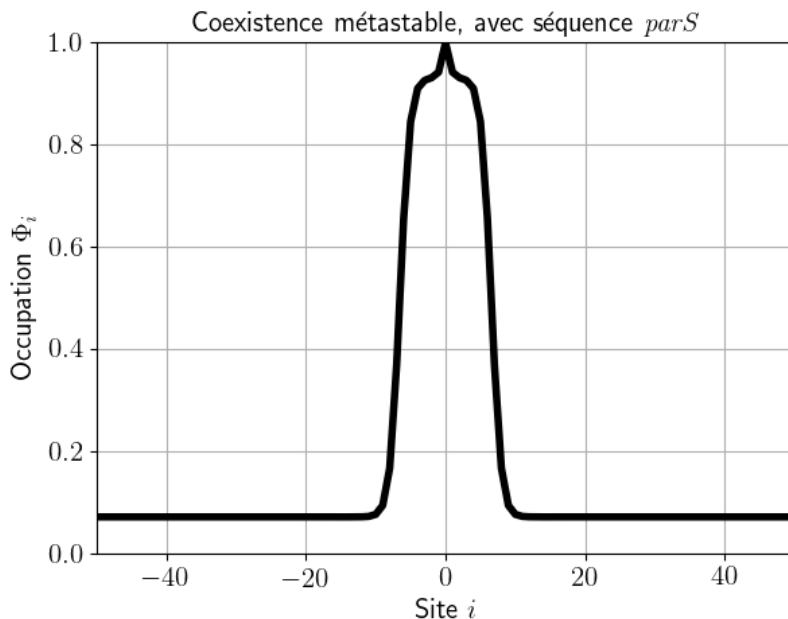


FIGURE 2.6 – **Profil d’occupation dans le régime de coexistence métastable pour $D = 1$** : avec une séquence *parS* au centre du réseau, cette distribution montre un plateau en haute occupation autour de la séquence *parS*.

simple régime de basse occupation : un tel système perd son abilité à former des complexes ParBS.

En terme d’organisation intracellulaire, le régime de coexistence métastable présente l’avantage de former des foyers à haute densité localisés avec peu de protéines, permettant d’éviter des effets d’encombrement. Il permet également de contrôler cette formation (séquence *parS*). En l’état, ce scénario semble crédible mais peu étayé à cause de la simplicité extrême du modèle de gaz sur réseau. L’influence de la conformation de l’ADN est absente tandis que le détail des interactions entre protéines est pauvrement décrit.

Dans la partie suivante, nous examinons la validité de notre approche en champ moyen. Nous montrons aussi que cette technique n’est pas sans intérêt dans les limites de basse et haute occupations.

2.3.7 Limite et succès du champ moyen

Bien que l’approximation du champ moyen permettent d’obtenir des résultats analytiques simples jusqu’à l’expression de la ligne de coexistence et spinodale, elle présente certains défauts. Elle prédit des transitions de phases quelle que soit la dimension $D = q/2$ du système, notamment pour $D = 1$. Or, le modèle d’Ising a une dimension ne présente pas de transition de phase : ce résultat, transposable au gaz sur réseau, est connu sous le nom de théorème de

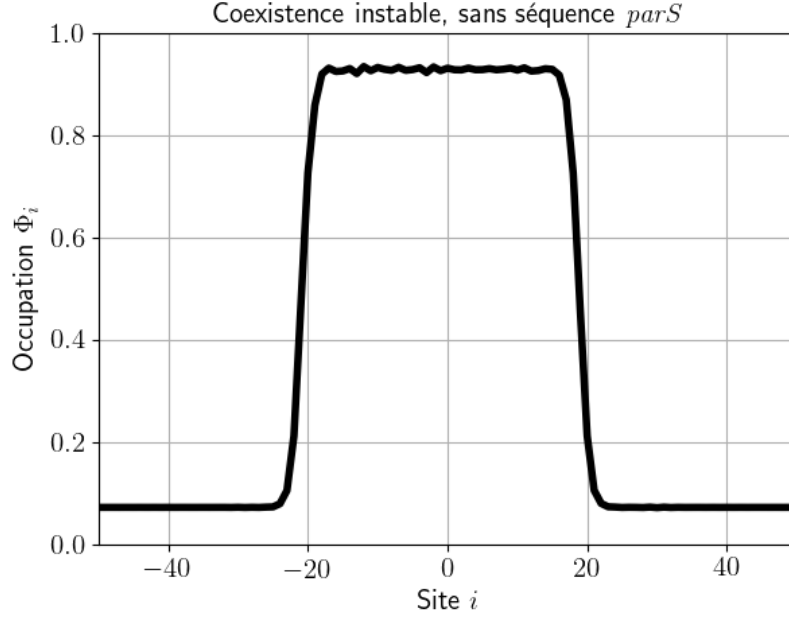


FIGURE 2.7 – **Profil d’occupation dans le régime de coexistence instable pour $D = 1$:** sans séquence *parS*, la distribution indique un plateau de haute occupation.

Landau. A une dimension, le champ moyen est ainsi fondamentalement faux et le gaz sur réseau possède une dimension critique inférieure $D = D_{\text{inf}} = 1$. Le cas $D = 1$ ne doit donc être utilisé que pour illustrer très généralement la phénoménologie des transitions de phase du premier ordre. C’est à Lars Onsager que l’on doit la résolution exacte du modèle d’Ising à deux dimensions sans champ magnétique ($h = 0$) [92]. Il montre que ce modèle présente des transitions de phase avec une température critique T_c^I dans le cas $D = 2$ donnée par :

$$\frac{k_B T_c^I}{J_I} = \frac{2}{\ln(1 + \sqrt{2})} \approx 2,269. \quad (2.68)$$

En utilisant les relations de liaison avec le modèle d’Ising, nous trouvons l’expression du couplage critique du gaz sur réseau 2D :

$$\tilde{J}_c^{2D} = 2 \ln(1 + \sqrt{2}) \approx 1,764 \left(\frac{k_B T_c^{2D}}{J} = \frac{1}{2 \ln(1 + \sqrt{2})} \approx 0,567 \right). \quad (2.69)$$

Les résultats exacts du cas général $h \neq 0$ n’ont pas encore été découverts à ce jour. Le même constat s’impose pour le modèle d’Ising à trois dimensions mais il existe néanmoins des résultats numériques pour la température critique à $D = 3$ [93] :

$$\frac{k_B T_c^I}{J_I} \approx 4,511, \quad (2.70)$$

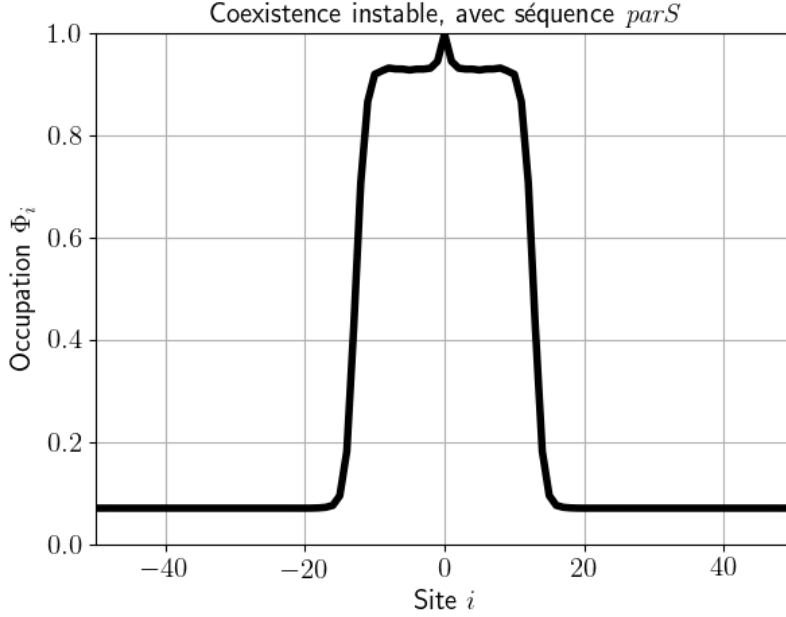


FIGURE 2.8 – **Profil d’occupation dans le régime de coexistence instable pour $D = 1$** : avec une séquence *parS* au centre du réseau, nous observons également un plateau de haute occupation au centre du réseau.

ce qui donne en langage gaz sur réseau :

$$\tilde{J}_c^{3D} \approx 0,887 \left(\frac{k_B T_c^{3D}}{J} \approx 1,128 \right). \quad (2.71)$$

La relation (2.52) nous permet de calculer le rapport entre les valeurs du couplage critique exact et en champ moyen à 2D :

$$\theta_{2D} = \frac{\tilde{J}_c^{MF}}{\tilde{J}_c^{2D}} = \frac{T_c^{2D}}{T_c^{MF}} \approx 0,567 \quad (2.72)$$

qui correspond à une erreur d’environ 43%, et à 3D :

$$\theta_{3D} = \frac{\tilde{J}_c^{MF}}{\tilde{J}_c^{3D}} = \frac{T_c^{3D}}{T_c^{MF}} \approx 0,752 \quad (2.73)$$

soit une erreur de 25%. Le champ moyen néglige les corrélations entre sites. Elle sous-estime ainsi les fluctuations thermiques et surestime la température critique. L’erreur due à cette approximation diminue avec la dimension. En effet, le champ moyen est efficace si la variation d’un Φ_i est dominée par sa moyenne Φ : c’est le critère de Ginzburg [88]. Or, cette différence entre écart-type et moyenne est fonction de la dimension D . On peut montrer que la prédiction champ moyen de la température critique devient exacte à partir de

$D = D_{\text{sup}} = 4$, qui définit ainsi une dimension critique supérieure.

Le système biologique constitué des protéines et de l'ADN devrait exister *a priori* à $D = 3$ ou à une dimension fractale effective comprise entre 2 et 3, si nous prenons en compte que les protéines existent sur une chaîne fluctuante. Pour ces valeurs de D , nous constatons que le champ moyen ne semble pas suffisant pour prédire correctement la température critique. Néanmoins, nous avons vu que le système biologique ne semble pas exister dans un régime proche de la température critique, mais plutôt dans un régime à basse occupation, loin du point critique. Il s'avère que le champ moyen est exact pour la courbe de coexistence des occupations très faibles. Pour le démontrer, il suffit de développer à basse et haute occupations la fonction de partition $\mathcal{Z}_{\text{SRLG}}$ et de déterminer à nouveau la courbe de coexistence et la ligne spinodale dans cette limite. En basse occupation et au premier ordre du développement, il suffit de prendre en compte deux configurations : la première où le réseau est vide et d'énergie nulle, la deuxième où il n'y a qu'une seule particule occupant un des N sites conférant une énergie $-\mu$ au réseau. Nous avons alors :

$$\mathcal{Z}_{\text{SRLG}}^{0+} \approx 1 + N \exp(\beta\mu) \quad (2.74)$$

Le potentiel grand canonique correspondant est :

$$\Omega_{\text{SRLG}}^{0+} = -k_B T \ln(1 + N \exp(\beta\mu)) \approx -Nk_B T \exp(\tilde{\mu}) \quad (2.75)$$

car, en basse occupation, μ diverge vers $-\infty$. En inversant la précédente équation et sachant que le potentiel grand canonique vérifie $\Omega_{\text{SRLG}} = -PV$:

$$\tilde{\mu}^{0+}(\Phi, T) \approx \ln \Phi \quad (2.76)$$

$$\tilde{P}^{0+}(\Phi, T) \approx \Phi \quad (2.77)$$

Dans la limite haute occupation, la fonction de partition est composée d'un premier terme d'énergie correspondant à un réseau complètement occupé ($\frac{Nq}{2}$ liaisons distinctes entre protéines pour N sites occupés) et un second terme correspondant à un réseau plein excepté un site (nous brisons q liaisons et nous avons N possibilités pour placer ce trou) :

$$\begin{aligned} \mathcal{Z}_{\text{SRLG}}^{1-} &= \exp \left[\beta \left(\frac{JNq}{2} + N\mu \right) \right] \\ &+ N \exp \left[\beta \left(\frac{JNq}{2} + N\mu - Jq - \mu \right) \right] + \dots \end{aligned} \quad (2.78)$$

Le potentiel grand canonique est alors :

$$\begin{aligned} \Omega_{\text{SRLG}}^{1-} &= -k_B T \left\{ \frac{\tilde{J}Nq}{2} + N\tilde{\mu} + \ln \left[1 + N \exp \left(-\tilde{J}q - \tilde{\mu} \right) \right] \right\} \\ &\approx -Nk_B T \left[\frac{\tilde{J}q}{2} + \tilde{\mu} + \exp \left(-\tilde{J}q - \tilde{\mu} \right) \right] \end{aligned} \quad (2.79)$$

Le potentiel chimique et la pression se calculent comme précédemment :

$$\tilde{\mu}^{1-}(\Phi, T) \approx -\tilde{J}q - \ln(1 - \Phi) \quad (2.80)$$

$$\tilde{P}^{1-}(\Phi, T) \approx -\frac{\tilde{J}q}{2} - \ln(1 - \Phi) + 1 - \Phi \quad (2.81)$$

En régime de coexistence, les densités Φ_b et Φ_h vérifient les égalités du potentiel chimique et de la pression :

$$\begin{cases} \tilde{\mu}^{0+}(\Phi_b, T) = \tilde{\mu}^{1-}(\Phi_h, T) \\ \tilde{P}^{0+}(\Phi_b, T) = \tilde{P}^{1-}(\Phi_h, T) \end{cases} \quad (2.82)$$

soit en développant :

$$\begin{cases} \tilde{J}q + \ln[\Phi_b(1 - \Phi_h)] = 0 \\ \Phi_b + \rho_l - 1 + \frac{\tilde{J}q}{2} + \ln(1 - \Phi_h) = 0 \end{cases} \quad (2.83)$$

En exprimant Φ_l en fonction de Φ_v avec la symétrie particule-trou (2.55), les deux équations deviennent, comme dans l'approche en champ moyen, identiques. Il ne reste qu'une seule équation :

$$\tilde{J}q + 2 \ln \Phi_b = 0 \quad (2.84)$$

qui nous permet d'exprimer $\tilde{J}_{\text{coex}}^{0+}$ en fonction de Φ dans la limite de basse occupation :

$$\tilde{J}_{\text{coex}}^{0+} \approx \frac{2}{q} \ln \Phi \left(T_{\text{coex}}^{0+} \approx -\frac{Jq}{2k_B \ln \Phi} \right) \quad (2.85)$$

Le développement à basse occupation de l'expression du couplage de transition en champ moyen (2.56) donne le même résultat. Pour l'expression de la ligne spinodale, il faut prolonger le développement de $\mathcal{Z}_{\text{SRLG}}^{0+}$ au deuxième ordre en $\exp(\tilde{\mu})$. Nous avons alors :

$$\begin{aligned} \mathcal{Z}_{\text{SRLG}}^{0+} &= 1 + N \exp(\tilde{\mu}) + \frac{Nq}{2} \exp(2\tilde{\mu} + \tilde{J}) \\ &\quad + \frac{N \times (N - 1 - q)}{2} \exp(2\tilde{\mu}) + \dots \dots \end{aligned} \quad (2.86)$$

Le troisième terme du développement correspond à la formation d'un dimère de ParB (avec $\frac{Nq}{2}$ possibilités distinctes), le quatrième à deux particules ParB qui n'interagissent pas ($\frac{N \times (N - 1 - q)}{2}$ possibilités distinctes restantes). Le grand potentiel s'écrit :

$$\begin{aligned} \Omega_{\text{SRLG}}^{0+} &= -k_B T \ln \mathcal{Z}_{\text{SRLG}}^{0+} \\ &\approx Nk_B T \left[\exp \tilde{\mu} + \frac{1}{2} \left(q \exp \tilde{J} - q - 1 \right) \exp 2\tilde{\mu} \right] \end{aligned} \quad (2.87)$$

2.3. GAZ SUR RÉSEAU À COURTE PORTÉE

en développant le logarithme au deuxième ordre en $\exp(\tilde{\mu})$. L'annulation de la dérivée première de $\Omega_{\text{SRLG}}^{0+}$ par rapport à Φ donne l'expression de la ligne spinodale dans la limite basse occupation :

$$\tilde{J}_{\text{spino}}^{0+} \approx \ln \left(1 + \frac{1 + \Phi}{q\Phi} \right) \quad \left(T_{\text{spino}} \approx \frac{J}{k_B \ln \left(1 + \frac{1+\rho}{q\rho} \right)} \right) \quad (2.88)$$

Nous retrouvons l'expression de la limite basse occupation du couplage de la ligne spinodale (2.57) lorsque q tend vers $+\infty$, qui est la limite attendue pour le champ moyen. La courbe de coexistence et la ligne spinodale du développement en basse et haute densités et du champ moyen sont tracées sur la figure 2.9. Nous montrons ainsi que le champ moyen donne un résultat exact pour la

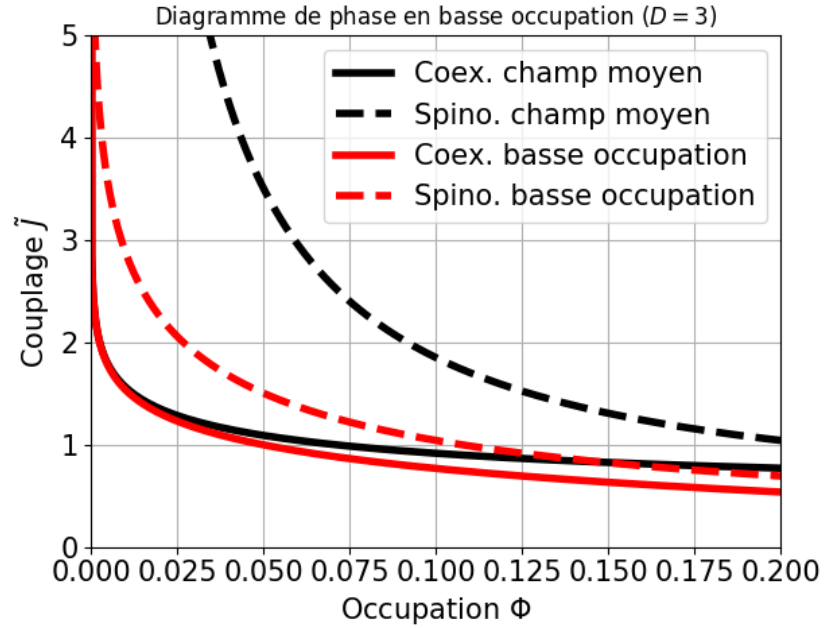


FIGURE 2.9 – **Diagramme de phase pour $D = 3$ en basse occupation** : courbes de coexistence en traits pleins et lignes spinodales en lignes discontinues. Noir pour les résultats dans l'approximation du champ moyen et rouge pour les résultats du développement en basse occupation. La courbe de coexistence du champ moyen est assez juste en basse occupation tandis que la ligne spinodale est corrigée.

courbe de coexistence dans la limite à basse occupation, qui est précisément le régime biologique ($\Phi_{\text{bio}} < 0,08$). L'expression de la ligne spinodale connaît néanmoins des corrections. Ces données tendent à renforcer notre hypothèse de métastabilité des protéines ParB sur l'ADN.

2.4 Polymère à 3D interagissant avec un bain de protéines

Dans cette nouvelle section, nous proposons une approche plus rigoureuse pour valider notre scénario métastable pour les protéines ParB. Nous prenons en compte de manière plus précise la distribution de l'ADN dans l'espace intracellulaire. Cette distribution s'avère capitale pour montrer que les complexes ParBS sont le résultat d'une séparation de phase. Enfin, nous parvenons à obtenir le diagramme de phase de cette séparation pour des polymères gaussiens et auto-évitants.

2.4.1 *Spreading et bridging*

Nous adoptons un modèle à gros grains, discret, de polymère (voir figure 2.10). La séquence d'ADN consiste en un polymère de N monomères

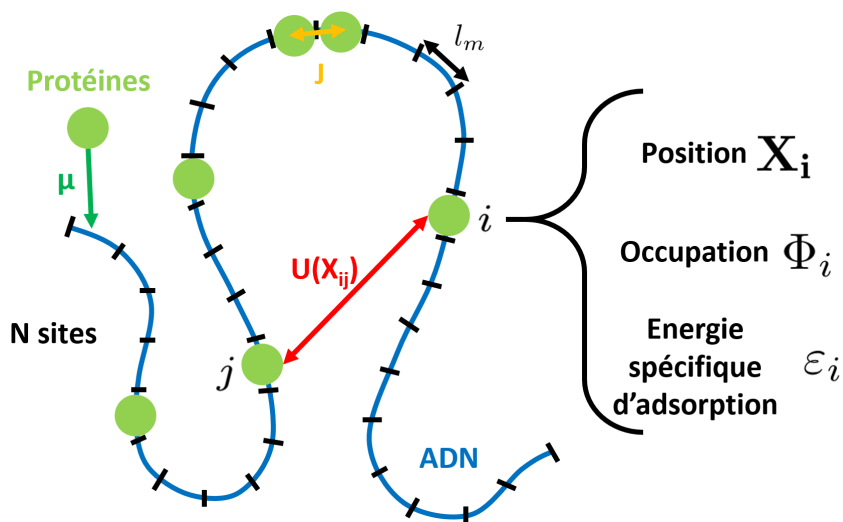


FIGURE 2.10 – Schéma du modèle de *spreading* et *bridging* : la molécule d'ADN est divisée en N sites. Chaque monomère ou site, de longueur l_m , possède trois attributs : sa position \mathbf{X}_i , son occupation Φ_i et une énergie spécifique d'adsorption ε_i . Le nombre de protéines sur ce réseau d'ADN est contrôlé par le potentiel chimique μ . Les protéines, une fois adsorbée, peuvent interagir via les interactions de *spreading*, d'intensité J , et de *bridging*, contrôlée par le potentiel $U(\mathbf{X}_{ij})$.

sans contrainte dans un volume V . Chaque monomère est capable d'adsorber une seule protéine et la longueur effective du monomère l_m (en nombre de paire de base, bp, par exemple) correspond à l'empreinte d'une protéine ParB sur l'ADN, soit $l_m = 16pb$. L'adsorption des protéines sur ce réseau unidimensionnelle est modélisée par un modèle de gaz sur réseau. Chaque site du réseau possède ainsi trois attributs :

2.4. POLYMÈRE À 3D INTERAGISSANT AVEC UN BAIN DE PROTÉINES

1. sa position, donnée par le vecteur \mathbf{X}_i dans l'espace à trois dimensions ;
2. une occupation locale Φ_i qui vaut 0 si une protéine est présente sur ce site, et 1 sinon ;
3. une énergie spécifique d'adsorption ε_i dont le signe est choisi tel que $\varepsilon_i < 0$ favorise l'adsorption sur le site i .

La vecteur bout-à-bout est simplement $\mathbf{R}_{\text{bb}} = \mathbf{X}_N - \mathbf{X}_1$ et la distance entre deux monomères i et j est $\mathbf{X}_{ij} = \mathbf{X}_j - \mathbf{X}_i$. Nous supposons que les protéines liées sont en équilibre avec un réservoir considéré infini et uniforme de protéines de température T et gouverné par un potentiel chimique μ . L'énergie d'un état microscopique $\{\Phi_i, \mathbf{X}_i\}$ est donnée par l'hamiltonien :

$$\mathcal{H}(\{\Phi_i, \mathbf{X}_i\}) = \mathcal{H}_P(\{\mathbf{X}_i\}) + \mathcal{H}_{\text{SRLG}}(\{\Phi_i\}) + \mathcal{H}_B(\{\Phi_i, \mathbf{X}_i\}). \quad (2.89)$$

L'hamiltonien \mathcal{H} est composé de trois termes. Le premier terme \mathcal{H}_P correspond à la contribution de l'ADN-polymère semi-flexible. Il ne dépend en conséquence que de l'ensemble des $\{\mathbf{X}_i\}$. Aucune hypothèse supplémentaire n'est faite sur la statistique de la chaîne : nous y reviendrons par la suite. Le deuxième terme ne dépend que des $\{\Phi_i\}$: il correspond à un gaz sur réseau à courte portée à une dimension et compte les interactions entre premiers voisins d'intensité J . Il s'écrit :

$$\mathcal{H}_{\text{SRLG}}[\Phi_i] = -J \sum_{i=1}^{N-1} \Phi_{i+1} \Phi_i - \sum_{i=1}^N (\mu - \varepsilon_i) \Phi_i. \quad (2.90)$$

La chaîne d'ADN constitue notre espace à une dimension. L'interaction à courte portée correspond à l'interaction de *spreading* entre protéines voisines le long de l'ADN. La dernière contribution s'écrit :

$$\mathcal{H}_B[\Phi_i, \mathbf{X}_i] = \frac{1}{2} \sum'_{i,j} \Phi_i U(X_{ij}) \Phi_j, \quad (2.91)$$

et donne le couplage entre les protéines adsorbées sur le réseau 1D du polymère et le polymère fluctuant dans l'espace à trois dimensions. Ce terme compte les interactions de *bridging* entre protéines liées voisines dans l'espace à trois dimensions, et non le long de la chaîne à une dimension, via le potentiel d'interaction U . Ce potentiel pourra être choisi répulsif à courte portée et attractif à portée intermédiaire, comme un potentiel de Lennard-Jones ou à puits carré. On gardera néanmoins une courte portée pour rester dans l'esprit d'une interaction entre premiers voisins. Le symbole $\sum'_{i,j}$ signifie que seuls les termes vérifiant $|i - j| \geq n_{\text{inf}}$ sont pris en compte dans la somme, où n_{inf} est la distance minimale, le long de la chaîne et en nombre de site, à partir de laquelle deux particules peuvent interagir à longue portée. Le paramètre n_{inf} correspond en fait à la longueur de persistance du polymère en nombre de site puisque les boucles plus petites que l_p sont inhibées par la rigidité du

polymère.

Nous définissons ainsi un cadre de travail très général pour des particules interagissant avec un polymère ou, dans notre cas, des protéines adsorbables sur l'ADN. Une approche très similaire a été définie [55] à la différence que le polymère fluctue dans un espace continue au lieu d'exister sur un réseau dans notre description. Dans cette même référence, ce modèle a fait l'objet d'une étude numérique extensive. Ces travaux sur des chaînes de taille finie n'apportent en revanche pas une vision complète du diagramme de phase. Nous souhaitons d'abord apporter une preuve théorique rigoureuse qu'une telle configuration peut aboutir à une vraie transition de phase thermodynamique, c'est-à-dire ici des séparations de phase entre états de haute et basse occupation en protéines sur l'ADN et ainsi à la formation de complexes de protéines, puis obtenir le diagramme de phase d'une telle séparation.

2.4.2 Simplification du *spreading* et *bridging*

La fonction de partition associée au *spreading* et *bridging*

$$\mathcal{Z} = \sum_{\{\Phi_i=0,1\}} \int_V \prod_{i=1}^N d\mathbf{X}_i e^{-\beta\mathcal{H}}. \quad (2.92)$$

est difficilement calculable analytiquement. Nous pouvons néanmoins calculer cette fonction de partition dans une approximation de basse densité en monomères à trois dimensions. L'expression précédente peut en effet s'écrire :

$$\mathcal{Z} = \mathcal{Z}_P \sum_{\{\Phi_i=0,1\}} e^{-\beta\mathcal{H}_{\text{SRLG}}} \langle e^{-\beta\mathcal{H}_B} \rangle_P, \quad (2.93)$$

avec

$$\mathcal{Z}_P = \int_V \prod_{i=1}^N d\mathbf{X}_i e^{-\beta\mathcal{H}_P} \quad (2.94)$$

la fonction de partition liée à la statistique du polymère, et

$$\langle e^{-\beta\mathcal{H}_B} \rangle_P = \mathcal{Z}_P^{-1} \prod_{i=1}^N \int_V d\mathbf{X}_i e^{-\frac{\beta}{2} \sum_{i,j}^{N'} \Phi_i U(X_{ij}) \Phi_j} e^{-\beta\mathcal{H}_P}. \quad (2.95)$$

la moyenne de \mathcal{H}_C dans l'ensemble statistique définie par \mathcal{H}_P . Ce terme peut être simplifié en utilisant un développement en clusters (*cluster expansion*) [91,

2.4. POLYMÈRE À 3D INTERAGISSANT AVEC UN BAIN DE PROTÉINES

94]. On peut en effet écrire :

$$\begin{aligned}
\langle e^{-\beta\mathcal{H}_B} \rangle_P &= \frac{1}{\mathcal{Z}} \int_V d\mathbf{X}_i \left(\prod'_{i,j}^N 1 + \Phi_i \Phi_j f_{ij} \right) \\
&= \frac{1}{\mathcal{Z}} \int_V d\mathbf{X}_i \left(1 + \sum'_{i,j}^N \Phi_i \Phi_j f_{ij} \right. \\
&\quad \left. + \sum'_{i,j,k,p}^N \Phi_i \Phi_j \Phi_k \Phi_p f_{ij} f_{kp} + \dots \right) \quad (2.96)
\end{aligned}$$

où $f_{ij} = e^{-\beta U(X_{ij})} - 1$ est la fonction de Mayer relative aux sites i et j . Les occupations locale Φ_i et Φ_j peuvent être extraites de l'argument de l'exponentielle car elles ne peuvent valoir que 0 ou 1. Cette nouvelle écriture permet de développer $\langle e^{-\beta\mathcal{H}_B} \rangle_P$ en série de termes faisant intervenir le même nombre de particules. L'ordre 0, égal à 1, correspond à un système sans interaction de *bridging*, tandis que le premier ordre ne compte que les interactions à deux corps et le deuxième à quatre particules. Il s'agit en réalité d'une généralisation du développement du viriel. En ne retenant que le premier ordre, nous avons donc :

$$\langle e^{-\beta\mathcal{H}_B} \rangle_P \approx 1 + \sum'_{i,j}^N \Phi_i \langle f_{ij} \rangle_P \Phi_j \quad (2.97)$$

avec par définition

$$\langle f_{ij} \rangle_P = \int_V d\mathbf{R} P_{ij}(\mathbf{R}) (e^{-\beta U(\mathbf{R})} - 1) \quad (2.98)$$

où $P_{ij}(\mathbf{R}) = \langle \delta(\mathbf{R} - \mathbf{X}_{ij}) \rangle_P$ est la fonction de distribution du polymère telle que :

$$\int_V d\mathbf{R} P_{ij}(\mathbf{R}) = 1 \quad (2.99)$$

Comme nous travaillons en régime de basse densité, il s'agit de la distribution du polymère à vide, c'est-à-dire sans particule adsorbée. Nous obtenons finalement une approximation pour la fonction de partition (2.93) :

$$\mathcal{Z} \approx \mathcal{Z}_P \times \mathcal{Z}_{\text{LRLG}} \quad (2.100)$$

où $\mathcal{Z}_{\text{LRLG}} = \sum_{\{\Phi_i=0,1\}} \exp(-\beta\mathcal{H}_{\text{LRLG}})$. Ce développement en clusters permet de définir un nouvel hamiltonien $\mathcal{H}_{\text{LRLG}}$ qui s'écrit :

$$\mathcal{H}_{\text{LRLG}} = \mathcal{H}_{\text{SRLG}} - \frac{1}{2} \sum'_{i,j}^N \Phi_i G_{ij} \Phi_j. \quad (2.101)$$

2.4. POLYMÈRE À 3D INTERAGISSANT AVEC UN BAIN DE PROTÉINES

Nous définissons ainsi un nouveau modèle de gaz sur réseau avec des interactions à courte portée, comptées dans le terme déjà connu $\mathcal{H}_{\text{SRLG}}$, et des interactions effectives à longue portée comptées dans le second terme de l'expression précédente (*Long Range Lattice Gas* ou LRLG) (voir figure 2.11). Cette défi-

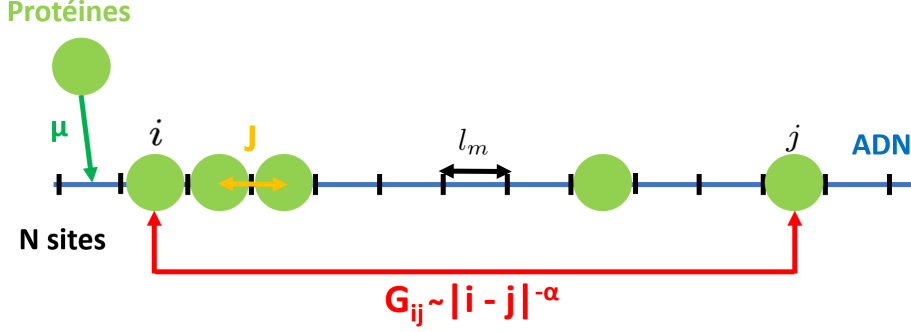


FIGURE 2.11 – Schéma du modèle de gaz sur réseau avec interactions à courte et longue portée : la molécule d'ADN, considérée comme un espace unidimensionnel, est divisée en N sites. Chaque monomère ou site, de longueur l_m , possède deux attributs : son occupation Φ_i et une énergie spécifique d'adsorption ε_i . Le nombre de protéines sur ce réseau d'ADN est contrôlé par le potentiel chimique μ . Les protéines, une fois adsorbée, peuvent interagir via les interactions de *spreading* à courte portée, d'intensité J , et des interactions effectives à longue portée, contrôlée par le noyau G_{ij} , qui dépendent de la distance linéaire entre deux sites.

inition mène à un noyau d'interactions à longue portée G_{ij} qui s'écrit :

$$G_{ij} = k_B T \langle f_{ij} \rangle_P = 4\pi k_B T \int_0^{+\infty} dR R^2 (e^{-\beta U(R)} - 1) P_{ij}(R) \quad (2.102)$$

pour des interactions isotropes. Le noyau G_{ij} prend en compte la connectivité de la chaîne grâce à la fonction de distribution du polymère P_{ij} . Ce noyau possède deux limites intéressantes.

1. Si nous brisons les liens entre les monomères du polymère, en considérant que ces monomères soient néanmoins tous occupés par des protéines ParB et que $J = 0$, nous obtenons un gaz réel de monomères. Dans notre approche, P_{ij} devient simplement $1/V$ et $G_{ij} \rightarrow B(T)$ avec :

$$B(T) = \int_0^{+\infty} dR R^2 (e^{-\beta U(R)} - 1) \quad (2.103)$$

qui est le coefficient de second ordre du développement du viriel appliqué à un fluide réel.

2. Dans les approches de type Flory [95, 96, 97], le terme d'énergie libre prenant en compte les interactions entre ligands sur un polymère fait intervenir un coefficient de second ordre du développement du viriel dans

2.4. POLYMÈRE À 3D INTERAGISSANT AVEC UN BAIN DE PROTÉINES

lequel la connectivité de la chaîne est simplement donnée par le volume effectif occupé par le polymère $P_{ij} \sim R_{\text{bb}}^{-3}$. Dans ce cas, le noyau d'interaction ne dépend plus de la position relative des deux ligands qui interagissent : il s'agit d'un modèle avec des interactions à portée infinie. Or, la limite thermodynamique n'est pas définie dans ce cas puisque ce noyau n'est pas normalisé par le nombre de monomères (voir Intermède sur le modèle de gaz sur réseau avec interactions à portée infinie).

Le noyau G_{ij} dépend aussi du choix de potentiel d'interaction entre protéines $U(R)$. Lorsque $U = 0$, le noyau G_{ij} devient nul et \mathcal{Z} est la fonction de partition d'un système d'un polymère et de protéines découplés.

Physiquement, le *spreading* et *bridging* et ce nouveau modèle de gaz sur réseau à longue portée présentent des différences importantes. Le modèle de *spreading* et *bridging* est un modèle de polymère fluctuant dans l'espace à trois dimensions ; des protéines peuvent se lier au polymère et interagir via deux types d'interactions. Notre simplification aboutit à une réduction en dimensionnalité puisque nous ne considérons que l'espace unidimensionnel du polymère dans notre gaz sur réseau. Les interactions de *bridging* entre protéines voisines dans l'espace sont réinterprétées comme des interactions effectives à longue distance sur la chaîne. Soulignons à nouveau que, pour tenir compte de la longueur de persistance du polymère, seul les termes tels que $|i - j| < n_{\text{inf}}$ sont pris en compte pour les interactions à longue portée. Ainsi, les premières valeurs du noyau G_{ij} , les plus importantes, telles que $|i - j|$ est faible, n'interviennent jamais dans l'hamiltonien $\mathcal{H}_{\text{LRLG}}$.

2.4.3 Critère d'existence pour les séparations de phase sur l'ADN

La propriété du modèle LRLG à présenter un comportement de transition de phase repose entièrement sur le terme des interactions à longue portée dans l'hamiltonien $\mathcal{H}_{\text{LRLG}}$. En effet, dans la limite $G_{ij} \rightarrow 0$, $\mathcal{H}_{\text{LRLG}}$ tend vers $\mathcal{H}_{\text{SRLG}}$. Or, nous avons montré que des interactions à courte portée dans un espace unidimensionnel sont insuffisantes pour provoquer la séparation de phase des espèces qui existent sur cette espace. Nous pouvons démontrer un critère d'existence des transitions ou des séparations de phase pour le modèle LRLG. Il entre en effet dans la classe d'universalité des modèles d'Ising avec interactions à longue portée. Ainsi, c'est le comportement asymptotique ($|i - j|$ grand) du noyau G_{ij} qui dicte l'existence des transitions de phase dans notre modèle. Dans notre cas, ce comportement asymptotique peut être déterminé analytiquement. La fonction de distribution du polymère à trois dimensions évolue à longue distance avec la loi d'échelle [97] :

$$P_{ij}(R) = c_0 \frac{R^g}{R_{ij}^{3+g}}, \quad (2.104)$$

2.4. POLYMÈRE À 3D INTERAGISSANT AVEC UN BAIN DE PROTÉINES

où c_0 est une constante, $R_{ij} = b|i - j|^\nu$ la distance quadratique moyenne entre les monomères i et j , avec b la longueur effective de Kuhn du polymère et ν l'exposant régissant la taille du polymère. L'exposant g gouverne le comportement du terme correctif dans la fonction de partition du polymère. Dans le cadre de simulation numérique sur un réseau à D dimensions, \mathcal{Z}_P s'écrit [98] :

$$\mathcal{Z} = C\bar{q}^N N^{\nu g} \quad (2.105)$$

où C est une constante et \bar{q} le nombre de coordination effectif des noeuds du polymère sur le réseau. En guise d'illustration, ces quantités peuvent être déterminées exactement dans le cas d'une statistique gaussienne. Pour cette dernière, le comportement complet de la fonction de distribution $P_{ij} = P_{ij}^G$ est connue exactement :

$$P_{ij}^G(R) = \left(\frac{3}{2\pi [R_{ij}^G]^2} \right)^{3/2} \exp \left[-\frac{3}{2} \left(\frac{R}{R_{ij}^G} \right)^2 \right], \quad (2.106)$$

avec $R_{ij}^G = b_G|i - j|^{\nu_G}$ où $b_G = \sqrt{2l_m l_p}$ et $\nu_G = 1/2$. La limite asymptotique de P_{ij}^G pour $R/R_{ij}^G \rightarrow \infty$ est simplement :

$$P_{ij}^G \xrightarrow{\frac{R}{R_{ij}^G} \rightarrow \infty} \left(\frac{3}{2\pi} \right) R_{ij}^{G-3} \quad (2.107)$$

d'où $g_G = 0$ et $c_0^G = \frac{3}{2\pi}$. La fonction de distribution complète pour un polymère gaussien est très proche de son comportement asymptotique pour $|i - j| > n_{\text{inf}}$. Ainsi, nous utiliserons la forme asymptotique, plus simple, de la statistique gaussienne par la suite. Pour un polymère auto-évitant à trois dimensions, ces quantités peuvent être estimées grâce à des calculs basés sur des énumérations exactes. Nous trouvons que $\nu_{\text{SAP}} = 0,588$, $g_{\text{SAP}} = 0,27$, $b_{\text{SAP}} = 1,17b_G$ et $c_0^{\text{SAP}} = 0,58$ (voir Annexe 1). Nous pouvons montrer que le noyau G_{ij} décroît algébriquement à longue distance :

$$G_{ij} \longrightarrow K|i - j|^{-\alpha}. \quad (2.108)$$

La constante K gouverne l'intensité des interactions à longue portée et dépend du choix du potentiel spatial $U(R)$:

$$K = \frac{4\pi c_0 k_B T}{b^{3+g}} \int_0^{+\infty} dR R^{2+g} (e^{-\beta U(R)} - 1). \quad (2.109)$$

L'exposant α contrôle la portée effective de ces interactions. Il ne dépend pas du choix de $U(R)$ mais seulement des exposants du polymère considéré :

$$\alpha = (3 + g)\nu. \quad (2.110)$$

2.4. POLYMÈRE À 3D INTERAGISSANT AVEC UN BAIN DE PROTÉINES

Ce dernier paramètre est le paramètre clé pour prédire les transitions de phase dans les systèmes composés de particules en interaction sur un polymère. F. J. Dyson a démontré un critère d'existence des transitions de phase dans les modèles d'Ising unidimensionnel avec interactions à portée algébriquement décroissante basé sur l'exposant α [99]. Ces transitions sont possibles si et seulement si :

$$1 < \alpha \leq 2. \quad (2.111)$$

La transition du cas $\alpha = 2$ intervient dans la résolution du problème de Kondo à l'aide d'un modèle d'Ising avec interaction à longue portée algébrique, et est du type Berezinsky-Kosterlitz-Thouless [10, 101]. Les exposants critiques du modèle sont classiques lorsque $1 < \alpha < 1,5$ [102]. Le modèle LRLG appartient à la classe d'universalité des modèles d'Ising à longue portée. Dans notre cas, α dépend uniquement des propriétés du polymère et nous pouvons donner les valeurs de α pour les statistiques archétypales. Pour un polymère gaussien, on trouve exactement $\alpha_G = 1,5$. Nous attendons la même valeur pour une phase globulaire à cause des effets d'écrantage des interactions d'auto-évitement à cause de la compaction de la chaîne [103]. La valeur pour un polymère auto-évitant peut être estimée à $\alpha_{SA} = 1,92$. Puisque l'ADN doit correspondre à un polymère auto-évitant corrigé par des interactions attractives entre monomères (via les protéines adsorbées ou d'autres types d'interaction), la valeur de l'exposant pour l'ADN α_{DNA} doit exister entre ces deux valeurs :

$$\alpha_G < \alpha_{DNA} < \alpha_{SAP}. \quad (2.112)$$

Nous montrons d'une part que les interactions à longue portée le long de l'ADN provenant des fluctuations conformationnelles du polymères sont suffisantes pour déclencher la séparation de phase des protéines liées à l'ADN. D'autre part, ces séparations de phase correspondent à des transitions de phase thermodynamiquement bien définies, similaires aux transitions ferromagnétiques ou liquide-vapeur.

2.4.4 Diagrammes de phase

Nous souhaitons désormais produire le diagramme de phase du modèle LRLG et connaître précisément le régime thermodynamique du système biologique. Nous allons utiliser la méthode variationnelle de type 2 présentée en première partie. Nous définissons donc comme hamiltonien de référence :

$$\mathcal{H}_0 = -\frac{J}{2} \sum_{\langle i,j \rangle} \Phi_i \Phi_j - \mu_0 \sum_{i=1}^N \Phi_i \quad (2.113)$$

Le système de référence est donc un gaz sur réseau avec interaction à courte portée dont la résolution exacte est connue (voir partie Solution exacte pour

2.4. POLYMÈRE À 3D INTERAGISSANT AVEC UN BAIN DE PROTÉINES

$D = 1$ dans la première section sur le gaz sur réseau à courté portée). Le potentiel grand canonique exact associé dans la limite thermodynamique est :

$$\begin{aligned}\Omega_0(X) &= -Nk_B T \ln \lambda_+ \\ &= -Nk_B T \ln \left\{ e^X \left[\cosh(X) + \sqrt{\sinh^2(X) + B} \right] \right\}\end{aligned}\quad (2.114)$$

où $X = \beta(J + \mu_0)/2$ et $B = \exp(-\beta J)$. Notre choix implique que l'hamiltonien variationnel $\Delta\mathcal{H}$ s'écrit :

$$\Delta\mathcal{H} = -\frac{1}{2} \sum_{i,j}^N \Phi_i G_{ij} \Phi_j - (\mu - \mu_0) \sum_{i=1}^N \Phi_i. \quad (2.115)$$

Ainsi, le système de référence gère les interactions à courte portée tandis que le terme variationnel contient les interactions à longue portée. Nous proposons d'appliquer cette méthode dans un premier temps sur un modèle plus simple que le gaz sur réseau à longue portée.

Gaz sur réseau avec interaction à portée infinie

Le modèle de gaz sur réseau avec interactions à portée infinie (ou IRLG pour Infinite Range Lattice Gas) est défini comme un cas spécial du modèle LRLG dans lequel l'expression du noyau G_{ij} est simplement :

$$G_{ij} \rightarrow \frac{K_\infty}{N}. \quad (2.116)$$

Le noyau G_{ij} devient donc une constante. L'hamiltonien correspondant est :

$$\mathcal{H}_{\text{IRLG}} = \mathcal{H}_{\text{SRLG}} - \frac{K_\infty}{N} \sum_{i < j} \Phi_i \Phi_j \quad (2.117)$$

Dans ce modèle, les protéines proches ou lointaines interagissent avec la même énergie K_∞/N . Le facteur de normalisation N^{-1} permet de garder l'énergie libre extensive et bien définie dans la limite thermodynamique. Le potentiel d'essai peut s'écrire $\Omega_V = \Omega_0 + \langle \Delta\mathcal{H} \rangle_0$ avec dans notre cas :

$$\langle \Delta\mathcal{H} \rangle_0 = -\frac{1}{2} N K_\infty \Phi_0^2 - N(\mu - \mu_0) \Phi_0 \quad (2.118)$$

dans la limite thermodynamique et avec $\Phi_0 = \langle \Phi_i \rangle_0$. En effet, on a $\langle \Phi_i \Phi_j \rangle_0 = \Phi_0^2 + \Phi_0(1 - \Phi_0)r^{|i-j|}$, le second terme en r donnant deux autres termes d'ordre N^0 et N^{-1} négligeables dans la limite $N \rightarrow +\infty$. La valeur de μ_0 est obtenue en inversant la relation :

$$\Phi_0 = \frac{1}{2} \left(1 + \frac{\sinh X}{\sqrt{\sinh^2 X + B}} \right) \quad (2.119)$$

2.4. POLYMÈRE À 3D INTERAGISSANT AVEC UN BAIN DE PROTÉINES

où $B = \exp(-\beta J)$ et $X = (\mu_0 + J)/2$, soit :

$$\mu_0 = 2k_B T \left[\ln \left(\sqrt{1 - A^2 + BA^2} + A\sqrt{B} \right) - \frac{1}{2} \ln(1 - A^2) \right] - J \quad (2.120)$$

où $A = 2\Phi_0 - 1$. L'étape d'optimisation donne en réalité le lien entre le potentiel chimique du modèle de gaz sur réseau avec interactions à portée infinie et le potentiel chimique d'essai μ_0^* :

$$\mu \approx \mu_{\text{IRLG}} = \mu_0^* - K_\infty \Phi \quad (2.121)$$

avec $X \rightarrow X^* = (\mu_0^* + J)/2$. On peut montrer qu'en minimisant Ω_V :

$$\Phi = -\frac{1}{N} \left(\frac{\partial \Omega_{\text{IRLG}}}{\partial \mu} \right) \approx -\frac{1}{N} \left(\frac{\partial \Omega_V^*}{\partial \mu} \right) = \Phi_0. \quad (2.122)$$

Nous avons finalement l'expression du potentiel grand canonique du modèle IRLG dans la limite thermodynamique :

$$\Omega_{\text{IRLG}} \approx \Omega_V^* = \Omega_0(X^*) + \frac{1}{2} N K_\infty \Phi^2. \quad (2.123)$$

Dans l'ensemble grand canonique, la pression est simplement reliée au potentiel grand canonique par la relation :

$$\Omega_{\text{IRLG}} = -P_{\text{IRLG}} V. \quad (2.124)$$

L'énergie libre canonique de Helmholtz est obtenue par la transformation de Legendre $\mathcal{F}_{\text{IRLG}} = \Omega_{\text{IRLG}} + N\mu_{\text{IRLG}}\Phi$:

$$\begin{aligned} \mathcal{F}_{\text{IRLG}} \approx & -\frac{1}{2} N K \phi^2 - N J \Phi + N k_B T \left[A \ln \left(\sqrt{1 - A^2 + BA^2} + A\sqrt{B} \right) \right. \\ & \left. - \ln \left(\sqrt{1 - A^2 + BA^2} + \sqrt{B} \right) + \frac{1 - A}{2} \ln(1 - A^2) \right] \end{aligned} \quad (2.125)$$

Nous pouvons montrer en utilisant une transformation de Hubbard-Stratonovich que ces résultats sont exacts dans la limite thermodynamique [104, 105]. Comme attendu, pour $K_\infty = 0$ et une valeur quelconque de J , la pression et le potentiel chimique ne montre aucun signe de transition de phase quel que soit la valeur de $K_\infty/(k_B T)$. C'est par contre le cas pour des K_∞ et des J différents de 0 et une température suffisamment faible. Nous pouvons déterminer l'expression de la courbe de coexistence et de la ligne spinodale de cette transition en utilisant la symétrie particule-trou. Nous trouvons pour la courbe de coexistence :

$$2k_B T_{\text{IRLG}}^{\text{coex}} \ln \left(\frac{\sqrt{1 - A^2 + B_{\text{co}} A^2} + A\sqrt{B_{\text{co}}}}{\sqrt{1 - A^2 + B_{\text{co}} A^2} - A\sqrt{B_{\text{co}}}} \right) - K_\infty A = 0 \quad (2.126)$$

2.4. POLYMÈRE À 3D INTERAGISSANT AVEC UN BAIN DE PROTÉINES

où $B_{\text{co}} = \exp[J/(k_B T_{\text{IRLG}}^{\text{coex}})]$. Cette équation donne la valeur de la température de transition $T_{\text{IRLG}}^{\text{coex}}$ pour un Φ donné. De même, la ligne spinodale vérifie :

$$4k_B T_{\text{IRLG}}^{\text{spino}} \left(\frac{1}{1 - A^2} - \frac{1}{1 - A^2 + B_{\text{sp}} A^2 + A \sqrt{B_{\text{sp}} - B_{\text{sp}} A^2 + B_{\text{sp}}^2 A^2}} \right) - AK_{\infty} = 0 \quad (2.127)$$

où $B_{\text{sp}} = \exp[J/(k_B T_{\text{IRLG}}^{\text{coex}})]$. Ces deux équations peuvent être résolues numériquement. La limite $\Phi \rightarrow \Phi_C = 1/2$ dans une de ces équations permet d'obtenir l'équation transcendante vérifiée par la température critique :

$$4k_B T_{\text{IRLG}}^{\text{C}} \sqrt{B_C} - K_{\infty} = 0. \quad (2.128)$$

Notons que nous revenons à une description en température à cause de la non-linéarité de nos équations variationnelles. Nous pouvons montrer que les exposants critiques du modèle IRLG sont classiques (voir Annexe 2). Nous comparons les données obtenues par cette méthode variationnelle avec les résultats en champ moyen. Ces derniers peuvent être déterminés en définissant un hamiltonien de référence identique à celui utilisé dans la partie sur le gaz sur réseau avec interactions à courte portée mais aussi, d'une manière intéressante, en prenant la limite $J \rightarrow 0$ dans nos expressions variationnelles. En effet, dans l'hamiltonien de référence eq. (2.113), prendre $J = 0$ revient à utiliser l'approximation du champ moyen dans un système sans interactions à courte portée. Ainsi, nous trouvons notamment pour l'énergie libre de Helmholtz :

$$\begin{aligned} \mathcal{F}_{\text{IRLG}}^{\text{MF}} = \lim_{J \rightarrow 0} \mathcal{F}_{\text{IRLG}}^{\text{VC}} = & -\frac{1}{2} N (2J + K_{\infty}) \Phi^2 \\ & + N k_B T [\Phi \ln \Phi + (1 - \Phi) \ln (1 - \Phi)] \end{aligned} \quad (2.129)$$

et pour la température critique du champ moyen :

$$T_{\text{IRLG}}^{\text{MF}} = \frac{(2J + K_{\infty})}{4k_B}. \quad (2.130)$$

Ces résultats indiquent que le modèle IRLG en champ moyen se comporte comme un gaz sur réseau avec interactions à courte portée d'intensité effective $J' = J + K_{\infty}/2$. Dans la figure 2.12, nous comparons l'évolution de la température critique en fonction des paramètres J et K_{∞} dans le cas $J = K_{\infty}$. Le champ moyen surestime la température critique exacte. Soulignons que la méthode variationnelle donne des résultats exacts pour le modèle IRLG dans la limite thermodynamique.

Gaz sur réseau avec interactions à longue portée

Nous illustrons nos résultats en choisissant pour le potentiel spatial d'interaction entre protéines $U(R)$ un potentiel à puits carré d'une profondeur u_0 et

2.4. POLYMÈRE À 3D INTERAGISSANT AVEC UN BAIN DE PROTÉINES

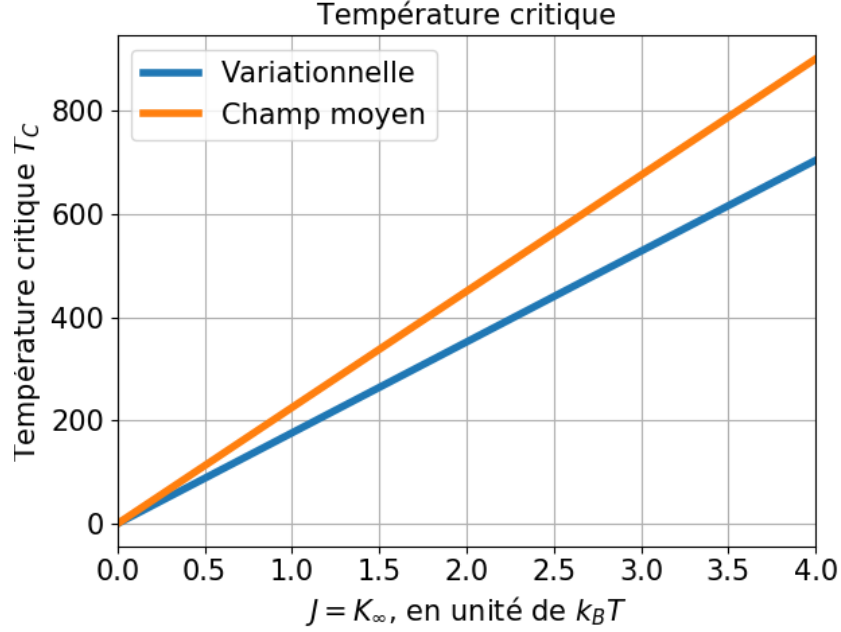


FIGURE 2.12 – **Température critique pour $J = K_\infty$** : en orange, les résultats en champ moyen ; les résultats exacts issus de la méthode variationnelle en bleu. Le champ moyen surestime la température critique.

d'une portée a , avec des interactions de sphères dures d'une portée σ [78, 77], soit :

$$U(R) = \begin{cases} +\infty, & \text{if } R < \sigma \\ -u_0, & \text{if } \sigma < R < a \\ 0, & \text{if } R > a \end{cases} \quad (2.131)$$

Le noyau d'interactions G_{ij} se comporte alors asymptotiquement comme $G_{ij} \rightarrow K_{SW}|i-j|^{-\alpha}$ pour $|i-j| \rightarrow +\infty$ où l'intensité K_{SW} est définie par :

$$K_{SW}(T) = 4\pi k_B T \frac{c_0}{3+g} \left(\frac{\sigma}{b}\right)^{3+g} \left\{ (e^{\beta u_0} - 1) \left[\left(\frac{a}{\sigma}\right)^{3+g} - 1 \right] - 1 \right\}. \quad (2.132)$$

Soulignons que K_{SW} dépend de la température. Ce choix nous permet d'illustrer simplement le comportement de l'intensité $K \rightarrow K_{SW}$ dans le cas d'un potentiel spatial U possédant une répulsion à courte portée et une attraction à plus grande portée. Une intensité K_{SW} positive traduit des interactions à longue portée effectives attractives : c'est le cas lorsque la température est suffisamment faible. Elle décroît ensuite avec T jusqu'à devenir négative et traduire des interactions répulsives : à haute température, le terme d'interaction $\exp(\beta u_0) - 1$ tend vers 0 et il ne reste que la partie répulsive du type sphère dure de U . Le régime d'intérêt, à faible température, correspond au cas attractif. K_{SW} augmente avec la profondeur du puits u_0 et la portée a ; elle

2.4. POLYMÈRE À 3D INTERAGISSANT AVEC UN BAIN DE PROTÉINES

décroit avec la courte portée σ mais aussi avec la longueur de Kuhn effective b et l'exposant g car la rigidité et l'auto-évitement du polymère limite le nombre possible de *bridging*. Nous avons cette fois :

$$\langle \Delta \mathcal{H} \rangle_0 = \frac{1}{2} \sum'_{i,j} \langle \Phi_i \Phi_j \rangle_0 - (\mu - \mu_0) \sum_{i=1}^N \langle \Phi_i \rangle_0. \quad (2.133)$$

L'expression de Ω_0 est donnée par l'équation (2.114) et μ_0^* par l'équation (2.120). En conséquence, l'étape d'optimisation donne pour l'approximation du potentiel chimique :

$$\mu \approx \mu_0^* + 2\Phi_0^* [S' - S] - S' - \Phi_0^* (1 - \Phi_0^*) (1 - 2\Phi_0^*) f(X^*) S''. \quad (2.134)$$

où :

$$f(X^*) = \beta \left(\frac{\partial \Phi_0}{\partial \mu_0} \right)_{\mu_0 = \mu_0^*}^{-1} = \frac{e^{-\beta J} \cosh \left[\frac{\beta}{2} (\mu_0^* + J) \right]}{\left\{ \sinh^2 \left[\frac{\beta}{2} (\mu_0^* + J) \right] + e^{-\beta J} \right\}^{\frac{3}{2}}} \quad (2.135)$$

est une fonction de $X^* = \beta(J + \mu_0^*)/2$. Les sommes S , S' et S'' dépendent de manière importante de K_{SW} et de l'exposant α . On a par exemple pour S :

$$\begin{aligned} S &\approx \sum_{k=n_{\text{inf}}}^{\infty} G_k = \sum_{k=1}^{\infty} G_k - \sum_{k=1}^{n_{\text{inf}}-1} G_k \\ &= K_{\text{SW}} \left[\zeta(\alpha) - \sum_{k=1}^{n_{\text{inf}}-1} k^{-\alpha} \right], \end{aligned} \quad (2.136)$$

où $\zeta(x)$ est la fonction de Riemann. Le premier terme $K_{\text{SW}}\zeta(\alpha)$ correspond à la contribution asymptotique dans le cas où $n_{\text{inf}} = 0$. Cette première partie est corrigé par le second terme qui prend en compte l'effet de la longueur de persistance. Ainsi, c'est dans $S(T)$ que la longueur de persistance n_{int} prend son importance : plus la longueur de persistance est grande, plus $S(T)$ et les contributions des interactions à longue portée sont faibles [59, 69]. De la même manière, on a :

$$S' = K_{\text{SW}} \left[\text{Li}_{\alpha}(r_{\text{LG}}) - \sum_{k=1}^{n_{\text{inf}}-1} \frac{r_{\text{LG}}^k}{k^{\alpha}} \right] \quad (2.137)$$

où $r_{\text{LG}} = \lambda_-/\lambda_+$, et

$$S'' = K_{\text{SW}} \left[\text{Li}_{\alpha-1}(r_{\text{LG}}) - \sum_{k=1}^{n_{\text{inf}}-1} \frac{r_{\text{LG}}^k}{k^{\alpha-1}} \right], \quad (2.138)$$

où

$$\text{Li}_s(z) \equiv \sum_{k=1}^{\infty} \frac{z^k}{k^s} \quad (2.139)$$

2.4. POLYMÈRE À 3D INTERAGISSANT AVEC UN BAIN DE PROTÉINES

est la fonction polylogarithme. Comme dans le cas du modèle IRLG, nous pouvons montrer qu'à l'optimisation :

$$\Phi \equiv \Phi_{\text{LRLG}} = -\frac{1}{N} \left(\frac{\partial \Omega_{\text{LRLG}}}{\partial \mu} \right) \approx -\frac{1}{N} \left(\frac{\partial \Omega_V^*}{\partial \mu} \right) = \Phi_0. \quad (2.140)$$

Ainsi, nous obtenons l'expression d'approximation du potentiel grand canonique dans l'approche variationnelle :

$$\begin{aligned} \Omega_{\text{LRLG}} \approx \Omega_V^* &= \Omega_0(X^*) + N\Phi^2(S - S') \\ &+ N\Phi^2(1 - \Phi)(1 - 2\Phi)f(X^*)S''. \end{aligned} \quad (2.141)$$

Nous obtenons l'expression de la pression P_{LRLG} à partir de l'expression de ce grand potentiel. Ce modèle montre aussi un comportement de transition de phase du premier ordre. L'égalité du potentiel chimique et de la pression dans les états de haute Φ_h et basse Φ_b occupation donne l'équation vérifiée par la courbe de coexistence. Avec la symétrie particule-trou $\Phi_h + \Phi_b = 1$, nous trouvons :

$$\begin{aligned} &2k_B T \ln \left(\frac{\sqrt{1 - A_b^2 + BA_b^2} + A_b\sqrt{B}}{\sqrt{1 - A_b^2 + BA_b^2} - A_b\sqrt{B}} \right) \\ &- 2\Phi_b(S - S'_b) + 2(1 - \Phi_b)(S - S'_h) \\ &+ \Phi_b(1 - \Phi_b)A_b [S''_b f(\Phi_b) + S''_h f(1 - \Phi_b)] = 0, \end{aligned} \quad (2.142)$$

avec $A_b = 2\Phi_b - 1$, $S'_{b/h} = S'(\Phi_{b/h})$ et $S''_{b/h} = S''(\Phi_{b/h})$. Cette équation donne la valeur de l'occupation de l'état de basse occupation Φ_b à la température T si T est inférieure à la température critique variationnelle T_c^{VC} . Au point critique $\Phi = \Phi_C = 1/2$ et $T = T_c^{\text{VC}}$, nous pouvons montrer que $S'_b, S'_h \rightarrow S'_c$ et $S''_b, S''_h \rightarrow S''_c$ car $r_{LG} \rightarrow (1 - \sqrt{B_C}) / (1 + \sqrt{B_C})$ avec $B_C = \exp[-J/(k_B T_c)]$. La même remarque s'impose pour $S''_b, S''_h \rightarrow S''_c$. De plus, $f \rightarrow 4\sqrt{B_C}$ et l'équation (2.142) devient :

$$2k_B \sqrt{B_C} T_c^{\text{VC}} - (S - S') + S'' \sqrt{B_C} = 0, \quad (2.143)$$

et nous obtenons finalement l'équation vérifiée par la température critique prédite par l'approche variationnelle :

$$T_c^{\text{VC}} = \frac{S_c - S'_c}{2k_B \sqrt{B_C}} - \frac{S''_c}{2k_B}. \quad (2.144)$$

avec $S \rightarrow S_c = S(T = T_c^{\text{VC}})$. Comme pour le modèle IRLG, la limite $J \rightarrow 0$ permet d'obtenir les résultats dans l'approximation du champ moyen. Nous trouvons pour l'énergie libre de Helmholtz :

$$\begin{aligned} \mathcal{F}_{\text{LRLG}}^{\text{MF}} &= -N [J + S(T)] \Phi^2 \\ &+ Nk_B T [\Phi \ln \Phi + (1 - \Phi) \ln (1 - \Phi)]. \end{aligned} \quad (2.145)$$

2.4. POLYMÈRE À 3D INTERAGISSANT AVEC UN BAIN DE PROTÉINES

Ce résultat est isomorphe à celui d'un gaz sur réseau avec un couplage effectif $J' = J + S(T)$ entre site premiers voisins. Nous soulignons la dépendance non-linéaire en température de la somme S qui nous empêche de déterminer l'expression analytique de la courbe de coexistence et de la ligne spinodale pour le modèle LRLG en champ moyen, contrairement aux modèles SRLG et IRLG. La courbe de coexistence vérifie alors :

$$T_{\text{coex}}^{\text{MF}} = \frac{(2\Phi - 1) [J + S(T_{\text{coex}}^{\text{MF}})]}{k_B \ln\left(\frac{\Phi}{1-\Phi}\right)} \quad (2.146)$$

tandis que la ligne spinodale dépend de l'équation suivante :

$$T_{\text{spino}}^{\text{MF}} = \frac{2 [J + S(T_{\text{coex}}^{\text{MF}})]}{k_B} \Phi(1 - \Phi) \quad (2.147)$$

Ces équations donnent la température de transition $T_{\text{coex}}^{\text{MF}}$ et la température spinodale $T_{\text{spino}}^{\text{MF}}$ du modèle LRLG en champ moyen pour un Φ donné. La température critique T_c^{MF} prédite par le champ moyen vérifie :

$$T_c^{\text{MF}} = \frac{J + S(T_{\text{coex}}^{\text{MF}})}{2k_B} \quad (2.148)$$

Les données disponibles pour le système ParABS nous permettent de paramétrer le modèle LRLG. En dehors de la température biologique de l'ordre de $T_{\text{bio}} = 300$ K et de son occupation moyenne $\Phi_{\text{bio}} = 0,08$ (voir la partie Scénario métastable), nous considérons trois types de paramètres (voir tableau 2.1) :

1. Les paramètres qui contrôlent le polymère et sa statistique. Notre modèle effectif d'ADN possède des monomères de taille l_m , égale à l'empreinte d'une protéine ParB sur l'ADN, soit $l_m = 16$ bp $\approx 5,44$ nm [22]. La longueur effective de Kuhn est $b = \sqrt{2l_p l_m} \approx 23,6$ nm avec la longueur de persistance $l_p \approx 51$ nm et donc $n_{\text{inf}} = l_p/l_m = 10$ [103]. Le paramètre n_{inf} contribue à diminuer lourdement l'importance du noyau G_{ij} puisque les premiers termes, les plus importants, sont filtrés. Le coefficient c_0 est égal à $[3/(2\pi)]^{3/2}$ pour un polymère gaussien, et à 0,58 pour un polymère auto-évitant (voir Annexe 1). Notons que le c_0 d'un polymère auto-évitant étant plus élevé, son K_{SW} sera plus fort. Cet effet est compensé dans le noyau G_{ij} par un exposant g plus important et donc une décroissance contrôlée par l'exposant α plus rapide.
2. Les paramètres estimables relatifs au potentiel spatial U : σ , la portée des interactions de sphère dure, et a , la portée du bridging. Nous choisissons naturellement $\sigma = l_m$, la longueur d'un monomère. Dans l'article [60], la portée a a été choisie telle que les interactions de bridging sont autant à courte portée que les interactions de spreading, soit $a = 2\sigma$. Nous discuterons de la pertinence de ce choix dans un prochain paragraphe.

2.4. POLYMÈRE À 3D INTERAGISSANT AVEC UN BAIN DE PROTÉINES

Paramètres	Symbole	Valeur
Température physiologique	T_{bio}	300 K
Occupation physiologique	Φ_{bio}	$\leq 0,08$
Longueur d'un monomère	l_m	16 bp ($\sim 5,4$ nm)
Longueur de persistance de l'ADN	l_p	150 bp (~ 50 nm)
Longueur minimale d'une boucle d'ADN	n_{inf}	10 sites
Longueur de Kuhn effective	b	23,6 nm
Exposant 1	g	SAP : 0,27, G : 0
Exposant 2	ν	SAP : 0,588, G : 0,5
Portée des sphères dures	σ	l_m
Portée de U(R)	a	2σ ($\sim 10,8$ nm)
Intensité du <i>bridging</i>	u_0	$u_0 = J$
Intensité du <i>spreading</i>	J	0 to $5k_B T$

TABLE 2.1 – **Paramètres du modèle LRLG.** L'exposant 1 correspond à l'exposant du polymère lié à la fonction de partition du polymère tandis que l'exposant 2 est l'exposant lié à la taille du polymère. G pour le cas du polymère gaussien et SAP pour un polymère auto-évitant. Les paramètres libres sont les intensités des interactions de *spreading* J et de *bridging* u_0 .

3. Les paramètres difficile à estimer en l'état : il s'agit des intensités des interactions de spreading et bridging, J et u_0 . Nous les considérons donc comme des paramètres de travail. Pour simplifier notre paramétrisation, nous avons tracé nos courbes pour $J = u_0$.

Nous commençons notre analyse en comparant l'évolution des températures critiques estimées par l'approximation du champ moyen et notre méthode variationnelle pour un polymère gaussien (figure 2.13) et auto-évitant (figure 2.13) en fonction de $J = u_0$. Il existe une valeur minimale J_C de J et u_0 telle que l'approche variationnel ne prédit pas de transition de phase à $T = 300$ K si $J < J_C$. Il faut donc des interactions de *spreading* et *bridging* suffisamment forte pour espérer observer une séparation de phase à température biologique. Consubstantiellement à notre méthode, les deux approximations prédisent des températures critiques similaires lorsque $J = u_0 \rightarrow 0$. La température critique du champ moyen croît ensuite plus rapidement que celle prédite par le calcul variationnel. Nous pouvons montrer que le champ moyen surestime la température critique. En effet, dans le cas où $J = 0$, $n_{\text{inf}} = 0$ et K_{SW} dans l'équation (2.132) est remplacé par un K^{LB} qui ne dépend plus de la température. Luijten et Blöte ont déterminé la valeur de l'intensité critique K_C^{LB} en utilisant des simulations Monte-Carlo [106]. Dans cette limite, notre modèle offre des résultats en champ moyen et nous avons pour le couplage critique à longue portée :

$$K_C^{\text{MF}} = \frac{2}{\zeta(\alpha)} \quad (2.149)$$

2.4. POLYMÈRE À 3D INTERAGISSANT AVEC UN BAIN DE PROTÉINES

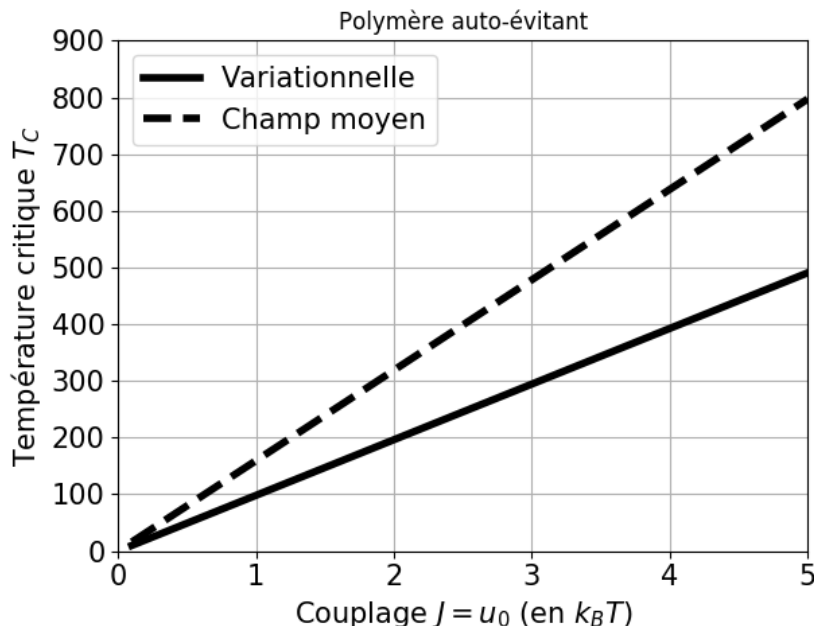


FIGURE 2.13 – **Température critique en fonction de $J = u_0$** : pour un polymère auto-évitant. Il existe un $J = J_C$ pour chaque type de polymère tel que pour tout $J = u_0 > J_C$, la température critique est supérieure à la température biologique $T_{\text{bio}} = 300K$. Il faut des interactions de *spreading* et *bridging* suffisamment fortes pour observer des séparations de phases *in vivo*.

En figure 2.15, nous comparons les résultats de Luijten et Blöte aux prédictions de cette valeur critique en fonction de l'exposant α en champ moyen. Comme attendu, le champ moyen surestime K_C^{LB} avec une erreur proche de celle estimée par le gaz sur réseau à courte portée à trois dimensions pour $\alpha = 1,5$. La température critique (l'intensité critique) diminue (augmente) avec α : la température critique augmente avec la portée des interactions en protéines ParB. De manière générale, la température critique croît avec les paramètres qui favorisent l'importance du noyau G_{ij} . En figure 2.16 et 2.17, nous montrons les diagrammes de phase du modèle LRLG pour les statistiques qui nous intéressent et pour des valeurs de J et u_0 raisonnables : $J = u_0 = 3k_B T$ pour la statistique gaussienne, $3,75k_B T$ pour l'auto-évitante. Nous avons indiqué les coordonnées du système biologique $(\Phi_{\text{bio}}, T_{\text{bio}})$: nous montrons ainsi que, pour des valeurs raisonnables d'intensité des interactions de *spreading* et *bridging*, le régime biologique est le régime de coexistence métastable en basse occupation.

Cependant, il s'avère que notre première estimation de la portée a des interactions de *bridging* n'est pas réaliste. Nous avons assimilé les protéines ParB à des sphères de rayon σ pouvant interagir entre elles dans l'espace dans une certaine couche de taille $a - \sigma$ via le potentiel U . Cette conception entre

2.4. POLYMÈRE À 3D INTERAGISSANT AVEC UN BAIN DE PROTÉINES

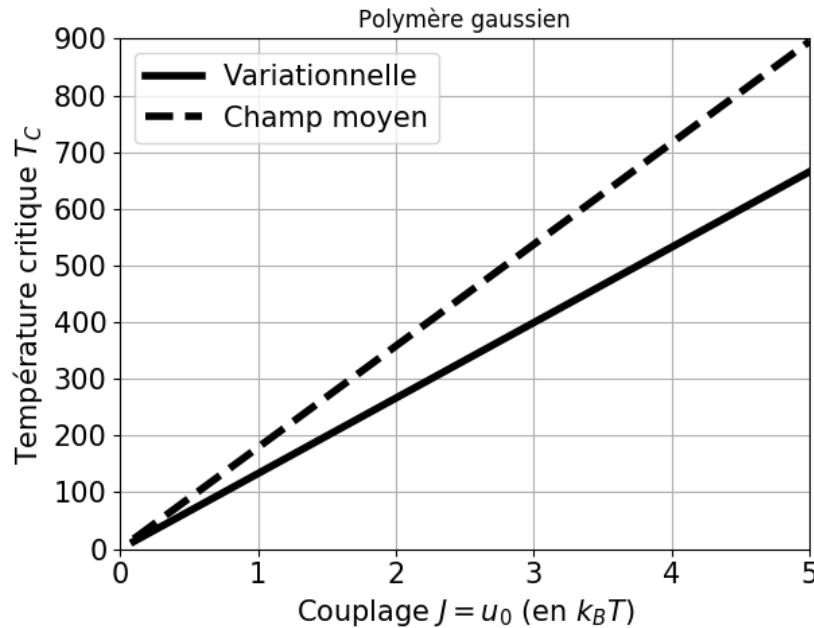


FIGURE 2.14 – Température critique en fonction de $J = u_0$: pour un polymère gaussien. Même remarque pour la figure 2.13.

dans une certaine mesure en contradiction avec la forme des protéines ParB. Chaque sous-unité des dimères de ParB possèdent en réalité un motif intrinsèquement non replié, une sorte de queue, d’une longueur linéaire de 156 pb ou $0,36 \times 156 \approx 56$ nm. La séquence d’acides aminés permettant les interactions de *bridging* entre deux dimères se situe sur ce motif. Loin d’être une structure quasi-cristalline, il faut imaginer le complexe comme un objet très poreux, semblable à une toile d’araignée. Nous avons ainsi sous-estimé la portée des interactions de *bridging*. Pour illustration, dans le cas où la portée a des interactions de *bridging* est de l’ordre de 35 nm, les nouveaux résultats n’entravent pas notre interprétation (voir figure 2.18 et 2.19). Néanmoins, les valeurs de J et u_0 requises pour exister dans le régime de coexistence métastable sont plus faibles : elles doivent être de l’ordre de $1 k_B T_r$ pour le polymère gaussien ou $1,75 k_B T_r$ pour l’auto-évitant, ce qui est en accord avec d’autres estimations de l’intensité J du *spreading* (voir section suivante sur le *looping* et *clustering*).

Nous avons montré que les protéines ParB pouvaient subir une séparation de phase unidimensionnelle et que le régime biologique des protéines ParB sur l’ADN pouvait être le régime de coexistence métastable en basse occupation. Dans la première partie sur le traitement en champ moyen du gaz sur réseau à D dimensions, nous avons émis une hypothèse similaire que nous avons appelé le scénario métastable. En effet, dans l’hypothèse d’une formation des complexes ParBS comme le résultat d’une séparation de phase, il s’agit du seul régime capable d’expliquer pourquoi il n’y a de complexes qu’en pré-

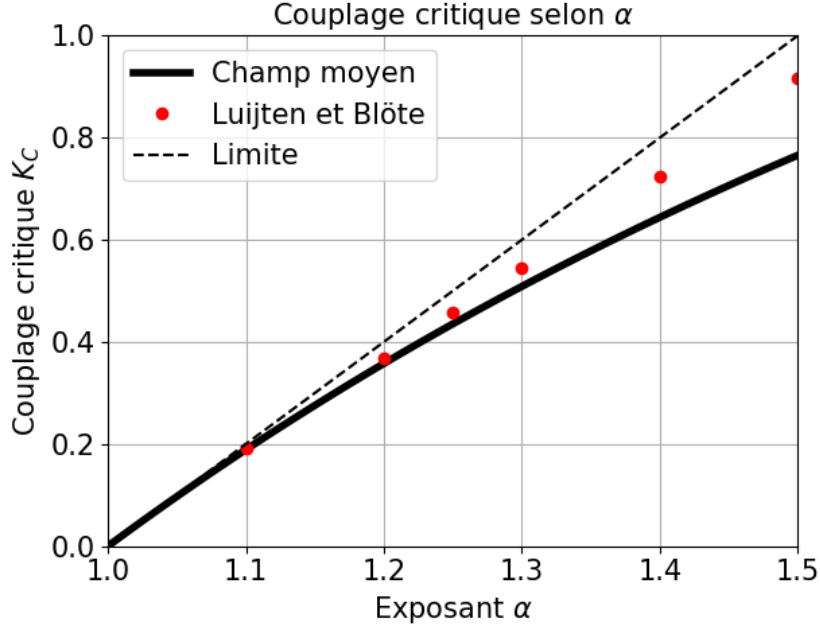


FIGURE 2.15 – **Temperature critique en fonction de la portée α du noyau G_{ij}** : les résultats de notre approche en champ moyen lorsque $G_{ij} \sim K|i - j|^{-\alpha}$ avec K sans dépendance en température sont en noir et trait plein. En trait discontinu, la limite du champ moyen lorsque $\alpha \rightarrow 1$. Les points rouges correspondent aux simulations de Luijten et Blöte.

sence des séquences *parS*. Nous pouvons réitérer cette hypothèse dans le cadre plus rigoureux du gaz sur réseau à longue portée. Dans la prochaine section, nous explorons les liens entre cette séparation de phase unidimensionnelle et la structure tridimensionnelle des complexes ParBS.

2.5 Lien entre séparation de phase 1D et 3D

Nous souhaitons montrer le lien entre la séparation de phase unidimensionnelle démontrée dans la section précédente et la structure des complexes ParBS, c'est-à-dire la séparation de phase tridimensionnelle des protéines ParB explicitement observée dans les expériences de SPT-PALM. Les interactions de *bridging* entre les protéines ParB dans l'espace depuis deux sites non adjacents doivent aboutir à une structure tridimensionnelle, comme dans notre modèle de gaz sur réseau à courte portée pour $D = 3$, dans lequel l'ADN forme un réseau tridimensionnel. Nous pouvons faire le lien entre ces deux dimensions grâce aux expériences de séquençage CHIP (voir chapitre Introduction et figure 1.2). Ces données montrent la statistique des protéines ParB formant effectivement un complexe tridimensionnel.

Dans cette section, nous présentons une nouvelle approche qui prend ses

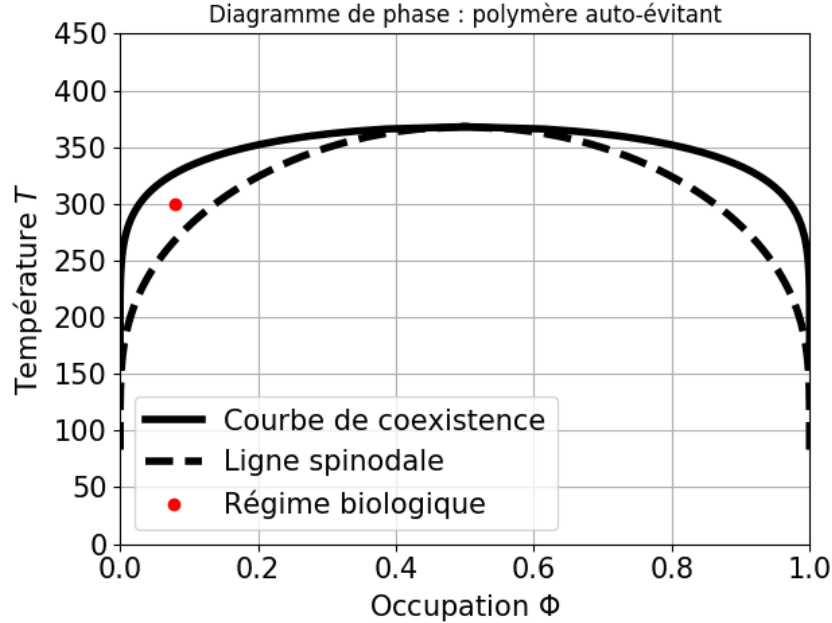


FIGURE 2.16 – **Diagramme de phase du modèle LRLG** : pour un polymère auto-évitant ($J = u_0 = 3,75k_B T_r$). La courbe de coexistence est en trait plein, la ligne spinodale en trait discontinu. Le point rouge correspond aux conditions biologiques de l'ADN et des protéines ParB.

racines dans le modèle de *spreading* et *bridging* en reformulant les interactions de *bridging* comme des phénomènes de création de boucles. Ces interactions sont considérées comme suffisantes pour former un complexe. Ce modèle unidimensionnel permet de lier le profil expérimental d'occupation des protéines ParB autour de la séquence *parS* à notre propre approche de gaz sur réseau avec interactions à longue portée.

2.5.1 *Looping et clustering*

Nous présentons succinctement le modèle de *spreading* et *bridging* sur réseau [55]. Le système physique consiste en un polymère de longueur L qui occupe les noeuds d'un réseau cubique à trois dimensions de maille l_m correspondant au $16pb$ que forme l'empreinte d'une protéine ParB sur l'ADN. Un nombre fini M de protéines ParB occupent ce réseau. Ces protéines peuvent diffuser le long de l'ADN. De plus, nous considérons que l'ADN constitue un réseau mouvant et dynamique. Les protéines ParB peuvent interagir au travers de deux types d'interactions :

1. l'interaction de *spreading* entre deux protéines voisines le long de l'ADN, d'intensité J déjà définie
2. et l'interaction de *bridging* entre deux protéines voisines dans l'espace

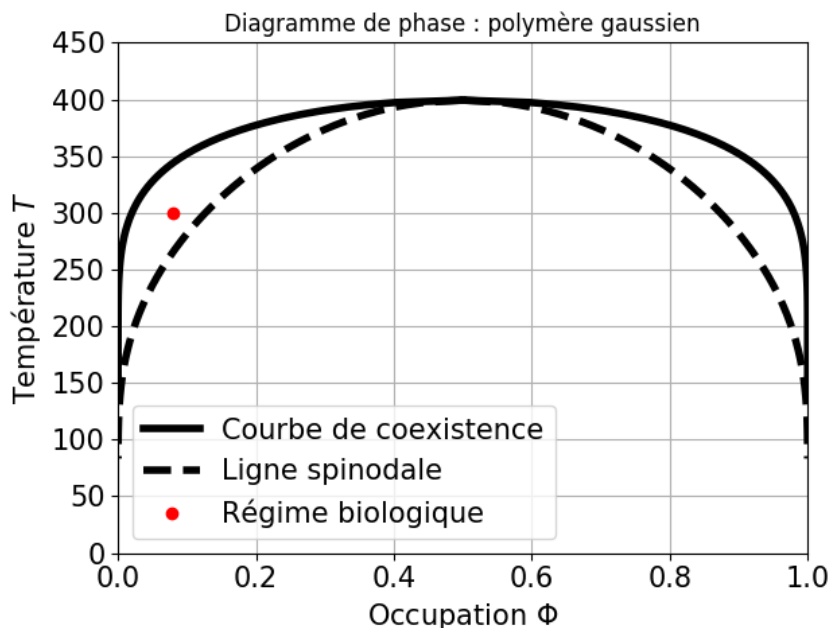


FIGURE 2.17 – **Diagramme de phase du modèle LRLG** : pour un polymère gaussien ($J = u_0 = 3k_B T_r$). La courbe de coexistence est en trait plein, la ligne spinodale en trait discontinu. Le point rouge correspond aux conditions biologiques de l'ADN et des protéines ParB.

mais pas adjacentes sur l'ADN, d'intensité J_B .

Cette approche du *spreading* et *bridging* sur réseau peut être vue comme un compromis entre notre approche de gaz sur réseau à courte portée et notre propre *spreading* et *bridging* dans l'espace continu à trois dimensions. C'est principalement l'interaction de *bridging* qui modifie la statistique du polymère puisque cette interaction est liée à la formation des boucles d'ADN. Le couplage J_B est considéré comme suffisamment grand pour observer la séparation de phase tridimensionnelle des protéines ParB. Nous rappelons que les interactions de *spreading* ne sont pas suffisantes thermodynamiquement pour engendrer une séparation de phase : l'intensité J_B contrôle pleinement ce mécanisme, comme le noyau d'interaction à longue portée G_{ij} . Notre paramètre de travail sera l'intensité du *spreading* J . La distribution des protéines sur l'ADN sera davantage sensible à ce paramètre puisqu'il faut casser un lien entre deux protéines adjacentes pour former une boucle.

Le *looping* et *clustering* [59] se construit en considérant qu'un complexe tridimensionnel de protéines ParB avec des boucles peut être représenté comme une chaîne unidimensionnelle avec des espaces non occupés correspondant aux boucles d'ADN qui sortent du complexe, et des espaces occupés par les protéines. Le *looping* et *clustering* est ainsi un modèle de protéines adsorbées sur une réseau unidimensionnel d'ADN qui interagissent entre elles à longue dis-

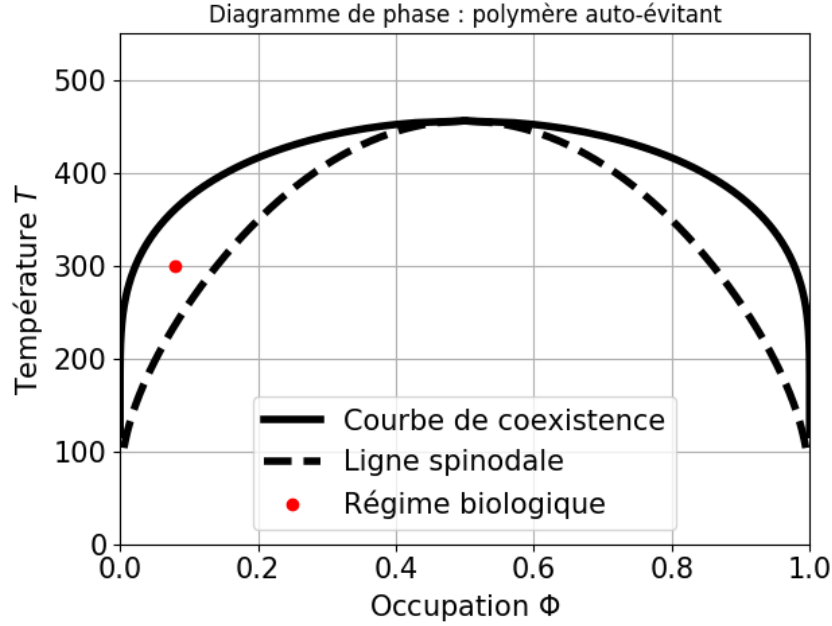


FIGURE 2.18 – **Diagramme de phase du modèle LRLG pour $a \sim 35nm$** : pour un polymère auto-évitant ($J = u_0 = 1,75k_B T_r$). La courbe de coexistence est en trait plein, la ligne spinodale en trait discontinu. Le point rouge correspond aux conditions biologiques de l'ADN et des protéines ParB.

tance. Il s'agit donc d'un modèle similaire au modèle de gaz sur réseau avec des interactions effectives à longue portée. Sa fonction de partition \mathcal{Z}_{LC} possède deux contributions :

1. La première contribution provient de l'entropie de positionnement des boucles sur la chaîne. Le plus simple est de considérer qu'une boucle se forme dès qu'un trou se forme entre deux clusters unidimensionnels. Le nombre de possibilité de placer n boucles dans un cluster de M protéines est simplement :

$$\Omega_P = \frac{(M-1)!}{(M-1-n)!n!} \quad (2.150)$$

L'entropie de positionnement des boucles s'écrit ainsi $S_P = k_B \ln \Omega_P$. Néanmoins, pour former une boucle dans un cluster unidimensionnel, il faut briser une interaction de spreading entre deux protéines ParB. Ainsi, cette contribution doit être corrigée par un facteur de Boltzmann à l'équilibre $\exp(-n\beta J)$.

2. La deuxième contribution retenue concerne la conformation des boucles. Pour la i -ème boucle de longueur l_i , le nombre de configurations possibles est donné par $z^{l_i/l_m} (l_i/l_m)^{-D\nu}$ où z est le nombre de coordination du réseau, D sa dimension et ν , le coefficient de Flory du polymère [97, 107]. En comptant toutes les boucles, considérées indépendantes entre elles, le

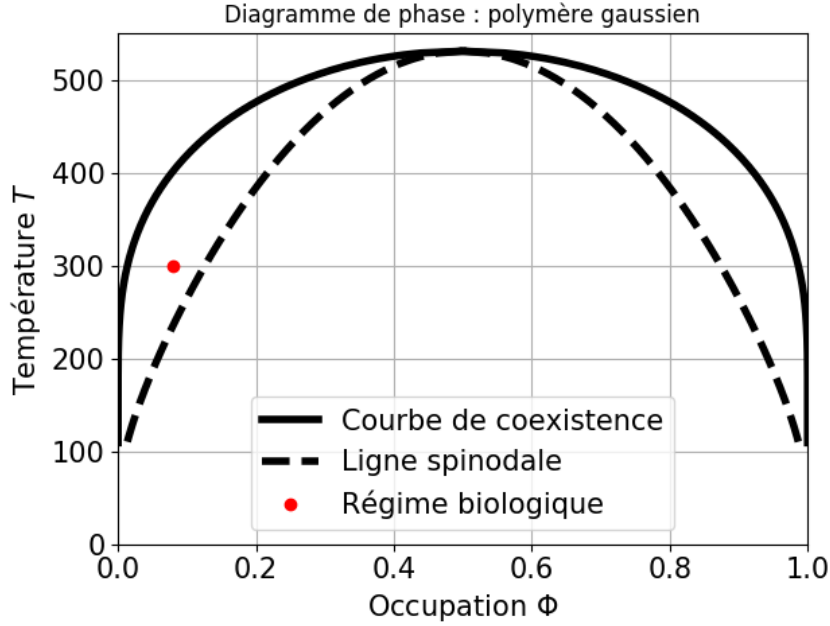


FIGURE 2.19 – Diagramme de phase du modèle LRLG pour $a \sim 35nm$: pour un polymère gaussien ($J = u_0 = 1k_B T_r$). La courbe de coexistence est en trait plein, la ligne spinodale en trait discontinu. Le point rouge correspond aux conditions biologiques de l'ADN et des protéines ParB.

nombre d'état possible s'écrit :

$$\Omega_C = \prod_{i=1}^n \left(\frac{l_i}{l_m} \right)^{-D\nu}. \quad (2.151)$$

Le terme extensif dépendant de z^{l_i} disparaît car, pour un polymère fini, une boucle à l'intérieur du complexe réduit la taille de la boucle que forme le polymère à l'extérieur du complexe. Nous supposons qu'il y a une compensation exacte entre le gain d'entropie provoqué par l'introduction d'un boucle dans le complexe et la perte d'entropie provoquée par cette introduction à l'extérieur du foyer. Nous trouvons ainsi une entropie conformationnelle :

$$S_C = -k_B D\nu \sum_{i=1}^n \ln \left(\frac{l_i}{l_m} \right). \quad (2.152)$$

La fonction de partition du *looping* et *clustering* s'écrit donc :

$$\mathcal{Z}_{\text{LC}} = \sum_{n=0}^{M-1} \prod_{j=1}^n \int_{l_p}^{l_{\max}} dl_j \Omega_{\text{C}} \Omega_{\text{P}} e^{-n\beta J} \quad (2.153)$$

$$= \sum_{n=0}^{M-1} \frac{(M-1)!}{(M-1-n)!n!} \exp(-n\beta J) \prod_{j=1}^n \int_{l_p}^{l_{\max}} dl_j l_j^{-D\nu} \quad (2.154)$$

où l_p est la longueur de persistance de l'ADN, c'est-à-dire la longueur minimale d'une boucle. En toute rigueur, la borne supérieure l_{\max} des intégrales devrait être la longueur restante d'ADN pour former des boucles, c'est-à-dire $L - (Ml_m + L_j)$ où $L_j = \sum l_k$ est la longueur cumulée de boucles. Cependant, la probabilité de former des boucles longues est faible tandis que L domine largement M . La longueur du plasmide F d'*E. Coli*, $L/l_m = N = 3750$ qui est bien supérieur au nombre $M = 300$ de protéines dans un complexe ParBS [22]. La limite $l_{\max} \rightarrow \infty$ donne :

$$\mathcal{Z}_{\text{LC}} \xrightarrow{l_{\max} \rightarrow \infty} [1 + \exp(-\beta J')]^{M-1} \quad (2.155)$$

avec $J' = J + \ln [l_p^{D\nu-1} (D\nu - 1)]$. Dans ce modèle, il y a une compétition entre l'intensité J du *spreading* et l'entropie S_{C} de fermeture, qui favorisent des complexes plus denses et moins de boucles, avec l'entropie S_{P} de positionnement des boucles, qui favorise au contraire des complexes moins denses avec plus de boucles d'ADN.

Cette compétition peut être simplement illustrée en appliquant une méthode de col à la fonction de partition \mathcal{Z}_{LC} . Dans la limite thermodynamique $n \gg 1$, nous pouvons approximer le terme entropique S_{P} en utilisant le formule de Stirling. Nous trouvons alors l'expression classique d'une entropie de mélange de n boucles placées sur un réseau de $M - 1$ sites disponibles :

$$S_{\text{P}}(\rho_\ell) \approx -k_B \{ [1 - \rho_\ell] \ln [1 - \rho_\ell] + \rho_\ell \ln [\rho_\ell] \}, \quad (2.156)$$

où $\rho_\ell = n/(M - 1)$ peut être vue comme une concentration de boucle sur la chaîne, avec $0 \leq \rho_\ell \leq 1$. La fonction de partition du *looping* et *clustering* peut ainsi s'écrire :

$$\mathcal{Z}_{\text{LC}} = \int_0^\infty d\rho_\ell \exp[-\beta(M-1)\mathcal{F}_{\text{eff}}(\rho_\ell)], \quad (2.157)$$

où

$$\mathcal{F}_{\text{eff}}(\rho_\ell) = \rho_\ell J' - T S_{\text{P}}(\rho_\ell) \quad (2.158)$$

est l'énergie libre effective du modèle de *looping* et *clustering* et $J' = J + \ln \alpha_0$, l'énergie d'activation d'une boucle renormalisée par le coût dans l'entropie des boucles avec $\alpha_0 = l_p^{D\nu-1} (D\nu - 1)$ (voir Annexe 1). Dans la limite $M \rightarrow \infty$, l'approximation précédente de la fonction de partition \mathcal{Z}_{LC} devient exacte et

peut être évaluée exactement par une méthode de col en minimisant l'énergie libre effective \mathcal{F}_{eff} . La solution, ρ_ℓ^* , de l'équation d'optimisation, $d\mathcal{F}_{\text{eff}}(\rho_\ell)/d\rho_\ell = 0$, est

$$\rho_\ell^* = \frac{1}{1 + e^{\beta J'}}. \quad (2.159)$$

La contribution entropique de \mathcal{F}_{eff} s'annule pour $\rho_\ell = 0$ et $\rho_\ell = 1$, et atteint son minimum pour $\rho_\ell = 1/2$, qui est le résultat exact pour ρ_ℓ^* pour une énergie renormalisée d'activation des boucles J' nulle car l'entropie de mélange S_P est alors maximisée. Pour $J' > 0$, ρ_ℓ^* décroît de $1/2$ jusqu'à 0 dans la limite $J' \rightarrow \infty$ avec $\rho_\ell^* \propto e^{-\beta J'}$. Notre méthode de col mène alors à :

$$\langle n \rangle^* = \frac{M - 1}{1 + e^{\beta J'}}, \quad (2.160)$$

qui s'avère être un résultat exact, grâce à des erreurs qui se compensent, pour le nombre moyen de boucle $\langle n \rangle$ pour un nombre fini de particules M . Nous obtiendrons cette expression dans la partie suivante en dérivant la fonction de partition \mathcal{Z}_{LC} par rapport à J . La compétition entre le coût d'énergie pour activer une boucle, J' , et le gain en entropie de mélange S_P détermine le nombre moyen de boucles, indépendant de la longueur de la chaîne à condition que la contrainte imposée par des longueurs de chaîne soit négligée dans nos intégrales sur les longueurs individuelles de boucles en prenant la limite $l_{\text{max}} \rightarrow \infty$. Pour $l_p = 10l_m$, $D = 3$, et $\nu = 1/2$ (chaîne idéale), $\alpha_0 = 1,581$ et $\ln(\alpha_0) = 0,4582$, qui représente une petite correction si $J > 1$.

2.5.2 Résultats

Le premier résultat illustre cette compétition. Nous pouvons en effet déterminer le nombre moyen de boucles par l'expression :

$$\langle n \rangle = \frac{1}{\mathcal{Z}_{\text{LC}}} \sum_{n=0}^{M-1} n \frac{(M-1)!}{(M-1-n)!n!} e^{-nJ} \left[\frac{l_0^{1-D\nu}}{D\nu-1} \right]^n \quad (2.161)$$

dans la limite où $l_{\text{max}} \rightarrow +\infty$. Nous trouvons ainsi :

$$\langle n \rangle = \frac{M - 1}{1 + e^{\beta J'}}. \quad (2.162)$$

Le nombre moyen de boucle dépend linéairement du nombre de protéines. Cette dépendance est corrigée par le terme $e^{\beta J'}$ au dénominateur. Plus le *spreading* est important, plus le nombre de boucle diminue : il y a bien compétition entre la formation des boucles et l'interaction de *spreading*. De la même manière, nous pouvons déterminer la distribution moyenne des protéines ParB à une distance

ou coordonnée génomique s de la séquence *parS*. Elle se calcule numériquement à partir de l'expression :

$$P_{\text{LC}}(s) = \frac{1}{\mathcal{Z}_{\text{LC}}} \sum_{n=0}^{M-1} \frac{(M-1)!}{(M-1-n)!n!} \exp(-n\beta J) \times \prod_{j=1}^n \int_{l_p}^{l_\infty} dl_j l_j^{-D\nu} P_n(s, \{l_i\}) \quad (2.163)$$

où $P_n(s, \{l_i\})$ est la distribution des protéines ParB pour un ensemble de longueurs de boucles $\{l_i\}$. Cette quantité est remplacé dans les faits par la probabilité $P_1(s, l)$ à une boucle sachant que s ne fait pas partie d'une boucle sortant du complexe ParBS. La longueur l de la seule boucle fictive est alors interprétée comme la longueur cumulée des n boucles réelles. Cette approximation a été confrontée à des résultats extraits de simulations Monte-Carlo (voir figure 2.20). Nous pouvons en effet définir un hamiltonien canonique du

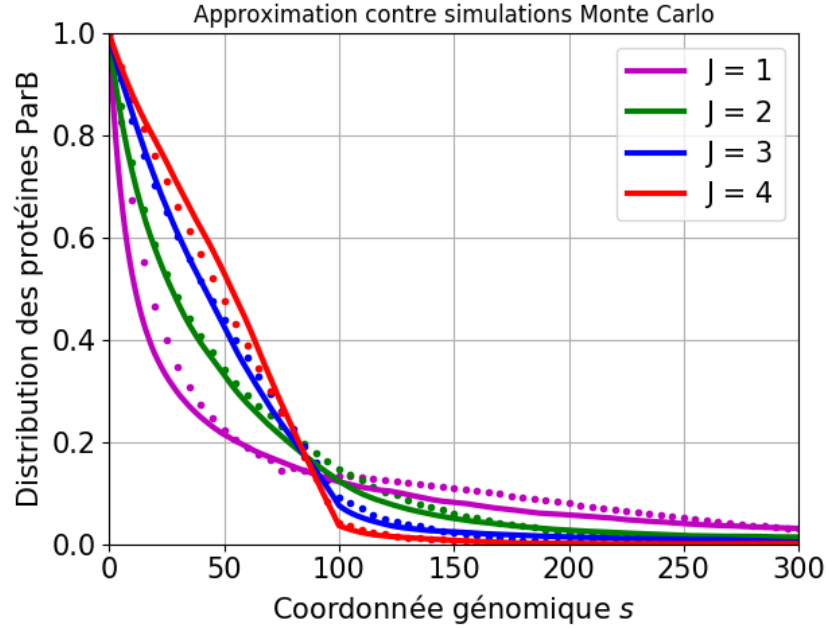


FIGURE 2.20 – Test de l'approximation utilisée avec des simulations Monte Carlo : pour plusieurs valeurs d'intensité des interactions de *spreading* J et $M = 100$. En trait plein, les profils théoriques donnés par (2.163). Les points indiquent dans la même couleur les données de simulations Monte Carlo basées sur l'hamiltonien (2.164).

looping et *clustering* à partir de la fonction de partition \mathcal{Z}_{LC} :

$$\mathcal{H}_{\text{LC}} = -J \sum_{i=1}^{N-1} \Phi_i \Phi_{i+1} + D\nu \sum_{i=1}^n \ln \left(\frac{l_i + l_p}{l_m} \right) \quad (2.164)$$

où Φ_i est l'occupation du site i . La première partie de l'hamiltonien prend en compte les interactions de spreading entre premiers voisins le long de la chaîne tandis que la seconde compte l'entropie des boucles. En figure 2.21, les paramètres utilisés sont $N = 4000$ sites, $M = 100$ protéines et $l_p = 10 * l_m = 160pb$ tandis que J peut varier. L'accord entre l'approximation précédemment mentionnée et les simulations Monte-Carlo est correct.

En figure 2.21, nous représentons la distribution moyenne des protéines ParB à la coordonnée génomique s pour $M = 400$ protéines⁵ et plusieurs valeurs de J . L'étendu de ces distributions dépend du nombre de protéines M

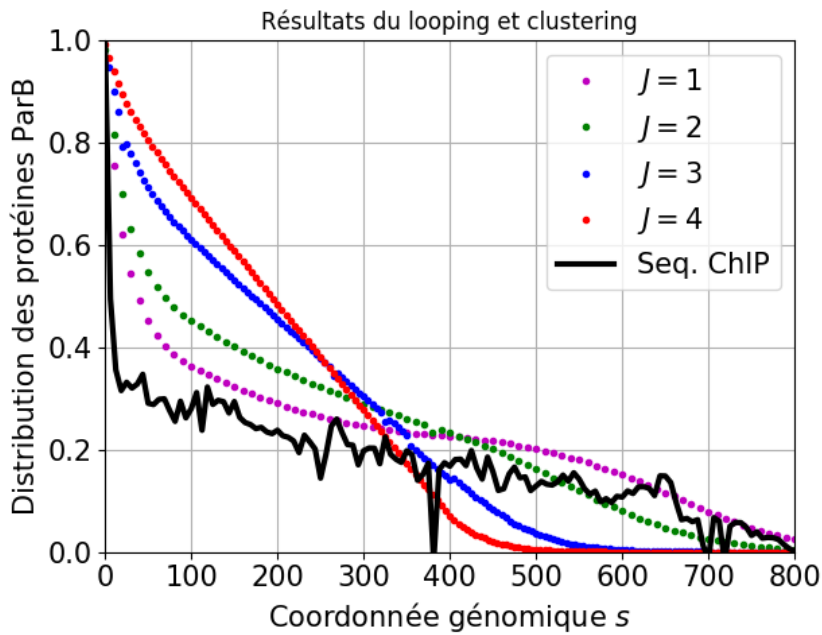


FIGURE 2.21 – **Résultats du looping et clustering** : distribution des protéines ParB autour des séquences *parS* pour différentes valeurs de J et $M = 400$ protéines. Le modèle *looping* et *clustering* reproduit correctement les données du séquençage ChIP en noir pour $J = 1$.

sur le réseau et du nombre moyen de boucle $\langle n \rangle$. Nous représentons aussi les données biologiques obtenue par séquençage ChIP. Pour de grandes valeurs de J , nous obtenons des profils triangulaires dont la base correspond à peu près à M : si l'interaction de *spreading* est forte, les protéines sont en moyenne concentrée de manière compacte autour des séquences *parS* dans un seul domaine qui peut glisser. Néanmoins, c'est le profil théorique obtenu avec $J = 1k_B T$ qui reproduit correctement le comportement de la courbe expérimentale. Dans le régime où J est comparable à $k_B T$, les liaisons de *spreading* entre protéines ParB peuvent être facilement rompues. Ainsi, ces résultats indiquent qu'un

⁵. Des estimations plus récentes estiment le nombre de protéines ParB par complexes ParBS à 400 protéines.

complexe ParBS est structurellement relâché et permet à l'ADN de former des boucles n'importe où sur la chaîne, et pas uniquement à la surface du complexe. Deux autres modèles ont permis de prédire des distributions de protéines sur l'ADN dans le cas des complexes ParBS. Dans le cadre du *spreading* et *bridging* présenté dans cette section, des simulations numériques ont montré que les protéines forment des profils triangulaires autour des séquences *parS* dans le régime des J forts [55]. Au contraire, dans le cas $J = 0$, le modèle de *stochastic binding*, qui considère que les protéines ParB se fixent de manière stochastique et dynamique dans une région déterminée autour d'une séquence *parS*, permet de capturer le comportement à longue distance des séquences *parS* de la distribution expérimentale des protéines ParB. Dans les figure 2.21 et 2.20, pour $M = 100$ protéines, nous retrouvons les profils triangulaires pour des $J \sim 3 - 4k_B T$ et un comportement algébrique du type

$$P_{LC} \sim s^{-D\nu} \quad (2.165)$$

quand $J \sim 1k_B T$, comportement asymptotique attendu dans le *stochastic binding* [22]. De manière intéressante, le looping et clustering apparaît comme une approche intermédiaire entre ces deux limites.

2.6 Conclusion

Dans ce chapitre, nous avons démontré que les protéines ParB pouvaient subir une séparation de phase entre états de haute et basse occupation sur l'espace unidimensionnel de l'ADN. La distribution moyenne des protéines ParB sur l'ADN révélée par les expériences de séquençage ChIP est cohérente avec ce mécanisme. Cette séparation de phase, provoquée par les interactions de *bridging*, aboutissent à la structure tridimensionnel des complexes ParBS tels qu'observés dans les expériences de SPT-PALM. Nous avons montré que le régime thermodynamique de ce système biologique était le régime de coexistence métastable qui est le seul régime capable d'expliquer l'activation des complexes ParBS en présence des séquences *parS*. Les estimations basées sur le modèle de *looping* et *clustering*, corroborées par les diagrammes de phase du modèle de gaz sur réseau à longue portée, permettent d'estimer l'intensité des interactions de *spreading* autour de la valeur de $1k_B T$.

D'une manière générale, notre modèle de gaz sur réseau à longue portée et notre cadre de travail du *spreading* et *bridging* pourraient posséder des applications dans d'autres problématiques biologiques, par exemple :

1. pour des systèmes très similaires au système de partition ParABS, tel que le système d'auto-organisation du nucléoïde PomXYZ qui intervient dans la division cellulaire de *Myxococcus xanthus* [23] ;
2. pour étudier les repliements des protéines ;

3. pour étudier l'organisation de la chromatine chez les cellules eucaryotes.

Une perspective immédiate est de déterminer le profil d'occupation des protéines ParB sur l'ADN en minimisant l'énergie libre calculée pour le modèle LRLG et de le comparer aux résultats du séquençage ChIP et du *looping* et *clustering*. L'approche complémentaire à notre développement en *cluster* qui consiste à déterminer la conformation du polymère ADN en fonction d'une séquence donnée de protéines adsorbées permettrait d'affiner nos diagrammes de phase. La fonction de distribution du polymère doit en toute rigueur dépendre de l'occupation moyenne sur la chaîne, brisant ainsi la symétrie particule-trou. Enfin, notre formulation du *spreading* et *bridging* pourrait permettre de donner des fondations rigoureuses aux approches de type Flory pour les polymères et les ligands, et apporter, grâce à la méthode variationnelle utilisée, un éclairage intéressant sur les polymères magnétiques [108] et les polyélectrolytes [109].

Dans le prochain chapitre, nous nous appliquons à ajouter la variable temporelle dans notre description au travers des expériences de photoblanchiment sur les complexes ParBS et de leur positionnement observé dans les expériences 3D-SIM.

2.7 Annexe 1 : estimation de c_0 pour un polymère auto-évitant

Nous simplifions notre analyse en imaginant que le polymère vit sur un réseau cubique simple de maille l_m , la longueur d'un monomère, à D dimensions. Si $Z(\mathbf{R})$ représente le nombre de polymères distincts de N monomères avec une distance bout à bout \mathbf{R} et $Z_{\text{tot}} = \sum_{\mathbf{R}'} Z(\mathbf{R}')$, le nombre total de configurations distinctes du polymère, alors la probabilité que ce polymère possède une distance bout à bout de \mathbf{R} est donnée par :

$$p(\mathbf{R}) = \frac{Z(\mathbf{R})}{Z_{\text{tot}}} \quad (2.166)$$

avec $\sum_{\mathbf{R}'} p(\mathbf{R}') = 1$ [97, 103]. Dans la limite continue, nous pouvons remplacer la somme de Z_{tot} par une intégrale $l_m^{-3} \int d\mathbf{R}'$, et la probabilité $p(\mathbf{R})$ par la fonction de distribution du polymère :

$$P(\mathbf{R}) = l_m^{-3} p(\mathbf{R}) = \frac{Z(\mathbf{R})}{\int d\mathbf{R}' Z(\mathbf{R}')}. \quad (2.167)$$

Ainsi, par définition :

$$l_m^3 P(\mathbf{R}) = \frac{Z(\mathbf{R})}{Z_{\text{tot}}}, \quad (2.168)$$

où dans la limite continue $Z_{\text{tot}} = l_m^{-3} \int d\mathbf{R}' Z(\mathbf{R}')$. A $D = 3$, la fonction de distribution du polymère respecte la relation d'échelle [97, 103] :

$$P(R) = R_{\text{ee}}^{-3} f_p(R/R_{\text{ee}}), \quad (2.169)$$

avec comme normalisation unitaire $\int d\mathbf{R} P(\mathbf{R}) = 1$, et où $R_{\text{ee}}(N) = \sqrt{\langle \mathbf{R}_{\text{ee}}^2 \rangle}$ est la racine de la distance bout à bout carrée moyenne. Le comportement asymptotique de la fonction de distribution du polymère est ainsi gouverné par trois exposants principaux, ν , g , et δ avec $R_{\text{ee}} \sim aN^\nu$ pour de grandes valeurs de N , $g = (\gamma - 1)/\nu$, et $\delta = (1 - \nu)^{-1}$:

$$f_p(x) = \begin{cases} c_0 x^g, & \text{for } x \rightarrow 0 \\ c_1 x^\omega \exp(-c_2 x^\delta), & \text{for } x \rightarrow \infty \end{cases} \quad (2.170)$$

où c_0, c_1, c_2 sont des constantes et $\gamma - 1$ est l'exposant qui gouverne le facteur d'amélioration qui apparaît dans la fonction de distribution d'un polymère auto-évitant (Self-Avoiding Polymer, SAP). Nous définissons le coût d'énergie libre pour former une boucle en fixant une extrémité du polymère à l'origine des coordonnées et en calculant la probabilité que l'autre extrémité se trouve dans une sphère de rayon l_m centrée sur la première extrémité :

$$\Delta \mathcal{F}_{\text{loop}} = -k_B T \ln \left[4\pi \int_0^{l_m} dR R^2 P(R) \right]. \quad (2.171)$$

2.7. ANNEXE 1 : ESTIMATION DE C_0 POUR UN POLYMÈRE
AUTO-ÉVITANT

Le cas d'un polymère gaussien peut nous servir d'illustration. La fonction de distribution de la distance bout à bout s'écrit dans le cas d'une chaîne idéale à trois dimensions :

$$P_N^G(R) = \left(\frac{3}{2\pi [R_{ee}^G]^2} \right)^{3/2} \exp \left[-\frac{3}{2} \left(\frac{R}{R_{ee}^G} \right)^2 \right], \quad (2.172)$$

avec $R_{ee}^G \sim aN^{1/2}$, c'est-à-dire que $\nu_G = 1/2$, $g_G = \omega_G = 0$, et $\delta_G = 2$. Dans ce cas :

$$\begin{aligned} \beta \Delta \mathcal{F}_{\text{loop}}^G &\approx -\ln \left[\frac{4\pi l_m^3}{3} P_N^G(0) \right] \\ &= -\ln \left[\frac{4\pi}{3} \left(\frac{3}{2\pi} \right)^{3/2} \right] + 3\nu_G \ln N. \end{aligned} \quad (2.173)$$

Cette dernière approximation est vraie quand $l_m/R_{ee}^G(N) \ll 1$.

Un polymère auto-évitant à $D = 3$ possède une distance bout à bout de la forme $R_{ee}^{\text{SA}} \sim a|i - j|^{3/5}$, c'est-à-dire que $\nu_{\text{SAP}} \approx 0,588 < \nu_{\text{Flory}} = 3/5 = 0,6$, $g_{\text{SAP}} \approx 0,27$ ($\gamma_{\text{SAP}} \approx 1,16$), et $\delta_{\text{SAP}} \approx 5/2$. En utilisant la forme asymptotique de la relation d'échelle contenant f_p pour de petits arguments, (2.171) mène dans ce cas à :

$$\beta \Delta \mathcal{F}_{\text{loop}}^{\text{SA}} = -\ln \left[\frac{4\pi c_0^{\text{SA}}}{3 + g_{\text{SA}}} \right] + 3\nu_{\text{SA}}(1 + g_{\text{SA}}/3) \ln N. \quad (2.174)$$

Pour estimer la constante c_0^{SA} pour un polymère auto-évitant, nous notons que le rapport du nombre Z_c de polymère auto-évitant formant une chaîne fermée et du nombre total de configurations Z_{tot} est donné par :

$$\frac{Z_c}{Z_{\text{tot}}} = \sum_{\text{nn}0'} p(\mathbf{R}) = (C - 1)p(l_m) = (C - 1)l_m^3 P(l_m) \quad (2.175)$$

dans le cas isotrope et où C est le nombre de coordination du réseau. La somme $\sum_{\text{nn}0'}$ est sur tous les sites du réseau qui sont premiers voisins (nearest-neighbors, nn) avec l'origine à l'exception du site lié au premier monomère. Pour un réseau cubique simple, on a $C - 1 = 5$. Le coût d'énergie libre pour créer un polymère fermé, qui correspond à une boucle, est alors :

$$\beta \Delta \mathcal{F}_c = -\ln \left[\frac{Z_c}{Z_{\text{tot}}} \right] \quad (2.176)$$

En utilisant la forme asymptotique pour de large N à trois dimensions $Z_{\text{tot}} \simeq C_{\text{tot}} \bar{\mu}^N N^{\gamma_{\text{SA}} - 1}$ et $Z_c \simeq C_c \bar{\mu}^N N^{-3\nu_{\text{SA}}}$ [97, 98, 103, 110], avec $\bar{\mu} \approx 4,68$ est le

nombre de coordination effectif d'un polymère auto-évitant sur un réseau, nous trouvons :

$$\beta \Delta \mathcal{F}_c^{\text{SA}} = -\ln \left[\frac{Z_c}{Z_{\text{tot}}} \right] = -\ln [C_c/C_{\text{tot}}] + 3\nu_{\text{SA}}(1 + g_{\text{SA}}/3) \ln N. \quad (2.177)$$

La quantité $\bar{\mu}$ prend simplement en compte le nombre effectif de sites disponibles pour construire une marche auto-évitante, nombre réduit à cause des sites premiers voisins déjà occupés. Elle est proche de $C - 1 = 5$ pour un réseau cubique simple. En utilisant des données issues d'énumérations exactes pour $N = 22$ [110], nous trouvons $C_c = 0,543$ et $C_{\text{tot}} = 0,2$. Il n'est pas clair jusqu'à quel point nous pouvons faire confiance à ces résultats car [98] trouve $C_{\text{tot}} = 1,2$ en utilisant des énumérations exactes jusqu'à $N = 36$, valeur à comparer à celle donnée précédemment de $0,2$ (aucune valeur de C_c n'est hélas fournie). Si C_{tot} et C_c évolue de la même manière pour N croissant alors nous pouvons espérer que la valeur du rapport C_c/C_{tot} obtenue pour $N = 22$ est proche de la valeur, inconnue, qu'on obtiendrait pour $N = 36$. En posant l'égalité des équations (2.174) pour $\Delta \mathcal{F}_{\text{loop}}^{\text{SA}}$ et (2.177) pour $\Delta \mathcal{F}_c^{\text{SA}}$, nous pouvons obtenir une estimation de la constante c_0^{SAP} :

$$c_0^{\text{SA}} = \frac{(3 + g_{\text{SA}}) C_c}{4\pi C_{\text{tot}}} = 0,7065. \quad (2.178)$$

En utilisation directement l'équation (2.175), nous pouvons trouver :

$$c_0^{\text{SA}} = \frac{1}{C - 1} \frac{C_c}{C_{\text{tot}}} = 0,543, \quad (2.179)$$

où $C = 6$ (réseau cubique simple). Une meilleure estimation pourrait vraisemblablement être obtenue en remplaçant $C - 1$ dans la dernière expression par le nombre effectif de coordination $\bar{\mu} = 4,68$:

$$c_0^{\text{SA}} = \frac{1}{\bar{\mu}} \frac{C_c}{C_{\text{tot}}} = 0,58. \quad (2.180)$$

Cette dernière valeur, $c_0^{\text{SA}} = 0,58$, est probablement notre meilleure estimation pour le moment et doit ainsi être utilisée pour établir le diagramme de phase pour des particules sur un polymère auto-évitant.

2.8 Annexe 2 : exposants critiques du modèle IRLG

Les exposants critique de modèle d'Ising ou le modèle de gaz sur réseau avec interactions à courte portée, traités dans l'approximation du champ moyen

sont connus. Ils définissent un ensemble de valeurs dites classiques que nous rappelons ci-dessous. Nous développons ici succinctement la méthode pour trouver ces exposants dans le cadre du modèle de gaz sur réseau avec interactions à portée infinie résolu avec notre méthode variationnelle. Nous rappelons que cette méthode est exacte pour le modèle IRLG dans la limite thermodynamique. La première étape consiste à écrire le potentiel grand canonique comme la transformée de Legendre de l'énergie libre de Helmholtz :

$$\omega_{\text{IRLG}}(\Phi, \mu, T) = \frac{\Omega_{\text{IRLG}}}{N} = f_{\text{IRLG}}(\Phi, T) - \mu\Phi \quad (2.181)$$

Développer l'énergie libre de Helmholtz par site $f_{\text{IRLG}}(\Phi, T) = \mathcal{F}_{\text{IRLG}}/N$ autour du point critique d'occupation $\Phi_c = \frac{1}{2}$ permet de réécrire le potentiel grand canonique ω_{IRLG} en utilisant l'énergie libre de Landau correspondante $a(\Phi, T)$:

$$\omega_{\text{IRLG}}(\Phi, \mu, T) \approx a(\Delta\Phi, T) - C\Delta\Phi \quad (2.182)$$

avec $\Delta\phi = \Phi - \Phi_c$. L'énergie libre $a(\Delta\Phi, T)$ peut être exprimée comme :

$$a(\Delta\Phi, T) = a_0 + \frac{a_2}{2}(\Delta\Phi)^2 + \frac{a_4}{4}(\Delta\Phi)^4 \quad (2.183)$$

avec :

$$C = \mu - \left(\frac{\partial f}{\partial \Phi} \right)_{\Phi_c} = \mu - \mu(\Phi_c, T) \quad (2.184)$$

$$a_0 = f(\Phi_c) - C\Phi_c \quad (2.185)$$

$$a_2 = \left(\frac{\partial^2 f}{\partial \Phi^2} \right)_{\Phi_c} = -K + 4\sqrt{B}k_B T \quad (2.186)$$

$$a_4 = \frac{1}{6} \left(\frac{\partial^4 f}{\partial \Phi^4} \right)_{\Phi_c} = \frac{8}{3}k_B T \sqrt{B}(3 - B) \quad (2.187)$$

La théorie de Landau pour les transitions de phase explique que la fonction de partition intégrale [83] :

$$Z = \int_{-\frac{1}{2}}^{\frac{1}{2}} d(\Delta\Phi) \exp[-N\beta \omega(\Delta\Phi, \mu, T)] \quad (2.188)$$

peut être remplacé par le maximum de l'intégrand atteint pour $\Delta\Phi = \bar{\Delta\Phi}$, la valeur d'équilibre de $\Delta\Phi$. Il s'agit en fait d'une méthode de col autour de la valeur $\bar{\Delta\Phi}$, d'où :

$$\begin{cases} \left(\frac{\partial \omega}{\partial \Delta\Phi} \right)_{\bar{\Delta\Phi}} = 0 \\ \left(\frac{\partial^2 \omega}{\partial (\Delta\Phi)^2} \right)_{\bar{\Delta\Phi}} > 0 \end{cases} \quad (2.189)$$

Ces équations sont les critères de la théorie de Landau pour les transitions de phase : la première permet de trouver l'expression de $\Delta\bar{\Phi}$ et la seconde, la stabilité de cette solution. Dans notre cas, le précédent système devient :

$$\begin{cases} \Delta\bar{\Phi}(a_2 + a_4\Delta\bar{\Phi}) = 0 \\ a_2 + 3a_4\Delta\bar{\Phi}^2 > 0 \end{cases} \quad (2.190)$$

La solution $\Delta\bar{\Phi}_1 = 0$ est stable si et seulement si le terme $a_2(T)$ est positif (instable sinon) et la solution $\Delta\bar{\Phi}_2 = \sqrt{\frac{-a_2}{a_4}}$ existe et est stable si $a_2(T)$ est négatif puisque a_4 est nécessairement positif pour assurer la convergence de l'intégrale précédente. Le cas $a_2(T_c) = 0$ mène à l'équation vérifiée par la température spéciale qui coïncide avec le changement de signe de $a_2(T)$, la température critique T_c :

$$T_c \sqrt{\exp\left(-\frac{J}{k_B T_c}\right)} = T_c \sqrt{B_C} = \frac{K}{4k_B} \quad (2.191)$$

La manière la plus simple de rendre compte du comportement mathématique de $a_2(T)$ autour de T_c est de donner son approximation linéaire en fonction de $T - T_c$:

$$a_2(T) \approx 4k_B \sqrt{B_C} \left(1 + \frac{2J}{K} \sqrt{B_C}\right) (T - T_c) = \alpha_2 (T - T_c) \quad (2.192)$$

où α_2 est une constante. Le terme $a_4(T)$ ne nécessite aucun développement puisque son comportement est plus simple. Il prend donc sa valeur à la température critique :

$$a_4(T) \approx a_4(T_c) = \alpha_4 = \frac{8}{3} k_B T_c \sqrt{B_C} (2 - B_C) \quad (2.193)$$

Ainsi, l'occupation d'équilibre quand $T < T_c$ est décrite par la relation :

$$\Delta\bar{\Phi} = \left(\frac{\alpha_2 T_c}{\alpha_4}\right)^{\frac{1}{2}} \left(\frac{T_c - T}{T_c}\right)^{\frac{1}{2}} = \left(\frac{\alpha_2 T_c}{\alpha_4}\right)^{\frac{1}{2}} \left(\frac{T_c - T}{T_c}\right)^{\beta} \quad (2.194)$$

où $\beta = \frac{1}{2}$ est l'exposant critique lié à l'occupation moyenne dans la théorie de Landau. Pour trouver l'exposant critique γ de la compressibilité isotherme χ , nous avons besoin de calculer la dérivée :

$$\frac{1}{\chi} = \frac{\Phi^2}{V_0} \frac{\partial^2 f}{\partial \Phi^2} \approx \frac{\Phi^2}{V_0} [a_2 + 3a_4(\Phi - \Phi_c)] \quad (2.195)$$

en utilisant le développement de $f_{\text{IRLG}}(\Phi, T)$ avec la forme de Landau. Avec l'expression précédente de $a_2(T)$, nous trouvons pour les valeurs près de Φ_c and T_c :

$$\chi \approx \frac{V_0}{\alpha_2 \Phi^2} \frac{1}{T - T_c} = \frac{V_0}{\alpha_2 \Phi^2} \left(\frac{1}{T - T_c} \right)^\gamma. \quad (2.196)$$

Nous obtenons ainsi la valeur de $\gamma = 1$ et avec la relation d'échelle reliant les exposants critiques :

$$\alpha + 2\beta + \gamma = 2 \quad (2.197)$$

nous pouvons déduire l'exposant critique lié à la chaleur spécifique $\alpha = 0$. En conclusion, nous avons trouvé des exposants critiques classiques pour le modèle de gaz sur réseau avec interactions à courte et longue portées. Pour aller plus loin, le comportement critique du même système mais traité dans l'approximation du champ moyen peut être trouvé en développant les précédentes quantités pour de faibles valeurs de J . Ce développement mène à :

$$T_c = \frac{2J + K}{4k_B} \quad (2.198)$$

$$\alpha_2 = 4k_B \quad (2.199)$$

$$\alpha_4 = \frac{8J + 4K}{3} \quad (2.200)$$

La limite $K \rightarrow 0$ permet de retrouver la valeur de la température critique T_c et des coefficients α_4 pour un gaz sur réseau avec interactions à courte portée à une dimension (voir section Champ moyen et approche variationnelle, Chapitre 2) . Nous avons notamment utilisé le coefficient $\alpha_4 = 8J/3$ pour déterminer une expression analytique de l'énergie libre de Helmholtz au voisinage du point critique (voir Résultats du modèle à un champ, Chapitre 3).

Chapitre 3

Cinétique et dynamique

3.1 Introduction

Nous avons jusqu'à présent décrit les protéines ParB en interaction sur l'ADN pour expliquer la formation des complexes ParBS par une séparation de phase thermodynamique, c'est-à-dire comme le résultat d'un mécanisme passif. Dans ce chapitre, nous concentrons notre étude sur l'interprétation des expériences de recouvrement de la fluorescence des complexes ParBS photoblanchis (voir Chapitre 1) dans une démarche exploratoire qui vise initialement à proposer un protocole numérique pour reproduire *in silico* ces expériences. Les résultats des expériences de FRAP apportent des informations essentielles à la compréhension du fonctionnement du système de partition ParABS dans sa totalité. Nous exposons nos propositions de la modélisation la plus phénoménologique à la plus fondamentale. Le premier niveau de lecture montre que les résultats des expériences de photoblanchiment ne sont pas anodins et qu'ils donnent des indices sur la nature des complexes ParBS : il s'agit d'un modèle de comptage de protéines permettant d'extraire des taux effectifs d'entrée et de sortie des complexes ParBS que nous comparons au taux attendu pour un mécanisme de pure diffusion. Le second niveau de modélisation, semi-phénoménologique, basé sur une équation de Smoluchowski, nous donne l'opportunité de développer une méthode de photoblanchiment *in silico*. Ce niveau contient peu de paramètres et rend possible une étude simple des courbes des récupérations obtenues *in silico* en fonction des coefficients de diffusion à l'intérieur et à l'extérieur des complexes ParBS. Le dernier niveau de modélisation se veut plus fondamental. Il s'agit d'une approche de gaz sur réseau à plusieurs espèces dans la limite continue où tous les paramètres, notamment les énergies d'interaction entre protéines sur l'ADN, sont bien définis. Cette approche est similaire aux approches fonctionnelles de théorie des champs utilisées par exemple pour étudier les séparations actives de phase induite par la motilité [111, 112]. Ce modèle général est étudié en deux dimensions afin d'alléger

fortement les calculs numériques. Nous montrons d’abord que l’activité des protéines ATPase ParA est nécessaire pour interpréter complètement les expériences de photoblanchiment car les complexes ParBS subissent un processus de maturation d’Ostwald. Nous montrons ainsi que le système ParABS est déjà un système actif pour le maintien de plusieurs complexes ParBS tels qu’ils apparaissent dans les expériences de SPT-PALM (voir Introduction) : ils existent hors de l’équilibre thermodynamique grâce à un mécanisme d’inversion de la maturation d’Ostwald des complexes ParBS [113, 114]. Le système ParABS s’inscrit ainsi dans le domaine de la matière molle active [115, 116, 117]. Nous concluons ce chapitre en étudiant le rôle de la protéine ParA dans la dynamique du système de partition, c’est-à-dire le positionnement des complexes ParBS. Dans le cadre de leur ségrégation et de leur positionnement (voir expérience de 3D-SIM dans l’Introduction), les complexes ParBS, de rayon $R = 18,5$ nm [63], peuvent être vus comme des particules colloïdales poreuses [118]. Cette dénomination rapproche le phénomène du positionnement des complexes ParBS de la physique des particules autopulsées [119, 120].

3.2 Approche phénoménologique

Nous assimilons l’espace intracellulaire au nucléoïde, c’est-à-dire le volume occupé par l’ADN dans le cytoplasme. Les transferts de protéines durant les expériences de photoblanchiment peuvent être décrits par un simple modèle de comptage de protéines (voir figure 3.1). Nous distinguons deux foyers :

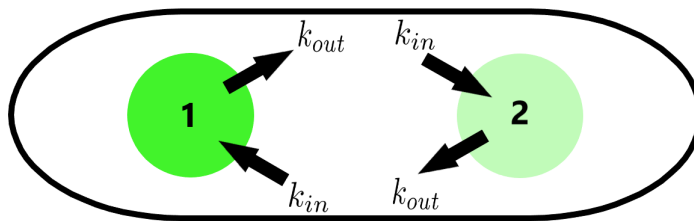


FIGURE 3.1 – **Représentation schématique du modèle de comptage.** Cette approche phénoménologique discrimine trois sous-espaces cytoplasmique : le complexe 1, à gauche, qu’on considère laissé intacte par le photoblanchiment : le complexe 2, de droite, partiellement photoblanchi; et le reste du volume cytoplasmique. Les complexes ParBS sont équilibrés et peuvent échanger des protéines avec le troisième milieu. Une protéine ParB peut sortir d’un complexe avec un taux k_{out} et peut y pénétrer avec un taux k_{in} .

le complexe 1 (gauche) reste intact tandis que le complexe 2 (droite) subit l'opération de photoblanchiment.

3.2.1 Modèle de comptage des protéines

Soient $S_1(t)$ et $S_2(t)$, les ratios moyens des protéines ParB encore fluorescentes dans les foyers 1 et 2, et soit $\sigma(t)$, le ratio moyen de ces protéines dans le reste du cytoplasme après photoblanchiment. Ces quantités sont normalisées par le ratio initial de protéines dans le foyer 1. Nous notons $k_{i,out}$ ($k_{i,in}$), la probabilité par unité de temps qu'une protéine ParB quitte (prénètre dans) le foyer i . L'évolution temporelle de ces ratios est donnée par les équations de bilan de matière suivantes :

$$\begin{aligned}\frac{dS_1}{dt} &= k_{1,in}\sigma - k_{1,out}S_1 \\ \frac{d\sigma}{dt} &= -(k_{1,in} + k_{2,in})\sigma + (k_{1,out} + k_{2,out})(S_1 + S_2) \\ \frac{dS_2}{dt} &= k_{2,in}\sigma - k_{2,out}S_2\end{aligned}\tag{3.1}$$

Ces équations régissent les échanges en protéines entre les deux foyers et l'espace cytoplasmique restant. Par conservation de la quantité de protéines ParB, on a :

$$\frac{dS_1}{dt} + \frac{d\sigma}{dt} + \frac{dS_2}{dt} = 0.\tag{3.2}$$

Vu les symétries engagées dans la ségrégation du matériel génétique, nous considérons que les deux complexes sont deux exemplaires du même objet et ainsi $k_{1,in} = k_{2,in} = k_{in}$ et $k_{1,out} = k_{2,out} = k_{out}$. Le système (3.1) devient alors :

$$\begin{aligned}\frac{dS_1}{dt} &= k_{in}\sigma - k_{out}S_1 \\ \frac{d\sigma}{dt} &= -2k_{in}\sigma + k_{out}(S_1 + S_2) \\ \frac{dS_2}{dt} &= k_{in}\sigma - k_{out}S_2\end{aligned}\tag{3.3}$$

Pour comparer directement nos résultats avec les données biologiques, nous faisons trois suppositions importantes :

1. l'intensité de la fluorescence est proportionnelle au nombre de protéines ParB, ainsi nous pouvons directement comparer l'intensité lumineuse dans le foyer 1, normalisée par son intensité initiale, et $S_1(t)$;
2. la dégradation spontanée des fluorophores est négligeable ;
3. il n'y a pas d'exclusion stérique entre les protéines ParB, nous ne tenons pas compte explicitement de l'agencement intrinsèque des protéines à

l'intérieur du foyer : il s'agit d'un modèle à zéro dimension où tous les effets physiques sont codés dans les coefficients phénoménologiques k_{in} et k_{out} .

3.2.2 Extraire k_{in} et k_{out} des données de FRAP

Les taux k_{in} et k_{out} peuvent être extraits des courbes de fluorescence en résolvant le système d'équations (3.3). Le photoblanchiment est partiel : seule une fraction des protéines ParB de la zone excitée est réellement photoblanchie. Dans la figure 1.3, le comportement de ces courbes pour $t \rightarrow \infty$ indique que la valeur stationnaire du ratio des protéines à l'intérieur d'un foyer est $S_1(t \rightarrow \infty) = S_2(t \rightarrow \infty) = S_\infty \approx 0,7$. Dans cette limite, nous pouvons aussi montrer que le rapport des ratios stationnaires est simplement :

$$\frac{S_\infty}{\sigma_\infty} = \frac{k_{in}}{k_{out}}. \quad (3.4)$$

Ce dernier résultat peut être mis en relation avec les résultats des expériences de SPT-PALM qui ont permis d'estimer le pourcentage de protéines à l'intérieur et à l'extérieur des complexes. L'estimation donne un rapport $2S_\infty/\sigma_\infty \approx 0,9$ soit $k_{in}/k_{out} = 4,5$. L'exploitation des données du photoblanchiment doit mener vers une conclusion similaire. La conservation du ratio total $S_{tot} = S_1(t) + S_2(t) + \sigma(t)$ permet d'écrire d'une part que

$$S_{tot} = \left(2 + \frac{k_{out}}{k_{in}}\right) S_\infty. \quad (3.5)$$

et d'autre part, en éliminant l'équation pour la variable $\sigma(t)$, de réduire le système (3.3) aux deux équations :

$$\begin{aligned} \frac{dS_1}{dt} &= k_{in}S_{tot} - (k_{in} + k_{out})S_1 - k_{in}S_2 \\ \frac{dS_2}{dt} &= k_{in}S_{tot} - (k_{in} + k_{out})S_2 - k_{in}S_1. \end{aligned} \quad (3.6)$$

Ces deux équations montrent une symétrie $1 \leftrightarrow 2$. On en déduit simplement l'équation vérifiée par la différence $S_-(t) = S_2(t) - S_1(t)$:

$$\frac{dS_-}{dt} = -k_{out}S_-(t) \quad (3.7)$$

L'ajustement d'une fonction exponentielle permet d'extraire la valeur du paramètre k_{out} : on trouve $k_{out} = (0,60176 \pm 0,021) \text{min}^{-1} = (0,01003 \pm 0,00035) \text{s}^{-1}$. Nous pourrions obtenir la valeur de k_{in} à partir de l'équation vérifiée par la

somme $S_+(t) = S_1(t) + S_2(t)$. Hélas, cette variable ne montre pas de dépendance temporelle directement exploitable. Les formes analytiques de $S_+(t)$ et $S_-(t)$ nous permettent néanmoins d'écrire :

$$\begin{aligned} S_1(t) &= S_\infty - \left[\left(S_\infty - \frac{1}{2}S_+(0) \right) e^{-2k_{in}t} + \frac{1}{2}S_-(0) \right] e^{-k_{out}t} \\ S_2(t) &= S_\infty - \left[\left(S_\infty - \frac{1}{2}S_+(0) \right) e^{-2k_{in}t} - \frac{1}{2}S_-(0) \right] e^{-k_{out}t}. \end{aligned} \quad (3.8)$$

où $S_-(0)$ est déterminée par notre première régression. C'est la partie droite de l'équation

$$(S_2(t) - S_\infty)e^{k_{out}t} = - \left[\left(S_\infty - \frac{1}{2}S_+(0) \right) e^{-2k_{in}t} - \frac{1}{2}S_-(0) \right] \quad (3.9)$$

qui nous sert à obtenir la valeur de k_{in} . Cette régression est moins précise que la première et nous trouvons $k_{in} = (2,61966 \pm 0,7068)min^{-1} = (0,043661 \pm 0,01178)s^{-1}$. Le rapport k_{in}/k_{out} est alors de :

$$\frac{k_{in}}{k_{out}} = 4,353 \pm 1,306, \quad (3.10)$$

mesure cohérente avec l'estimation de 4,5 des expériences de SPT-PALM. Nous confrontons ces ajustements avec les données expérimentales dans la figure 3.2.

3.2.3 Temps de résidence typiques à l'intérieur et à l'extérieur des complexes

Les mesures de k_{in} et k_{out} nous permettent d'estimer les temps typiques d'existence des protéines ParB à l'intérieur τ_{out} et à l'extérieur du foyer τ_{in} . Nous avons en effet :

$$\begin{aligned} \tau_{in} &= k_{in}^{-1} \simeq 22.9 \text{ s} \\ \tau_{out} &= k_{out}^{-1} \simeq 99.7 \text{ s}. \end{aligned} \quad (3.11)$$

Ces temps doivent être comparés au temps typique donné par la diffusion d'une protéine ParB d'un complexe à l'autre. Pour une distance radiale approximative entre complexes $l_c = 0.5\mu m$ et un coefficient de diffusion des protéines ParB dans le cytoplasme de l'ordre de $D_V^B = 0.5\mu m^2/s$ [63], nous obtenons un temps typique à trois dimensions $\tau_{diff} = l_c^2/(6D_V^B) \approx 0.08s$, dominé par les deux temps typiques déduits des expériences de photoblanchiment et donc en désaccord avec les taux cinétiques k_{in} et k_{out} . Ce résultat indique qu'un

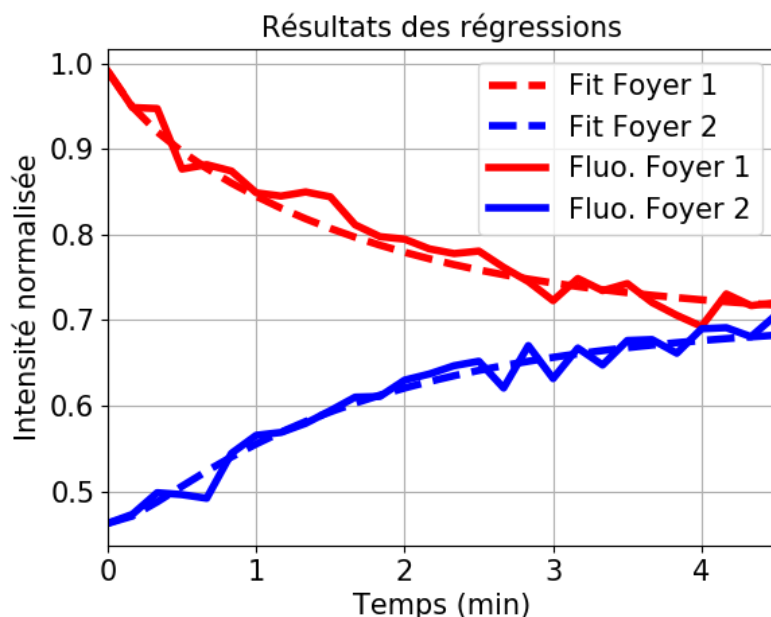


FIGURE 3.2 – **Résultats des ajustements** : Intensités lumineuses des complexes photoblanchi (en bleu) et non-photoblanchi (en rouge) en traits pleins. Mêmes couleurs pour les courbes de tendance issues du modèle de comptage en traits discontinus.

mécanisme simple de diffusion ne suffit pas à expliquer les temps de recouvrement ou de perte de fluorescence de plusieurs minutes observés dans le cas des protéines ParB, et que d'autres mécanismes doivent être pris en compte. Nous développons par la suite deux scénarios qui peuvent expliquer ces différences.

1. Le premier est la structure tridimensionnelle des protéines ParB. Les séquences d'acide animés permettant les interactions entre protéines se situent sur des motifs non repliés des ParB. Ces motifs sont d'une longueur linéaire approximativement égale à 56 nm (voir Chapitre 2). Ainsi, il faudrait imaginer un complexe ParB S comme un objet auto-assemblé très poreux semblable à une toile d'araignée ou une cage (*nucleation* et *caging*) qui complique l'entrée ou la sortie des protéines ParB. Dans le scénario de séparation de phase, les protéines ParB montrent des coefficients de diffusion collectif en accord avec ce schéma (voir Approche fondamentale ci-dessous) et les données expérimentales. Des expériences de molécules individuelles ont en effet permis de mettre en évidence la distribution du coefficient de diffusion des protéines ParB [63]. Il apparaît deux populations de protéines : une population lente des protéines en foyer, de coefficient de diffusion majoré, à cause d'une forte incertitude expérimentale, à $D_L^B \approx 0.02 \mu\text{m}^2/\text{s}$, et une population rapide des protéines, de coefficient de diffusion $D_V^B = 0,5 \mu\text{m}^2/\text{s}$, diffusant librement dans l'espace cytoplasmique.

2. Le second est l'existence de barrière d'énergie à la frontière des complexes. Les interactions entre protéines ParA et ParB pourraient en être à l'origine. Ces deux espèces possèdent une interaction attractive tandis que les protéines ParB catalysent l'hydrolyse de l'ATP par les protéines ParA : ces deux éléments mènent à la formation d'une ceinture de ParA autour d'un complexe ParBS (voir Approche fondamentale). Les barrières d'énergie pour entrer ΔE_{in} et sortir ΔE_{out} d'un complexe ParBS sont liées aux valeurs des taux k_{in} et k_{out} . On a en effet :

$$\Delta E_{\text{in}} = -k_B T \ln(k_{\text{in}}/k_{\text{in}}^0) \quad (3.12)$$

$$\Delta E_{\text{out}} = -k_B T \ln(k_{\text{out}}/k_{\text{out}}^0) \quad (3.13)$$

avec, dans le cas tridimensionnel, $k_{\text{in}}^0 \sim 6D_V^B/l_c^2$ et $k_{\text{out}}^0 \sim 6D_L^B/l_f^2$ où $l_f \simeq 0.037 \mu\text{m}$ est le diamètre d'un complexe ParBS. Ces expressions permettent d'obtenir $\Delta E_{\text{in}} \sim 5.616 k_B T$ et $\Delta E_{\text{out}} \sim 9.075 k_B T$.

Pour illustrer ces deux scénarios, nous proposons dans la prochaine partie une modélisation plus avancée de l'évolution de la concentration des protéines ParB fluorescentes durant une expérience de photoblanchiment.

3.3 Approche intermédiaire

Dans cette partie, nous explorons l'idée d'interpréter les données des expériences de photoblanchiment en intégrant une équation de Smoluchowski pour la concentration des protéines ParB avec un potentiel effectif $U(\vec{r})$ capturant ou excluant les protéines ParB dans l'espace. Ce nouvel élément pourrait changer la cinétique des protéines et mener à des temps de récupération de l'ordre de plusieurs minutes. L'origine du potentiel U pourrait être les interactions entre protéines ParB et entre protéines ParA et ParB : ces espèces possèdent en effet des interactions attractives et leurs distributions quasi-statiques pourraient jouer le rôle d'un potentiel effectif.

3.3.1 Modèle de diffusion avec un potentiel spatial

Dans cette approche, l'espace intracellulaire est décrit soit par un segment de longueur L pour le cas à une dimension ou par un objet d'une forme semblable à une bactérie, de longueur principale $L = 1\mu\text{m}$ et de largeur l à deux ou trois dimensions. La paramètre $l = 0,3\mu\text{m}$ est choisi de telle sorte que le L/l respecte visuellement le rapport d'aspect d'un bacille (voir figure 3.10 ci-dessous par exemple). Le système est centré en 0. Nous utilisons une équation de conservation de la matière pour décrire l'évolution spatio-temporelle de la concentration en protéine ParB fluorescente $\rho(x, t)$:

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{j} = 0 \quad (3.14)$$

pour laquelle le flux \vec{j} est défini par :

$$\vec{j} = -D(\rho) \left(\vec{\nabla}\rho + \rho\beta\vec{\nabla}U \right). \quad (3.15)$$

où $\beta = (k_B T)^{-1}$. Il s'agit d'une équation de Smoluchowski pour la concentration $\rho(x, t)$ avec un potentiel $U(\vec{r})$ [121]. Nous avons choisi des conditions de flux nul aux bords du système ($\vec{\nabla}j = \vec{0}$). Le coefficient de diffusion des protéines dépend de la densité locale. Pour deux complexes, il y a environ 600 protéines ParB et un complexe possède un rayon de $R = l_f/2 = 18.5$ nm [22, 63]. En supposant que le complexe est un segment de longueur $2R$ ou un disque ou une sphère de rayon R (selon la dimension), on peut estimer la concentration à l'intérieur ρ_l (état *liquide, dense* ou *haute densité*) avec un coefficient de diffusion D_L^B et à l'extérieur ρ_v du complexe (état *vapeur, dilué* ou *basse densité*) avec une diffusion D_V^B . La fonction $D(\rho)$ est alors défini comme la fonction suivant :

$$D(\rho) = \begin{cases} D_V^B, & \text{if } \rho < \rho_v \\ D_L^B, & \text{if } \rho > \rho_v \end{cases} \quad (3.16)$$

3.3.2 Photoblanchiment *in silico*

Ce modèle continu nous permet de simuler une expérience de photoblanchiment. Nous appelons photoblanchiment *in silico* une simulation numérique mimant une expérience de FRAP dans un système biologique. Concrètement, il s'agira d'intégrer une équation avec une condition initiale bien choisie. Nous détaillons d'abord notre procédé pour ensuite présenter et commenter les résultats de nos simulations.

Méthode

Le photoblanchiment *in silico* consiste en trois étapes :

1. **Initialisation** : deux copies du potentiel $\tilde{U} = \beta U$ sont placées aux positions $x = \pm 0,25\mu m$, correspondant aux positions 1/4 et 3/4 de la bactérie [54], pour créer et maintenir deux complexes ParBS bien localisés. L'obtention de la distribution des protéines dans cette configuration est simplifié par le fait que nous connaissons la solution stationnaire de l'équation de Smoluchowski. La solution stationnaire correspond à un flux nul dans le système, d'où

$$\rho_{ss}(\mathbf{r}) = \rho_0 \exp(-\tilde{U}(\mathbf{r})) \quad (3.17)$$

avec ρ_0 , concentration de normalisation donnée par :

$$M = \int_V \rho_{ss}(\mathbf{r}') d^3\mathbf{r}' \quad (3.18)$$

3.3. APPROCHE INTERMÉDIAIRE

où M est le nombre de protéines ParB. Aussi, ρ_0 peut être vue comme la concentration moyenne des protéines ParB dans l'espace intracellulaire, d'où $\rho_0 \approx 600\mu m^{-1}$ à une dimension par exemple [22]. Cette distribution montre deux maxima autour des positions $x = \pm 0,25\mu m$ (voir figure 3.3).

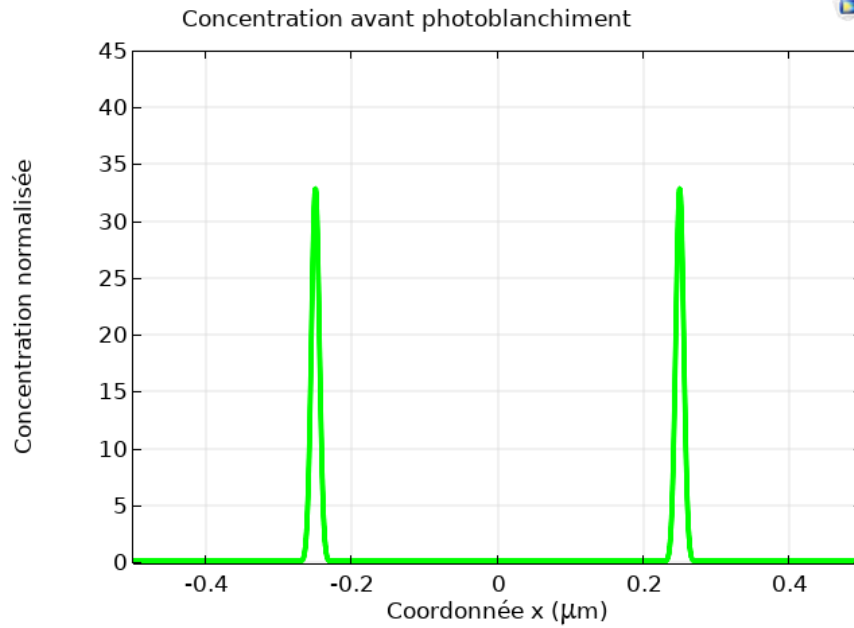


FIGURE 3.3 – **Profil stationnaire avant photoblanchiment** : le champ de concentration montre des maxima aux positions du potentiel \tilde{U} .

2. **Photoblanchiment** : Ce profil stationnaire est perturbé mais le photoblanchiment n'est pas parfait [63]. La partie droite, définie par $x > 0$, est multipliée par un taux de fluorescent rémanente $\gamma_F = 1 - \gamma_{PB}$ qui dépend simplement du taux de photoblanchiment γ_{PB} , tandis que la partie gauche ($x < 0$) est laissée intacte. D'après les données expérimentales (voir figure 1.3), nous avons un photoblanchiment de l'ordre de 60% dans le complexe photoblanchi, d'où $\gamma_{PB} = 0,6$ et $\gamma_F = 0,4$. Cette opération crée un profil photoblanchi et divise l'espace intracellulaire en trois sous-espaces : le complexe 1 pleinement fluorescent (F1), le complexe 2 photoblanchi (F2) et la phase basse densité (bulk). Pour une équation de Smoluchowski, la solution stationnaire $\rho_{ss}(\mathbf{r})$ étant connue analytiquement, le photoblanchiment aboutit à la fonction suivante :

$$\Phi_{ini}(\mathbf{r}) = \begin{cases} \rho_{ss}(\mathbf{r}) & \text{si } x \leq 0 \\ \gamma_F \rho_{ss}(\mathbf{r}), & \text{si } x > 0. \end{cases} \quad (3.19)$$

représentée en figure 3.4.

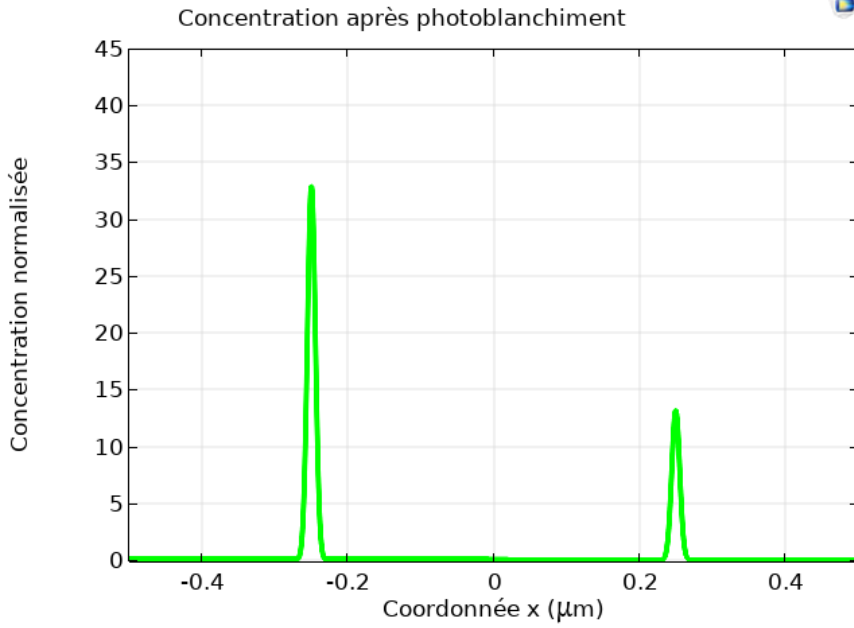


FIGURE 3.4 – **Profil stationnaire après photoblanchiment** : la concentration des protéines ParB est réduite de 60% dans la partie droite.

3. **Récupération** : Ce profil photoblanchi est utilisée comme distribution initiale d'une simulation avec les deux copies du potentiel \tilde{U} aux mêmes positions qu'à l'étape 1. Pour comparer nos résultats avec les données expérimentales, on calcule $S_{1/2}(t)$ et $\sigma(t)$, les proportions de protéines ParB dans le complexe 1 ou 2 et dans la phase basse densité, définies dans notre approche par :

$$S_{1(2)}(t) = S_{\text{tot}}^{-1} \int_{F_{1(2)}} \Phi(x, t) dx^D, \quad (3.20)$$

$$\sigma(t) = S_{\text{tot}}^{-1} \int_{\text{bulk}} \Phi(x, t) dx^D, \quad (3.21)$$

où S_{tot} est l'intégrale de $\rho(x, t)$ dans tout l'espace cellulaire. Dans les faits, nous représenterons les quantités $S_{1(2)}^*(t)$ et $\sigma^*(t)$, les ratios de protéines normalisés par le ratio initial dans le complexe 1 non-photoblanchi pour comparer directement nos résultats aux données des expériences de FRAP, c'est-à-dire que :

$$S_{1(2)}^*(t) = \frac{S_{1(2)}(t)}{S_{1(2)}(t=0)} \quad (3.22)$$

$$\sigma^*(t) = \frac{\sigma(t)}{\sigma(t=0)} \quad (3.23)$$

3.3. APPROCHE INTERMÉDIAIRE

Pour explorer les deux hypothèses définies en fin de section précédente, nous considérons deux versions du potentiels $\tilde{U}(\mathbf{r})$. Nous n'écrivons ici que les définitions à une dimension mais ces formes sont facilement généralisables.

1. Le potentiel \tilde{U} est un simple puit attractif (voir figure 3.5). Sa forme analytique \tilde{U}_{SB} (SB pour Sans Barrière) est définie par :

$$\tilde{U}_{\text{SB}}(x) = \tilde{f}_{\text{SB}}(x - x_c) + \tilde{f}_{\text{SB}}(x + x_c) \quad (3.24)$$

où $\pm x_c = 0,25\mu\text{m}$ sont les positions d'équilibre des complexes ParBS sur le nucleoïde, et avec :

$$\tilde{f}_{\text{SB}}(x) = \begin{cases} B_1 x^2 + B_2, & \text{si } -R \leq x \leq R \\ 0, & \text{sinon} \end{cases} \quad (3.25)$$

où les constantes B_1 et B_2 sont choisies telles que 90% des protéines ParB se trouvent dans un rayon $R = 18,5\text{nm}$ des positions $\pm x_c = 0,25\mu\text{m}$ et $\tilde{U}_{\text{SB}}(R) = 0$. La partie attractive correspond au complexe ParBS.

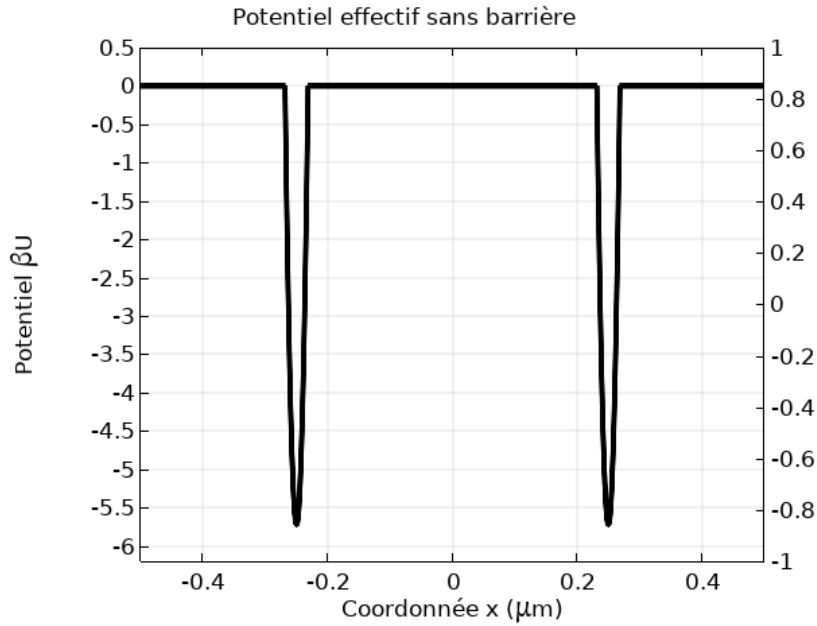


FIGURE 3.5 – Simple puit attractif utilisé dans nos simulations (cas à une dimension).

2. Le potentiel \tilde{U} possède en plus deux domaines répulsifs ou barrières d'énergie (see figure 3.6). Nous considérons en pratique :

$$\tilde{U}_{\text{B}}(x) = \tilde{f}_{\text{B}}(x - x_c) + \tilde{f}_{\text{B}}(x + x_c) \quad (3.26)$$

avec :

$$\tilde{U}_B(x) = \begin{cases} C_1 x^4 + C_2 x^2 + C_3, & \text{si } -X_B \leq x \leq X_B \\ 0, & \text{sinon} \end{cases} \quad (3.27)$$

où X_B est la plus grande racine du polynôme que nous avons défini. Les

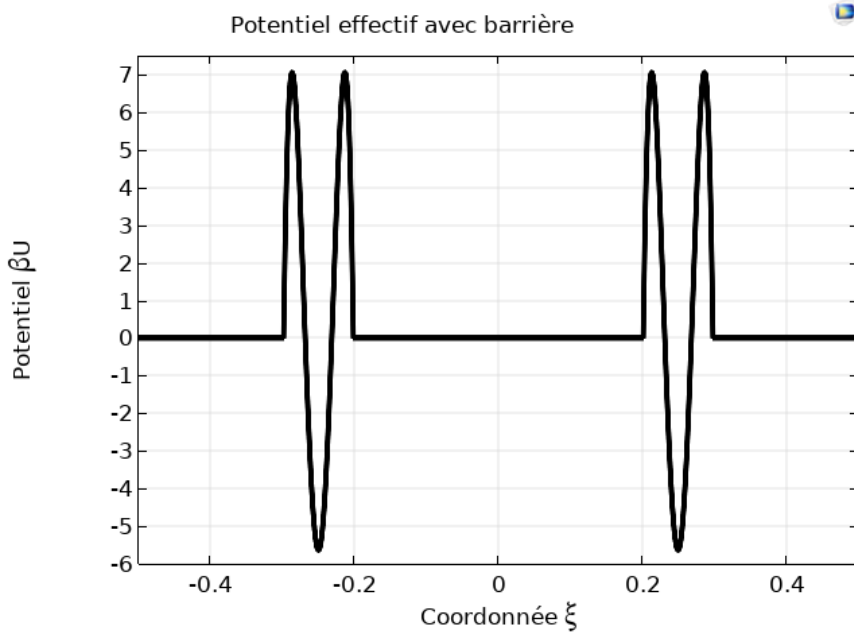


FIGURE 3.6 – Potentiel avec barrière utilisé dans nos simulations. Les domaines attractifs et répulsifs respectent les valeurs déterminées par le modèle de comptage.

coefficients C_1 , C_2 et C_3 sont choisis tels que $\tilde{U}_B(R) = 0$ (i.e. la partie attractive correspond au complexe lui-même), et tels que les valeurs moyennes de la partie attractive de \tilde{U}_B et des barrières vérifient (à une dimension) :

$$\frac{1}{2R} \int_{-R}^R \tilde{U}_B(x) dx \approx \Delta \tilde{E}_{in} - \Delta \tilde{E}_{out} = \Delta \tilde{E}, \quad (3.28)$$

où $\Delta \tilde{E} = \beta \Delta E < 0$ dans notre convention, et

$$\frac{1}{X_B - R} \int_R^{X_B} \tilde{U}_B(x) dx \approx \Delta \tilde{E}_{in}, \quad (3.29)$$

où $\Delta \tilde{E}_{in} = \beta \Delta E_{in}$ et $\Delta \tilde{E}_{out} = \beta \Delta E_{out}$ proviennent des estimations basées sur les expériences de FRAP (voir section précédente).

Résultats

- Pour les valeurs mesurées par fluorescence des coefficients de diffusion $D_L^B = 0.02 \mu\text{m}^2/\text{s}$ et $D_V^B = 0.5 \mu\text{m}^2/\text{s}$ [63], nous observons des temps de récupération de la fluorescence de l'ordre de quelques secondes dans le cas sans barrière (voir figure 3.7). Nous obtenons un résultat remarquablement similaire si le coefficient de diffusion $D(\rho)$ est simplement remplacé par D_V^B , c'est-à-dire si les phases haute et basse densité ont le même coefficient de diffusion (voir figure 3.7).

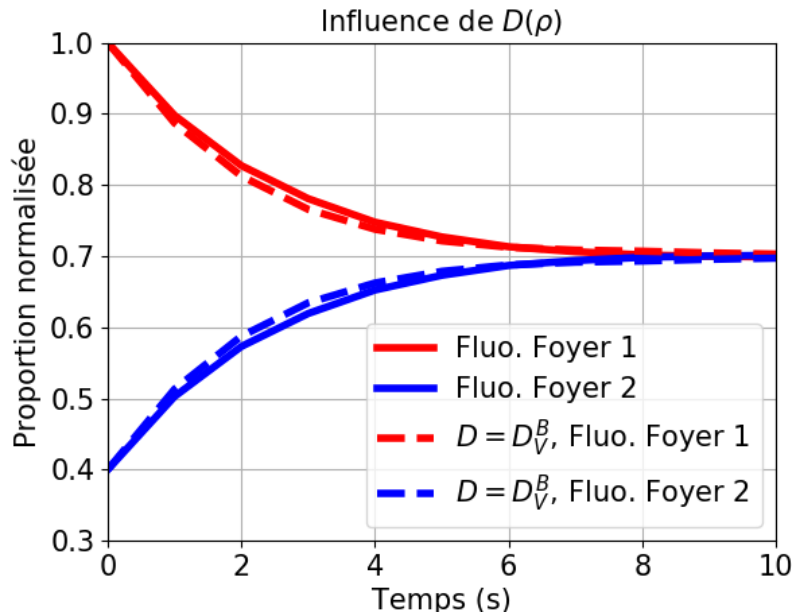


FIGURE 3.7 – **Courbe de recouvrement** *in silico* avec les paramètres mesurés par les expérimentateurs. En trait plein : courbes de recouvrement en bleu et perte en rouge prédites pour un système à deux coefficient de diffusion sans barrière. En trait discontinu : les protéines ne diffusent plus qu'avec un seul coefficient de diffusion rapide D_V^B . On trouve sensiblement le même résultat, c'est-à-dire un temps de récupération de l'ordre de quelques secondes.

- La valeur du coefficient de diffusion des protéines à l'intérieur des complexes est capitale. En effet, choisir une diffusion bien plus lente que la valeur déduite des expériences modifie conséquemment la cinétique de récupération. Par exemple, pour la valeur $D_L^B = 1.5 \times 10^{-4} \mu\text{m}^2/\text{s}$ au lieu de $0,02 \mu\text{m}^2/\text{s}$, nous arrivons à tracer des courbes de récupération très similaires aux résultats des expériences de FRAP (voir figure 3.8). Le même constat s'impose à $D = 2$ pour $D_L^B = 4.5 \times 10^{-4} \mu\text{m}^2/\text{s}$ (voir également figure 3.8). La dimensionalité semble jouer un rôle important et il est raisonnable de penser que les courbes de récupération pour $D = 3$, qui nécessitent des temps de calcul longs, reproduiraient le com-

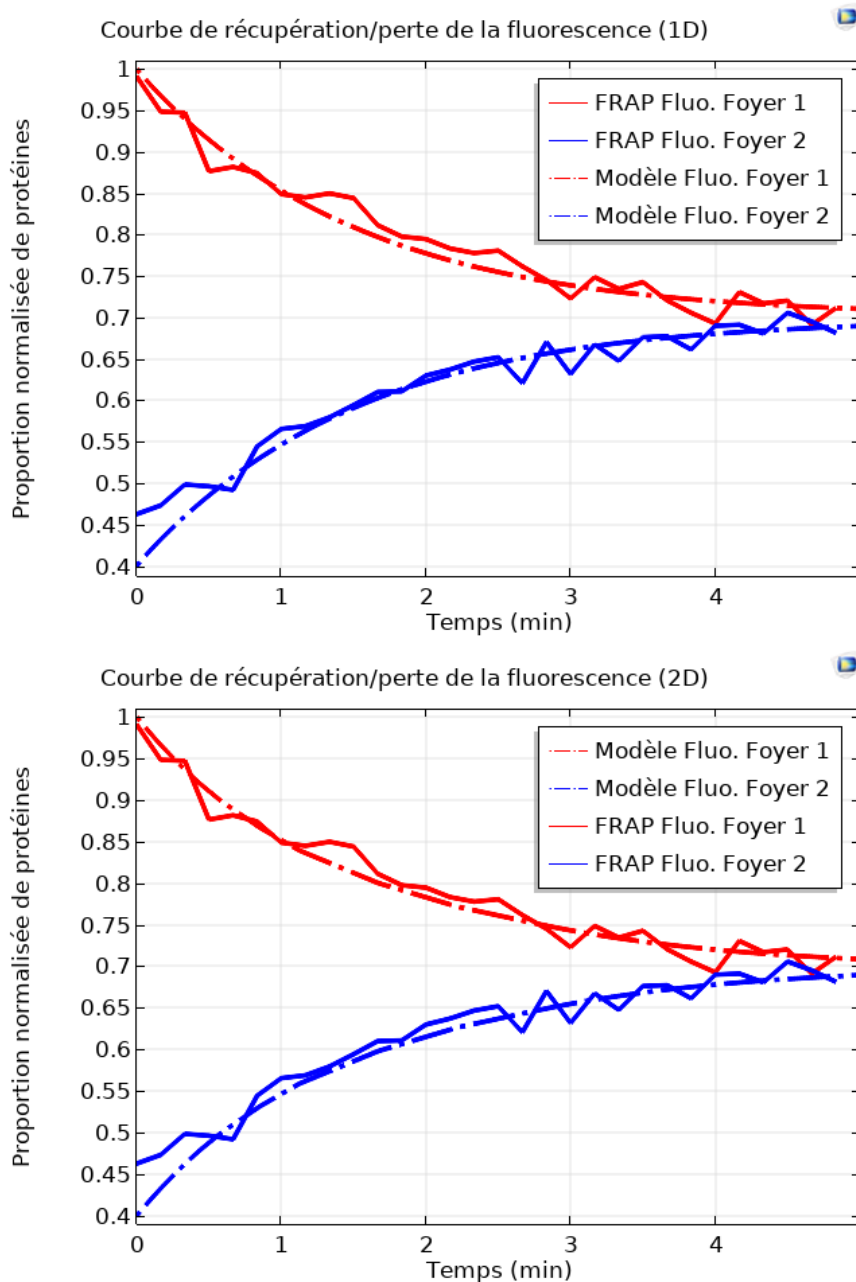


FIGURE 3.8 – **Courbe de recouvrement** *in silico* avec un coefficient de diffusion de la phase condensée D_L^B faible. En haut, à une dimension, $D_L^B = 1.5 \times 10^{-4} \mu\text{m}^2/\text{s}$; en bas, à deux dimensions, $D_L^B = 4.5 \times 10^{-4} \mu\text{m}^2/\text{s}$. En trait plein : courbes de recouvrement en bleu et perte en rouge issues des expériences. En trait discontinu : courbes de recouvrement en bleu et perte en rouge issues des simulations.

portement expérimental pour un coefficient de l'ordre de $10^{-3} \mu\text{m}^2/\text{s}$. Un temps de récupération de l'ordre de la minute peut donc s'expliquer

3.3. APPROCHE INTERMÉDIAIRE

par une diffusion lente à l'intérieur des complexes.

- Pour le potentiel avec barrière, nous rappelons que les expériences de FRAP nous permettent d'évaluer les taux de sortie $k_{out} = (0.01003 \pm 0.00035)s^{-1}$ et d'entrée d'une protéine ParB des foyers $k_{in} = (0.043661 \pm 0.01178)s^{-1}$. Pour $D = 1$, les estimations des niveaux d'énergie du potentiel sont données par :

$$\Delta\tilde{E}_{in} = -\ln(k_{in}/k_{in}^0) \quad (3.30)$$

$$\Delta\tilde{E}_{out} = -\ln(k_{out}/k_{out}^0) \quad (3.31)$$

avec $k_{in}^0 = 2D_v/l_c^2$ où $l_c = 0.5 \mu\text{m}$, et $k_{out}^0 = 2D_l/l_f^2$ où $l_f = 0.037 \mu\text{m}$. Nous trouvons alors $\Delta\tilde{E}_{in} = 4.452$, $\Delta\tilde{E}_{out} = 7.977$ et $\Delta\tilde{E} = -3.525$. Dans nos simulations, nous devons utiliser la valeur $\Delta\tilde{E} = -3.650$ pour concentrer 90% des particules dans les foyers tandis que la valeur moyenne des énergies de barrière est exactement $\Delta\tilde{E}_{in}$. Ces définitions mènent aux courbes de récupération/perte de la fluorescence dans la figure 3.9 en bon accord avec les expériences de FRAP.

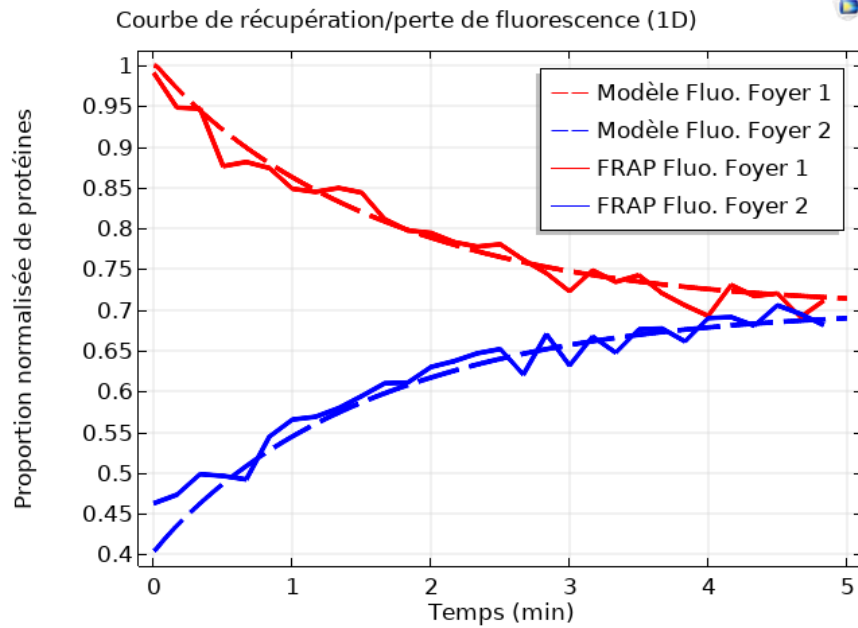


FIGURE 3.9 – **Courbe de recouvrement** *in silico* prédite avec potentiel avec barrière. En trait plein : courbes de recouvrement en bleu et perte en rouge issues des expériences. En trait discontinu : courbes de recouvrement en bleu et perte en rouge issues des simulations.

En l'état, nous ne pouvons faire de distinction entre un mécanisme de diffusion faible à l'intérieur des foyers et des barrières d'énergie effectives provenant des interactions entre protéines ParB et ParA. De plus, l'origine, si elle existe, du potentiel U reste obscure. Dans la prochaine partie, nous développons une

approche fondamentale bien définie. Le potentiel U trouve alors son origine dans les interactions entre protéines sur un réseau et la formation de phase haute occupation par séparation de phase.

3.4 Approche fondamentale

Avec cette nouvelle approche, nous voulons obtenir les équations qui régissent la cinétique des protéines ParB adsorbées sur l'ADN fluctuant dans l'espace intracellulaire. Lors d'une expérience de photoblanchiment, nous comptons deux sous populations de protéines : les protéines fluorescentes et les protéines photoblanchies. Notre approche fondamentale doit donc distinguer au moins deux espèces, et davantage si le modélisateur veut enquêter sur le rôle des protéines ParA dans la formation ou la survie des complexes ParBS. Ainsi, nous généralisons le modèle de gaz sur réseau décrit au chapitre précédent à une seule espèce pour plusieurs espèces chimiques. Nous parlerons alors de modèle à P champs pour parler de notre approche cinétique basée sur un gaz sur réseau à P espèces.

3.4.1 Gaz sur réseau continu à une et deux espèces

Nous avons déjà déterminé la forme de l'énergie libre de Helmholtz pour une seule espèce dans le cas inhomogène et dans l'approximation du champ moyen. Le champ moyen présente certains défauts mais il contient l'essentiel de la thermodynamique des transitions de phase. Nous pouvons de fait mapper notre modèle LRLG à une seule dimension avec un modèle de gaz sur réseau à courte portée à une dimension effective D . Nous présentons d'abord le calcul permettant d'obtenir cette énergie libre dans le cas d'un gaz sur réseau à deux espèces. La limite thermodynamique nous permettra d'obtenir les équations cinétiques pour les protéines ParB sur l'ADN.

Construction du modèle

Comme dans le modèle de gaz sur réseau à une espèce, nous supposons que l'ADN comble l'espace intracellulaire de telle sorte que nous puissions définir un réseau cubique de N sites de maille l_m , où l_m est l'empreinte d'une protéine sur l'ADN. Ce réseau d'ADN est mis en contact avec un réservoir de deux espèces chimiques : l'espèce A et l'espèce B. Par simplification, nous supposons également que les espèces A et B possèdent la même empreinte sur l'ADN de telle sorte que les deux espèces voient le même réseau. Nous associons à chaque site du réseau les quantités :

1. a_i , l'occupation locale des particules de l'espèce A, qui vaut 1 si une particule de type A se trouve sur le site i et 0 sinon ;

3.4. APPROCHE FONDAMENTALE

2. b_i , l'occupation locale des particules de l'espèce B, qui vaut 1 si une particule de type B se trouve sur le site i et 0 sinon ;
3. ϵ_i^a , une énergie spécifique d'adsorption pour les exemplaires de l'entité A au site i , définie attractive quand négative ;
4. de même, ϵ_i^b , une énergie spécifique d'adsorption pour les exemplaires de l'entité B au site i , définie attractive quand négative.

Les nombres de particules du type A ou B sur le réseau sont ajustés en fonction des potentiels chimiques respectifs μ_a de l'espèce A et μ_b de l'espèce B. Toutes les particules peuvent interagir via des interactions locales, c'est-à-dire sur le même site, ou via des interactions entre particules adsorbées sur des sites premiers voisins. Ainsi, nous définissons :

1. le couplage J_{aa} entre particules de type A sur des sites premiers voisins ;
2. le couplage J_{bb} entre particules de type B sur des sites premiers voisins ;
3. le couplage complémentaire J_{ab} entre une particule de type A et une particule de type B sur des sites premiers voisins ;
4. une interaction locale J_I , dite intrinsèque, entre une particule de type A et une particule de type B sur le même site.

Avec ces définitions, nous sommes en mesure d'écrire l'hamiltonien grand-canonique du gaz sur réseau à deux espèces :

$$\begin{aligned} \mathcal{H} = & - \sum_{\langle i,j \rangle} \frac{J_{aa}}{2} a_i a_j + \frac{J_{bb}}{2} b_i b_j + J_{ab} a_i b_j \\ & - \sum_{i=1}^N a_i (\mu_a - \epsilon_i^a) + b_i (\mu_b - \epsilon_i^b) + J_I a_i b_i \end{aligned} \quad (3.32)$$

où $\langle i, j \rangle$ indique une somme sur tous les couples premiers voisins sur le réseau. Le facteur 1/2 devant les interactions corrige la double prise en compte des couples voisins dans la somme. Notons que le terme d'interaction entre particules de type différent ne possède pas ce facteur. Pour déterminer l'énergie libre de Helmholtz inhomogène en champ moyen du système, nous choisissons d'appliquer une transformation de Hubbard-Stratonovich sur sa fonction de partition \mathcal{Z} . Nous rappelons que cette approche inhomogène repose sur le principe de l'équilibre local thermodynamique [83] : le système, soumis à des conditions qui le dévient de l'équilibre thermodynamique, peut être divisé en parties mésoscopiques, coïncidant avec les sites du réseau et considérées à l'équilibre. Les propriétés locales telles que la température, le potentiel chimique intrinsèque ou la pression sont ainsi bien définies et sont déterminées avec les relations classiques de la thermodynamique. Ainsi, nous pouvons écrire

dans un formalisme vectoriel :

$$\begin{aligned} \mathcal{Z} = \sum_{\{a_i\}} \sum_{\{b_i\}} \exp \left\{ \frac{\beta}{2} [(\mathbf{a} + \gamma \mathbf{b})^T J_1 (\mathbf{a} + \gamma \mathbf{b}) + \mathbf{a}^T J_2 \mathbf{a}] \right\} \\ \times \exp [\beta (\boldsymbol{\psi}_a^T \mathbf{a} + \boldsymbol{\psi}_b^T \mathbf{b} + J_I \mathbf{a}^T \mathbf{b})] \end{aligned} \quad (3.33)$$

avec \mathbf{a} et \mathbf{b} , de i -ème composantes respectives a_i et b_i , sont les vecteurs occupations des espèces A et B. Les vecteurs $\boldsymbol{\psi}_a$ et $\boldsymbol{\psi}_b$ ont pour i -èmes composantes $\mu_a - \epsilon_i^a$ et $\mu_b - \epsilon_i^b$. Les termes γ , J_1 et J_2 permettent d'agencer les termes de \mathcal{Z} pour préparer le découplage induit par la transformation de Hubbard-Stratonovich. Nous avons $\gamma = \frac{J_{bb}}{J_{ab}}$ tandis que les quantités $J_1 = J_{aa} - J_2$ et $J_2 = \frac{J_{ab}^2}{J_{bb}}$ sont vues comme des matrices d'adjacence, c'est-à-dire comme des propriétés du réseau. Ainsi, en introduisant deux champs fictifs $\boldsymbol{\phi}_1 - \frac{1}{\gamma} \boldsymbol{\psi}_b$ et $\boldsymbol{\phi}_2 - \boldsymbol{\psi}_a + \frac{1}{\gamma} \boldsymbol{\psi}_b$, nous écrivons \mathcal{Z} après avoir intégré sur les champs a et b comme (voir Chapitre 2, Illustration des régimes thermodynamiques) :

$$\mathcal{Z} = C \int_{\mathbb{R}^{2N}} \prod_{i=1}^N d\phi_{1i} d\phi_{2i} \exp[-\beta S(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\psi}_a, \boldsymbol{\psi}_b)] \quad (3.34)$$

avec C une constante qui dépend des déterminants des matrices J_1 et J_2 . S est la fonction d'essai définie par :

$$\begin{aligned} S = \frac{1}{2} \left(\boldsymbol{\phi}_1 - \frac{1}{\gamma} \boldsymbol{\psi}_b \right)^T J_1 \left(\boldsymbol{\phi}_1 - \frac{1}{\gamma} \boldsymbol{\psi}_b \right) \\ + \frac{1}{2} \left(\boldsymbol{\phi}_2 - \boldsymbol{\psi}_a + \frac{1}{\gamma} \boldsymbol{\psi}_b \right)^T J_2 \left(\boldsymbol{\phi}_2 - \boldsymbol{\psi}_a + \frac{1}{\gamma} \boldsymbol{\psi}_b \right) \\ - \frac{1}{\beta} \sum_{i=1}^N \ln \{ 1 + \exp[\beta(\phi_{1i} + \phi_{2i})] \\ + \exp[\beta\gamma\phi_{1i}] + \exp[\beta((1 + \gamma)\phi_{1i} + \phi_{2i} + J_I)] \} \end{aligned} \quad (3.35)$$

Une méthode de col au premier ordre permet d'optimiser cette fonction d'essai. Au premier ordre, nous avons alors $\mathcal{Z} \approx \exp[\beta S(\boldsymbol{\phi}_1^*, \boldsymbol{\phi}_2^*, \boldsymbol{\psi}_a, \boldsymbol{\psi}_b)]$ avec par optimisation :

$$\left(\frac{\partial S}{\partial \boldsymbol{\phi}_1} \right)_{\boldsymbol{\phi}_1^*, \boldsymbol{\phi}_2^*} = \vec{0} \quad (3.36)$$

$$\left(\frac{\partial S}{\partial \boldsymbol{\phi}_2} \right)_{\boldsymbol{\phi}_1^*, \boldsymbol{\phi}_2^*} = \vec{0} \quad (3.37)$$

Les occupations moyennes $a_{mi} = \langle a_i \rangle$ et $b_{mi} = \langle b_i \rangle$ exprimées au site i et le potentiel grand canonique sont liés dans l'approximation du champ moyen par

3.4. APPROCHE FONDAMENTALE

les relations :

$$\mathbf{a}_m \approx - \left(\frac{\partial \Omega}{\partial \psi_a} \right)_{\phi_1^*, \phi_2^*} \quad (3.38)$$

$$\mathbf{b}_m \approx - \left(\frac{\partial \Omega}{\partial \psi_b} \right)_{\phi_1^*, \phi_2^*} \quad (3.39)$$

où \mathbf{a}_m et \mathbf{b}_m sont des vecteurs de i -ème élément $\langle a_i \rangle$ et $\langle b_i \rangle$. Nous trouvons ainsi les équations vérifiées par les champs fictifs ϕ_1^* , ϕ_2^* à l'optimisation :

$$\langle a_i \rangle = \frac{e^{\beta(\phi_{1i}^* + \phi_{2i}^*)} e^{\beta[(1+\gamma)\phi_{1i}^* + \phi_{2i}^* + J_I]}}{1 + e^{\beta(\phi_{1i}^* + \phi_{2i}^*)} + e^{\beta\gamma\phi_{1i}^*} + e^{\beta[(1+\gamma)\phi_{1i}^* + \phi_{2i}^* + J_I]}} \quad (3.40)$$

$$\langle b_i \rangle = \frac{e^{\beta\gamma\phi_{1i}^*} + e^{\beta[(1+\gamma)\phi_{1i}^* + \phi_{2i}^* + J_I]}}{1 + e^{\beta(\phi_{1i}^* + \phi_{2i}^*)} + e^{\beta\gamma\phi_{1i}^*} + e^{\beta[(1+\gamma)\phi_{1i}^* + \phi_{2i}^* + J_I]}} \quad (3.41)$$

Inverser ces dernières relations nous permet de trouver les expressions analytiques des champs ϕ_{1i}^* et ϕ_{2i}^* en fonction des occupations moyennes :

$$e^{\beta\phi_{1i}^*} = \left(\frac{1 - \langle a_i \rangle - \langle b_i \rangle + \langle a_i \rangle K - \langle b_i \rangle K - \sqrt{\Delta_1}}{2(\langle b_i \rangle - 1)K} \right)^{\frac{1}{\gamma}} \quad (3.42)$$

$$e^{\beta\phi_{2i}^*} = \frac{1 - \langle a_i \rangle - \langle b_i \rangle - \langle a_i \rangle K + \langle b_i \rangle K - \sqrt{\Delta_2}}{2(\langle a_i \rangle - 1)K e^{\beta\phi_{1i}^*}} \quad (3.43)$$

avec $K = \exp(\beta J_I)$ et avec les discriminants :

$$\Delta_1 = (\langle b_i \rangle + \langle b_i \rangle K - \langle a_i \rangle K + \langle a_i \rangle - 1)^2 - 4\langle b_i \rangle(\langle b_i \rangle - 1)K \quad (3.44)$$

$$\Delta_2 = (\langle a_i \rangle + \langle a_i \rangle K - \langle b_i \rangle K + \langle b_i \rangle - 1)^2 - 4\langle a_i \rangle(\langle a_i \rangle - 1)K \quad (3.45)$$

A ce stade, nous sommes capable d'écrire une expression analytique du potentiel grand canonique. Nous pouvons aussi changer d'ensemble statistique et calculer l'énergie libre de Helmholtz à nombres de particules de type A et B constants en utilisant la transformation de Legendre :

$$\mathcal{F} = \Omega + \mathbf{a}_m^T \boldsymbol{\mu}_a + \mathbf{b}_m^T \boldsymbol{\mu}_b \quad (3.46)$$

Ainsi, de manière générale, nous obtenons :

$$\begin{aligned} \mathcal{F} \approx & - \sum_{\langle i, j \rangle} \frac{J_{aa}}{2} \langle a_i \rangle \langle a_j \rangle + \frac{J_{bb}}{2} \langle b_i \rangle \langle b_j \rangle + J_{ab} \langle a_i \rangle \langle b_j \rangle \\ & + \sum_{i=1}^N \langle a_i \rangle \epsilon_i^a + \langle b_i \rangle \epsilon_i^b + \langle a_i \rangle \phi_{2i}^* + \langle b_i \rangle \gamma \phi_{1i}^* \\ & - k_B T \sum_{i=1}^N \ln[1 + e^{\beta(\phi_{1i}^* + \phi_{2i}^*)} + e^{\beta\gamma\phi_{1i}^*} + e^{\beta[(1+\gamma)\phi_{1i}^* + \phi_{2i}^* + J_I]}] \end{aligned} \quad (3.47)$$

Nous avons déjà déterminé l'énergie libre de Helmholtz inhomogène pour un seul type de particules (voir Chapitre 2, Illustration des régimes thermodynamiques). Pour des particules d'occupation $\{\Phi_i\}$ sur le même réseau et interagissant avec un couplage J , nous aurions simplement :

$$\begin{aligned} \mathcal{F}_{\text{IE}}(\{\Phi_i\}) = & -\frac{J}{2} \sum_{\langle i,j \rangle} \langle \Phi_i \rangle \langle \Phi_j \rangle + \sum_{i=1}^N \langle \Phi_i \rangle \epsilon_i \\ & + k_B T \sum_{i=1}^N \langle \Phi_i \rangle \ln \langle \Phi_i \rangle + (1 - \langle \Phi_i \rangle) \ln (1 - \langle \Phi_i \rangle) \end{aligned} \quad (3.48)$$

L'ajout d'un second type de particule complexifie considérablement le terme entropique de l'énergie libre. Néanmoins, il existe deux comportements remarquables selon la limite de l'interaction intrinsèque J_I .

1. Limite d'indépendance locale : lorsque J_I tend vers 0, c'est-à-dire lorsque deux particules de type différents n'interagissent pas sur un même site, nous pouvons montrer que :

$$\begin{aligned} \lim_{J_I \rightarrow 0} \mathcal{F} = & - \sum_{\langle i,j \rangle} \frac{J_{aa}}{2} \langle a_i \rangle \langle a_j \rangle + \frac{J_{bb}}{2} \langle b_i \rangle \langle b_j \rangle + J_{ab} \langle a_i \rangle \langle b_j \rangle \\ & + \sum_{i=1}^N \langle a_i \rangle \epsilon_i^a + \langle b_i \rangle \epsilon_i^b + k_B T \sum_{i=1}^N \langle a_i \rangle \ln \langle a_i \rangle + \langle b_i \rangle \ln \langle b_i \rangle \\ & + k_B T \sum_{i=1}^N (1 - \langle a_i \rangle) \ln (1 - \langle a_i \rangle) \\ & + k_B T \sum_{i=1}^N (1 - \langle b_i \rangle) \ln (1 - \langle b_i \rangle) \end{aligned} \quad (3.49)$$

Dans cette limite, le terme entropique se réduit simplement à celui du gaz sur réseau à une seule espèce. Si les interactions entre espèces différentes sur des sites premiers voisins est nulle ($J_{ab} = 0$), l'énergie libre s'écrit comme la somme de deux énergies libres indépendantes, l'une pour l'espèce A et l'autre pour l'espèce B. Nous formons ainsi deux sous-systèmes indépendants.

2. Limite d'exclusion locale : lorsque J_I tend vers $-\infty$, c'est-à-dire lorsqu'une particule d'un type sur un site exclue les particules d'autres types,

l'énergie libre s'écrit :

$$\begin{aligned}
 \lim_{J_I \rightarrow -\infty} \mathcal{F} = \mathcal{F}_{2E} = & - \sum_{\langle i,j \rangle} \frac{J_{aa}}{2} \langle a_i \rangle \langle a_j \rangle + \frac{J_{bb}}{2} \langle b_i \rangle \langle b_j \rangle + J_{ab} \langle a_i \rangle \langle b_j \rangle \\
 & + \sum_{i=1}^N \langle a_i \rangle \epsilon_i^a + \langle b_i \rangle \epsilon_i^b + \langle a_i \rangle \ln \langle a_i \rangle + \langle b_i \rangle \ln \langle b_i \rangle \\
 & + k_B T \sum_{i=1}^N (1 - \langle a_i \rangle - \langle b_i \rangle) \ln (1 - \langle a_i \rangle - \langle b_i \rangle) \quad (3.50)
 \end{aligned}$$

L'exclusion réciproque des particules se traduit dans le terme entropique par une borne de définition supérieure de la variable $\langle a_i \rangle + \langle b_i \rangle$ qui ne peut excéder 1 sur le site i .

Pour l'application au système de partition ParABS, nous nous placerons dans cette deuxième limite. En effet, le système nous invite à considérer ce cas puisque, par exemple, deux protéines ParB, ou une protéine ParB et une protéine ParA, ne peuvent occuper le même site sur l'ADN pour des raisons d'encombrement stérique.

Gaz sur réseau dans la limite continue

Nous traduisons nos expressions dans la limite thermodynamique pour déterminer les équations régissant la cinétique des particules de type A et B sur le réseau. Nous considérons que la taille typique L du système est grande devant le paramètre de maille l_m du réseau dans le cas de l'exclusion locale (définie ci-dessus). Ainsi, les occupations $\langle a_i \rangle$ et $\langle b_i \rangle$ et les énergie spécifiques d'adsorption ϵ_i^a et ϵ_i^b deviennent des fonctions de l'espace continue, soit respectivement $a(\mathbf{r})$, $b(\mathbf{r})$, $\epsilon^A(\mathbf{r})$ et $\epsilon^B(\mathbf{r})$. Le produit du type $\langle a_i \rangle \langle a_j \rangle$ sur sites premiers voisins se développe en utilisant un développement de Taylor. Par exemple, à une dimension, le produit $\langle a_i \rangle \langle a_{i\pm 1} \rangle$ devient $a(x)a(x \pm l_m)$ et s'écrit au deuxième ordre :

$$a(x)a(x \pm l_m) \approx a(x) \left(a(x) \pm l_m \frac{\partial a}{\partial x} + \frac{l_m^2}{2} \frac{\partial^2 a}{\partial x^2} \right) \quad (3.51)$$

Ainsi, on peut montrer qu'à une dimension D quelconque, on a par exemple :

$$-\frac{J_{AA}}{2} \sum_{\langle i,j \rangle} a_i a_j = \int_V \frac{J_{AA}}{2} \{ l_m^2 [\nabla a(\mathbf{r})]^2 - q a(\mathbf{r}) \} \frac{d^D \mathbf{r}}{l_m^D} \quad (3.52)$$

où V est le volume de la bactérie, pour des choix de flux normal $\mathbf{n} \cdot \nabla a = \mathbf{n} \cdot \nabla b = 0$ nul sur les bords, où \mathbf{n} est le vecteur normal aux bords du système. Nous devons normaliser l'intégrale par le facteur l_m^D pour conserver une énergie libre

extensive. La fonctionnelle d'énergie $\mathcal{F}[\rho(\mathbf{r})]$ du gaz sur réseau à deux espèces s'écrit dans la limite thermodynamique et dans la limite de l'exclusion locale :

$$\begin{aligned} \mathcal{F}_{2E} = & \int_V \frac{J_{aa}}{2} \{l_m^2[\nabla a]^2 - qa^2\} + \frac{J_{bb}}{2} \{l_m^2[\nabla b]^2 - qb^2\} \\ & + J_{ab} \{m^2[\nabla a \nabla b] - qab\} + a\epsilon^a + b\epsilon^b \\ & + k_B T [a \ln a + b \ln b + (1 - a - b) \ln(1 - a - b)] \frac{d^D \mathbf{r}'}{l_m^D} \end{aligned} \quad (3.53)$$

De la même manière, nous pouvons montrer simplement en prenant $a \rightarrow \Phi$ et $b \rightarrow 0$ que l'énergie libre du gaz sur réseau à une espèce s'écrit dans la limite thermodynamique :

$$\begin{aligned} \mathcal{F}_{1E} = & \int_V \frac{J}{2} [l_m^2(\nabla \Phi)^2 - q\Phi^2] + \Phi\epsilon \\ & + k_B T [\Phi \ln \Phi + (1 - \Phi) \ln(1 - \Phi)] \frac{d^D \mathbf{r}'}{l_m^D} \end{aligned} \quad (3.54)$$

Nous sommes désormais en position de dériver des équations cinétiques de ces énergies libres.

Equations dérivées de l'énergie libre

Nous supposons que le système constitué du réseau d'ADN et des protéines respecte le principe de l'équilibre thermodynamique locale [83]. Nous considérons que l'occupation totale sur le réseau est conservée. Pour le système de partition ParABS, cette dernière supposition revient à dire que le rapport du nombre de protéines M et du nombre de site d'ADN N reste conservé, c'est-à-dire que M croît de la même manière que N lors de la réplication de l'ADN. Concrètement, la bactérie ne doit pas connaître de perturbation trop intense au risque de changer de régime thermodynamique. Sous ces hypothèses, nous pouvons considérer qu'un flux local $\mathbf{j}_{a/b}(\mathbf{r}, t)$ de particules de type A ou B est une réponse linéaire aux variations locales du potentiel chimique correspondant $\mu_{a/b}(\mathbf{r}, t)$, d'où :

$$\mathbf{j}_a(\mathbf{r}, t) = -\frac{L_a}{T} \nabla \mu_a(\mathbf{r}, t) \quad (3.55)$$

$$\mathbf{j}_b(\mathbf{r}, t) = -\frac{L_b}{T} \nabla \mu_b(\mathbf{r}, t) \quad (3.56)$$

où L_a et L_b sont des coefficients de transport, dits d'Onsager, reliés à la diffusion des particules dans le système [83, 123, 124]. Ces coefficients dépendent a priori des occupations a et b . Nous expliciterons cette dépendance ultérieurement. Les expressions précédentes peuvent être vues comme une généralisation

3.4. APPROCHE FONDAMENTALE

de la loi de Fick. Les potentiels chimiques respectifs μ_a et μ_b dérivent simplement de la fonctionnelle \mathcal{F}_{2E} :

$$\begin{aligned} \mu_a = l_m^D \frac{\delta \mathcal{F}_{2E}}{\delta a} &= -J_{aa} l_m^2 \nabla^2 a - J_{ab} l_m^2 \nabla^2 b - J_{aa} q a - J_{ab} q b \\ &+ \epsilon^a + k_B T \ln \left(\frac{a}{1-a-b} \right) \end{aligned} \quad (3.57)$$

$$\begin{aligned} \mu_b = l_m^D \frac{\delta \mathcal{F}_{2E}}{\delta b} &= -J_{bb} l_m^2 \nabla^2 b - J_{ab} l_m^2 \nabla^2 a - J_{bb} q b - J_{ab} q a \\ &+ \epsilon^b + k_B T \ln \left(\frac{b}{1-a-b} \right) \end{aligned} \quad (3.58)$$

Les flux \mathbf{j}_a et \mathbf{j}_b dépendent à la fois de l'occupation des particules de même type, via les termes en J_{aa} et J_{bb} , et de l'occupation des particules du type B au travers des termes d'interaction en J_{ab} et au terme d'origine entropique d'exclusion locale $\ln(1-a-b)$. La conservation des occupations correspond aux équations de continuité :

$$\frac{\partial a(\mathbf{r}, t)}{\partial t} + \nabla \cdot \mathbf{j}_a(\mathbf{r}, t) = 0 \quad (3.59)$$

$$\frac{\partial b(\mathbf{r}, t)}{\partial t} + \nabla \cdot \mathbf{j}_b(\mathbf{r}, t) = 0 \quad (3.60)$$

En injectant l'expression des flux \mathbf{j}_a et \mathbf{j}_b dans l'expression précédente, nous trouvons les équations aux dérivées partielles décrivant l'évolution des champs d'occupation $a(\mathbf{r}, t)$ et $b(\mathbf{r}, t)$:

$$\frac{\partial a(\mathbf{r}, t)}{\partial t} = \nabla \cdot \left[\frac{L_a}{T} \nabla \mu_a(\mathbf{r}, t) \right] \quad (3.61)$$

$$\frac{\partial b(\mathbf{r}, t)}{\partial t} = \nabla \cdot \left[\frac{L_b}{T} \nabla \mu_b(\mathbf{r}, t) \right] \quad (3.62)$$

Ces équations couplées sont chacune du même type que l'équation de Cahn-Hilliard pour les mélanges [125]. Les conditions aux bords associées à cette équation sont choisies de telle sorte qu'elles prohibent tout flux de particules en dehors du domaine d'intégration. Notre approche numérique (voir Annexe) découple les champs d'occupation des champs de potentiel chimique et nécessite ainsi deux types de conditions au bords de Neumann :

$$\mathbf{n} \cdot \nabla \mu_A(\mathbf{r}, t)|_{\text{bords}} = \mathbf{n} \cdot \nabla \mu_B(\mathbf{r}, t)|_{\text{bords}} = 0 \quad (3.63)$$

$$\mathbf{n} \cdot \nabla a(\mathbf{r}, t)|_{\text{bords}} = \mathbf{n} \cdot \nabla b(\mathbf{r}, t)|_{\text{bords}} = 0 \quad (3.64)$$

Nous avons ainsi quatre équations couplées à résoudre. Avant d'appliquer ce modèle pour interpréter les expériences de photoblanchiment dans la prochaine section, nous proposons d'illustrer ces équations dans le cadre du modèle à une seule espèce.

Résultats du modèle à un champ

Comme dans la précédente sous-section, nous pouvons déterminer l'équation aux dérivées partielles qui gouvernent l'évolution spatio-temporelle du champ d'occupation $\Phi(\mathbf{r}, t)$ d'une espèce chimique sur un réseau à partir de l'équation (3.54) de l'énergie libre de Helmholtz dans la limite continue. Cette équation s'écrit :

$$\frac{\partial \Phi(\mathbf{r}, t)}{\partial t} = \nabla \left[\tilde{L}(\Phi) \nabla \mu(\mathbf{r}, t) \right], \quad (3.65)$$

avec

$$\mu(\mathbf{r}, t) = l_m^D \frac{\delta \mathcal{F}_{1E}}{\delta \Phi} = -Jl_m^2 \nabla^2 \Phi - Jq\Phi + \epsilon + k_B T \ln \left(\frac{\Phi}{1 - \Phi} \right) \quad (3.66)$$

Numériquement, nous cherchons à résoudre ces deux équations couplées. Ces équations s'accompagnent ainsi des conditions aux bords de Neumann pour le potentiel chimique μ et l'occupation Φ :

$$\mathbf{n} \cdot \nabla \mu(\mathbf{r}, t)|_{\text{bords}} = \vec{0} \quad (3.67)$$

$$\mathbf{n} \cdot \nabla \Phi(\mathbf{r}, t)|_{\text{bords}} = \vec{0} \quad (3.68)$$

Nous pouvons proposer une forme analytique du coefficient d'Onsager L . Pour des variations infinitésimales $\delta\Phi$ autour de l'occupation uniforme Φ_0 telle que le système existe dans un état fluide ou métastable, notre équation est simplement réduite à une équation de diffusion. En négligeant les termes non-linéaires et en annulant les champs ϵ , nous obtenons en effet :

$$\frac{\partial \delta\Phi(\mathbf{r}, t)}{\partial t} = D(\Phi_0) \nabla^2 \delta\Phi(\mathbf{r}, t) \quad (3.69)$$

$$(3.70)$$

où D est le coefficient collectif de diffusion de l'espèce sur réseau :

$$D(\Phi_0) = Lk_B \left(\frac{1}{\Phi_0} + \frac{1}{1 - \Phi_0} \right) \quad (3.71)$$

$$(3.72)$$

Nous pouvons déterminer les valeurs de L en fonction des coefficients de diffusion dans les phases de haute D_L et basse D^V occupation dans ces limites :

$$\lim_{\Phi_0 \rightarrow 0^+} L = L^V = \frac{\Phi_0 D^V}{k_B} \quad (3.73)$$

$$\lim_{\Phi_0 \rightarrow 1^-} L = L^L = \frac{(1 - \Phi_0) D^L}{k_B} \quad (3.74)$$

3.4. APPROCHE FONDAMENTALE

Le coefficient D^V est le coefficient de diffusion de l'espèce dans un régime dilué et D^L dans un régime dense. Entre ces deux limites, et notamment dans des régimes de coexistence, nous admettons que $\tilde{L}(\Phi)$ suit la simple formule d'interpolation :

$$\tilde{L}(\Phi) = k_B L(\Phi) \approx \frac{L^V L^L}{L^V + L^L} = \frac{D^V D^L \Phi (1 - \Phi)}{D^V \Phi + D^L (1 - \Phi)} \quad (3.75)$$

Nous appliquons ces équations à la formation des complexes ParBS : $\Phi(\mathbf{r}, t)$ décrit l'évolution du champ de protéines ParB sur l'ADN et $\tilde{\epsilon} = \beta\epsilon$ représente le champ d'adsorption spécifique lié aux séquences *parS*. Les simulations suivantes ont été réalisées à deux dimensions, d'où $q = 4$, dans un espace reproduisant l'espace cytoplasmique de longueurs caractéristiques $L = 1\mu m$ et $l = 0.3\mu m$. Le paramètre de maille l_m , empreinte des protéines ParB sur l'ADN, correspond à une longueur $l_m = 5.44nm$. Les coefficients de diffusion utilisés sont $D^V = D_B^V = 0,5\mu m^2/s$ et $D^L = D_B^L = 0,02\mu m^2/s$ [63]. Le profil initial utilisé est un profil bruité de moyenne $\Phi_0 = 0.08$, valeur utilisée dans le premier chapitre pour la séparation de phase unidimensionnelle des protéines ParB sur l'ADN. Nous reproduisons une expérience de trempe sans et avec une séquence *S*. Si la séquence *parS* est présente, elle est positionnée au milieu de l'espace d'intégration, en $\mathbf{r}_0 = \vec{0}$:

$$\epsilon(\mathbf{r}) = \varepsilon \left[\exp \left(-\frac{\mathbf{r}^2}{2l_m^2} \right) \right] \quad (3.76)$$

où $\varepsilon = -5k_B T$. Cette valeur intermédiaire ε permet d'éviter que le schéma numérique ne propose des valeurs de Φ supérieure à 1 en argument du terme entropique en logarithme. Nous représenterons les résultats des résolutions numériques au temps $t = 1s$. Nous utilisons le logiciel de simulation COMSOL Multiphysics pour obtenir nos résultats (voir Annexe 1). Nous choisissons trois valeurs successives de \tilde{J} qui placent successivement le système dans :

1. un régime fluide, au-dessus de la courbe de coexistence (voir figures 3.10 et 3.11), pour $\tilde{J} = 1$;
2. un régime de coexistence métastable, sous la courbe de coexistence mais au-dessus de la ligne spinodale (voir figures 3.12 et 3.13), pour $\tilde{J} = 2$;
3. et dans un régime de coexistence instable, sous la ligne spinodale (voir figures 3.14 et 3.15), pour $\tilde{J} = 3$. A $t = 1 s$, le système biologique subit encore un processus de maturation d'Ostwald. Aux temps longs, le profil stationnaire montre un foyer de protéines ParB localisé à un bord de l'espace d'intégration et un complexe ParBS mineur autour de la séquence *parS* centrée (voir figure 3.16). Il s'agit de l'équilibre thermodynamique dans le régime de coexistence instable. Ce résultat est commenté dans la partie Photoblanchiment *in silico* sans protéine ParA.

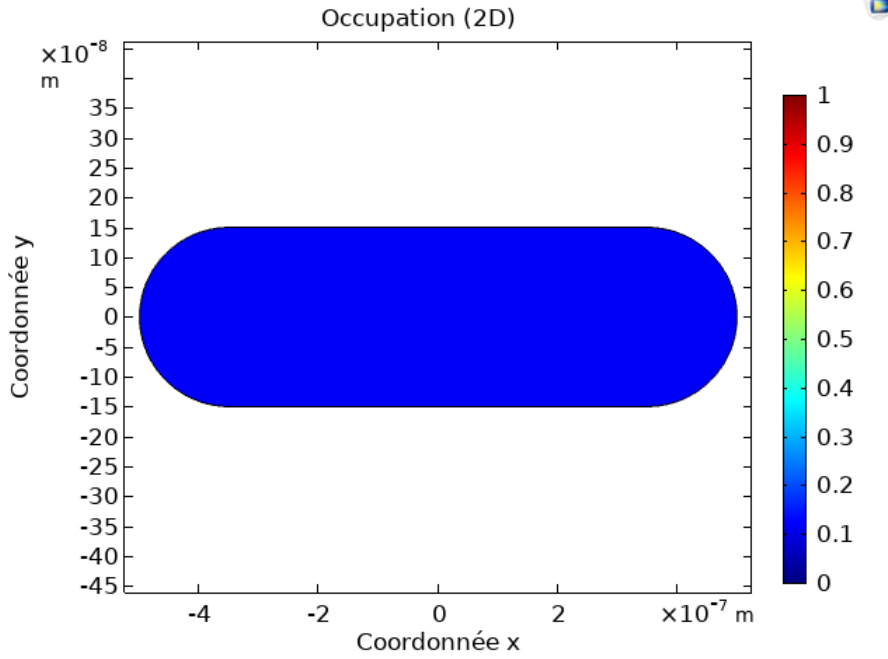


FIGURE 3.10 – **Profil d’occupation dans le régime de basse occupation pour $D = 2$:** sans séquence *parS*, la distribution d’équilibre est homogène en moyenne. Il s’agit également de la distribution stationnaire.

Les états stationnaires reproduisent qualitativement les résultats des expériences de SPT-PALM explicités dans le précédent chapitre : seul le régime de coexistence métastable montre des complexes ParBS et un espace cytoplasmique dépeuplé en protéines ParB avec une séquence *parS* et un profil homogène sans séquence *parS*.

Nous souhaitons soumettre notre méthode numérique à un examen plus précis. Il est possible de déterminer une forme analytique intégrable de l’énergie libre de Helmholtz aux environs du point critique ($\Phi_c = 1/2$, $T_c = Jq/(4k_B)$). Cette forme intégrable nous permettra de calculer l’expression analytique des interfaces entre états de haute et basse occupation et de la confronter aux résultats d’une simulation dans le cas unidimensionnel. En effet, avec un développement de Landau autour de la valeur Φ_c (voir Annexe 2 du Chapitre 1), nous pouvons écrire [83] :

$$\mathcal{F}_{1E} = \int_L -\frac{J}{2} l_m^2 \partial_x^2 \Phi + \frac{C}{2} (\Phi - \Phi_b)^2 (\Phi - \Phi_h)^2 \frac{dx}{l_m} \quad (3.77)$$

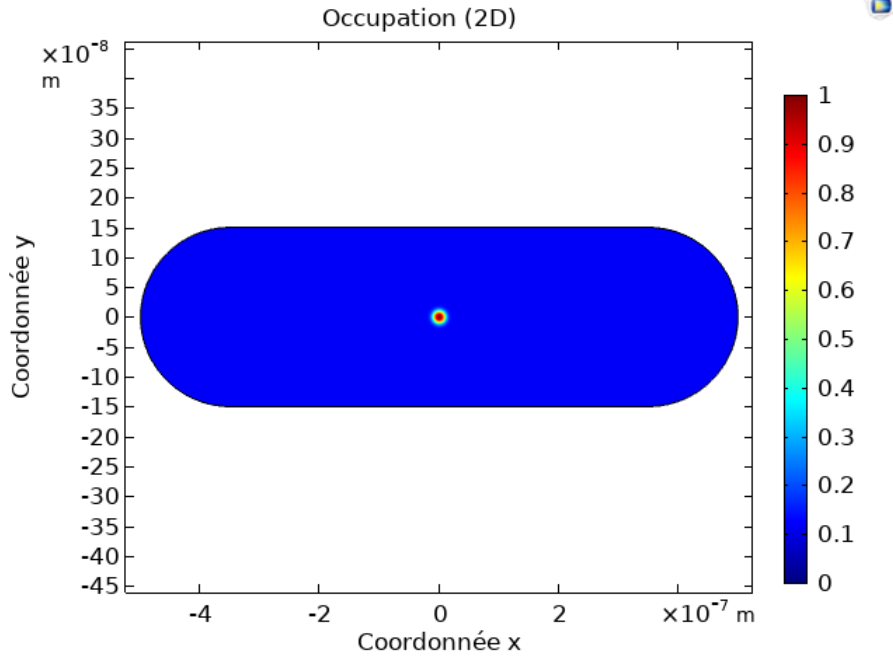


FIGURE 3.11 – **Profil d’occupation dans le régime de basse occupation pour $D = 2$** : avec une séquence *parS* au centre du réseau, la distribution des particules est quasiment homogène si ce n’est un maximum local autour de la séquence *parS*. Il s’agit également de la distribution stationnaire.

où $C = \frac{8k_B T}{3}$ et $\epsilon = 0$. Les occupations Φ_b et Φ_h sont les occupations de transition de la courbe de coexistence dans l’approximation de Landau et s’écrivent :

$$\Phi_b = \Phi_c - \frac{1}{4} \sqrt{\frac{3Jq}{k_B T} - 12} \quad (3.78)$$

$$\Phi_h = \Phi_c + \frac{1}{4} \sqrt{\frac{3Jq}{k_B T} - 12} \quad (3.79)$$

A l’équilibre thermodynamique, l’énergie libre \mathcal{F} est minimisée selon l’équation $\delta F / \delta \Phi = 0$ qui mène à l’égalité :

$$dx = \sqrt{\frac{Jl_m^2}{C}} \frac{1}{(\Phi - \Phi_b)(\Phi - \Phi_h)} d\Phi \quad (3.80)$$

Dans le cas le plus simple à une dimension, intégrer la précédente équation mène à l’expression du profil d’occupation suivant :

$$\Phi(x) = \frac{\Phi_b}{1 + \exp\left(-\frac{x}{\kappa}\right)} + \frac{\Phi_h}{1 + \exp\left(\frac{x}{\kappa}\right)} \quad (3.81)$$

pour une interface centrée en $x = 0$ et avec

$$\kappa = \sqrt{\frac{Jl_m^2}{C}} \frac{1}{\Phi_h - \Phi_b} \quad (3.82)$$

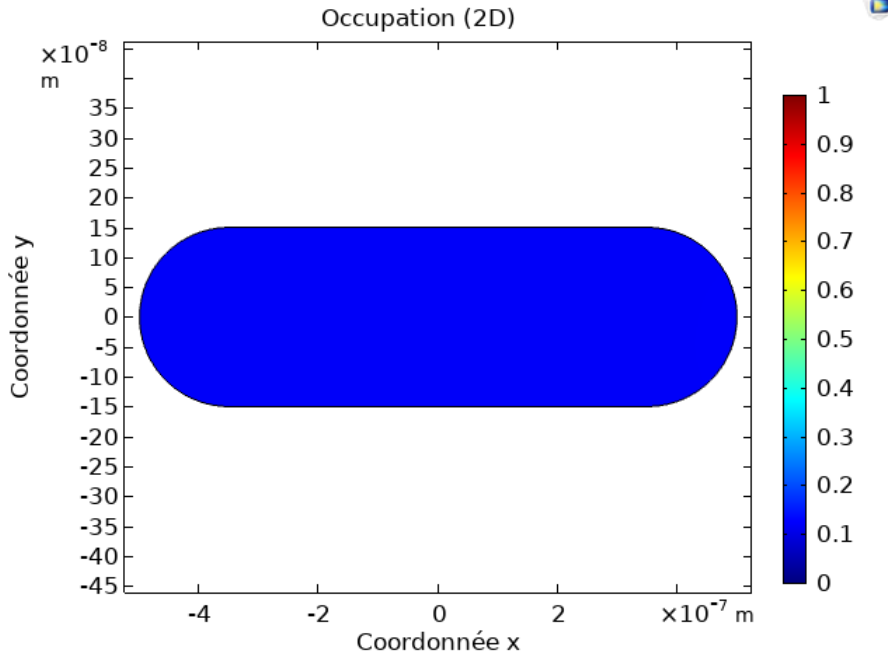


FIGURE 3.12 – **Profil d’occupation dans le régime de coexistence métastable pour $D = 2$** : sans séquence *parS*, la distribution d’équilibre est homogène en moyenne. Il s’agit de la distribution stationnaire dans ce cas.

La figure 3.17 confronte cette expression au résultat d’une simulation à une dimension dans un système de taille $L = 1\mu\text{m}$ centré en 0. Avec une occupation moyenne $\Phi_m = 0,5$ et un couplage $\tilde{J} = 2,1$, le système se situe dans une régime de coexistence instable très proche du point critique à une dimension ($\Phi_c = 0,5$, $\tilde{J}_c = 2$). Notre simulation est en excellent accord avec les résultats théoriques attendus dans cette limite.

3.4.2 Photoblanchiment *in silico* sans protéine ParA

Nous utilisons le modèle à plusieurs espèces pour reproduire les résultats des expériences de photoblanchiment. Nous utilisons une méthode très similaire à celle utilisée dans la partie sur le photoblanchiment *in silico* avec l’équation de Smoluchowski.

Méthode

Comme dans la méthode exposée pour l’équation de Smoluchowski, le photoblanchiment *in silico* avec ce modèle comprend trois étapes.

1. L’étape d’initialisation utilise modèle à un champ car nous ne faisons pas encore la différence entre les protéines ParB fluorescentes et photoblanchies. Soit $B(\mathbf{r}, t)$, l’occupation des protéines ParB sur l’ADN dans

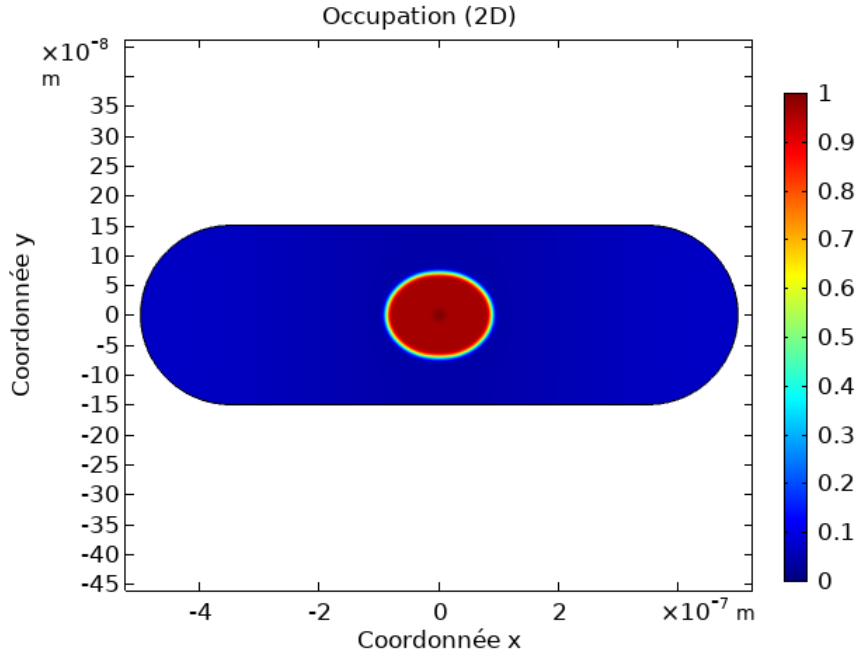


FIGURE 3.13 – **Profil d’occupation dans le régime de coexistence métastable pour $D = 2$** : avec une séquence *parS* au centre du réseau. Cette distribution montre un plateau en haute occupation bien localisée autour de la séquence *parS*.

l’espace cytoplasmique. Le champ $B(\mathbf{r}, t)$ est régie par l’équation :

$$\frac{\partial B}{\partial t} = \nabla \cdot \left[\tilde{L}_B(B) \nabla \mu_B(\mathbf{r}, t) \right]. \quad (3.83)$$

avec :

$$\mu_B = -\tilde{J}_{BB} l_m^2 \nabla^2 B - \tilde{J}_{BB} q B + \tilde{\epsilon}(\mathbf{r}) + \ln \left(\frac{B}{1-B} \right) \quad (3.84)$$

où $\tilde{J}_{BB} = \beta J_{BB}$ est l’énergie d’interaction entre protéines ParB adsorbées sur des sites premiers voisins et s’apparente donc au paramètre J utilisé dans le Chapitre 2, et $\tilde{\epsilon} = \beta \epsilon$. La fonction $\tilde{L}_B(B)$ tient compte des deux limites connues en haute D_L^B et basse occupation D_V^B du coefficient de diffusion collectif des protéines ParB :

$$\tilde{L}_B(B) = \frac{D_V^B D_L^B B(1-B)}{D_V^B B + D_L^B (1-B)}. \quad (3.85)$$

Avec une première simulation dont le profil initial est homogène, nous produisons un profil stationnaire pour deux séquences *parS* aux positions $1/4$ et $3/4$ de l’espace d’intégration, c’est-à-dire que :

$$\epsilon(\mathbf{r}) = \varepsilon \left\{ \exp \left[-\frac{(\mathbf{r} - \mathbf{r}_1)^2}{2l_m^2} \right] + \exp \left[-\frac{(\mathbf{r} - \mathbf{r}_2)^2}{2l_m^2} \right] \right\} \quad (3.86)$$

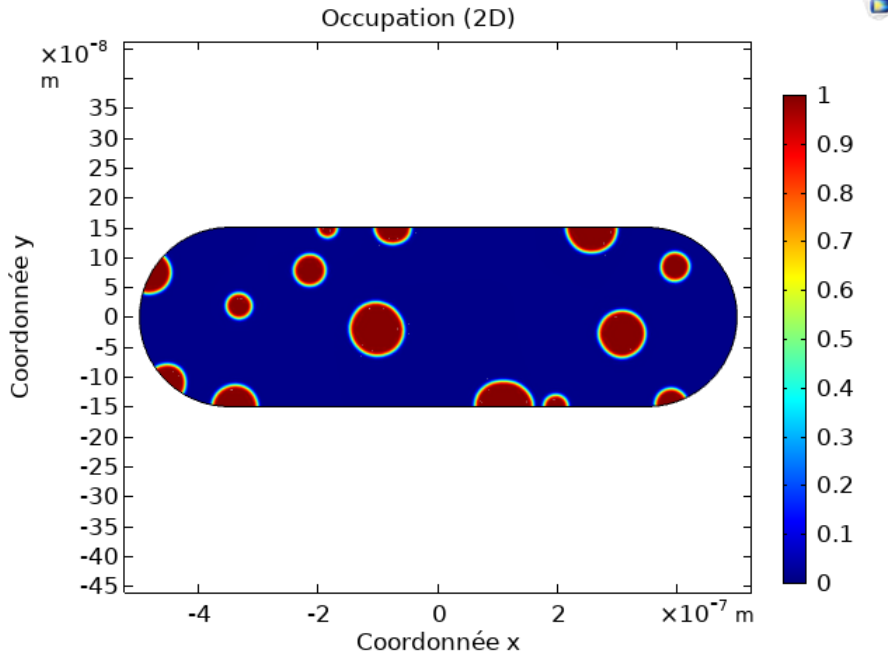


FIGURE 3.14 – **Profil d’occupation dans le régime de coexistence instable pour $D = 2$** : sans séquence *parS*, la distribution indique plusieurs plateaux en haute occupation (en rouge foncé). La séparation de phase a bien lieu. A $t = 1$ s, le système subit encore une maturation d’Ostwald. La distribution stationnaire d’une telle configuration est donnée par la figure 3.16 sans le complexe minoritaire correspondant à l’occupation locale autour de la séquence *parS*.

avec $\mathbf{r}_1 = (1/4, 0, 0) \mu\text{m}$ et $\mathbf{r}_2 = (-1/4, 0, 0) \mu\text{m}$, les positions du complexe *ParBS* en coordonnées cartésiennes.

2. L’étape de photoblanchiment va engendrer deux profils d’occupation : le premier pour l’espèce photoblanchie, le second pour l’espèce fluorescente. Soient $B_F(\mathbf{r}, t)$ et $B_P(\mathbf{r}, t)$ les occupations des protéines *ParB* fluorescentes et photoblanchies. Si on note $B_s(\mathbf{r})$, le profil stationnaire de $B(\mathbf{r}, t)$, nous avons alors :

$$B_F(\mathbf{r}, t = 0) = \begin{cases} \gamma B_s & \text{si } x \leq 0 \\ \gamma_F B_s, & \text{si } x > 0 \end{cases} \quad (3.87)$$

d’une part, et d’autre part :

$$B_P(\mathbf{r}, t = 0) = \begin{cases} (1 - \gamma) B_s & \text{si } x \leq 0 \\ (1 - \gamma_F) B_s, & \text{si } x > 0. \end{cases} \quad (3.88)$$

où $\gamma_F = 0,4$ est le taux de fluorescence restante dans le complexe photoblanchi, et γ un taux très proche de 1 ($\gamma = 0,99$ dans nos simulations)

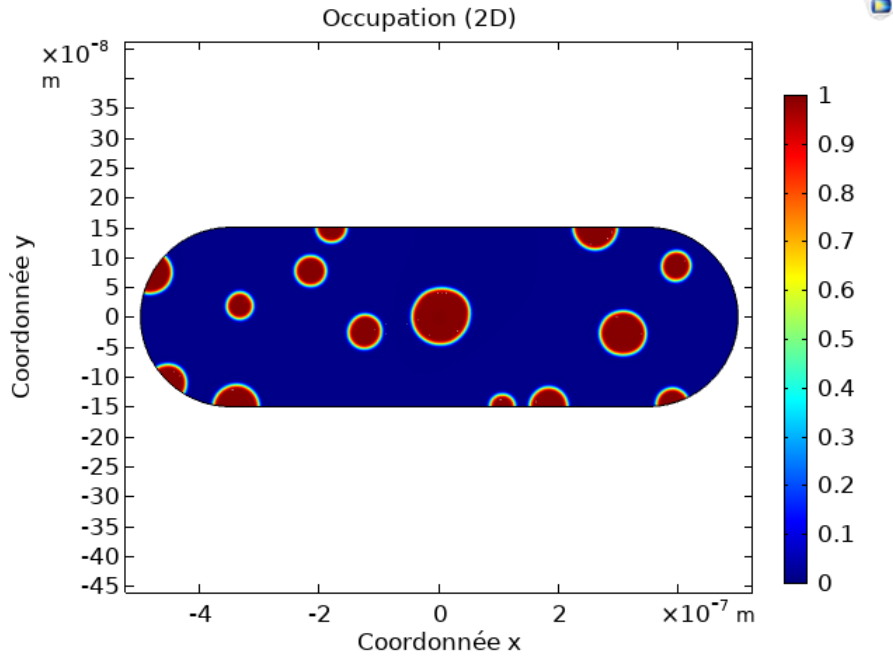


FIGURE 3.15 – **Profil d’occupation dans le régime de coexistence instable pour $D = 2$** : avec une séquence *parS* au centre du réseau. Nous observons plusieurs plateaux en haute occupation (comme dans le cas sans séquence *parS*) qui indique que la séparation de phase a lieu. Le système subit une maturation d’Ostwald qui mène à la distribution stationnaire en figure 3.16.

empêchant au champ B_P de prendre des valeurs nulles, interdites par les termes entropiques dans notre approche numérique qui découple les équations pour B et μ_B .

3. Pour l’étape de recouvrement de la fluorescence, nous utilisons les équations couplées suivantes, issues du modèle de gaz sur réseau à deux espèces dans le cas de l’exclusion locale :

$$\frac{\partial B_F}{\partial t} = \nabla \cdot \left[\tilde{L}_B(B_F + B_P) \nabla \mu_F \right] \quad (3.89)$$

$$\frac{\partial B_P}{\partial t} = \nabla \cdot \left[\tilde{L}_B(B_F + B_P) \nabla \mu_P \right]. \quad (3.90)$$

avec d’une part part, le potentiel chimique lié à l’espèce fluorescente :

$$\begin{aligned} \mu_F = & -\tilde{J}_{BB} l_m^2 \nabla^2 (B_F + B_P) - \tilde{J}_{BB} q (B_F + B_P) + \tilde{\epsilon}(\mathbf{r}) \\ & + \ln \left(\frac{B_F}{1 - B_F - B_P} \right) \end{aligned} \quad (3.91)$$

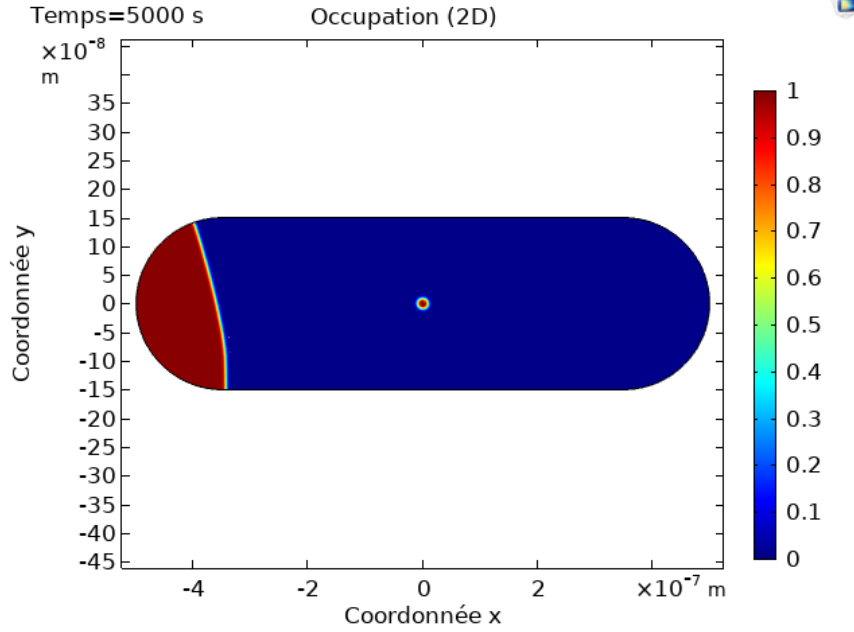


FIGURE 3.16 – **Profil d’occupation dans le régime de coexistence instable pour $D = 2$** : avec une séquence *parS* au centre du réseau aux temps longs, ici $t = 5000$ s. Les protéines ParB forment un complexe sans séquence *parS* au bord de l’espace d’intégration, où la courbure de la cellule est la plus importante.

et d’autre part, le potentiel chimique lié à l’espèce photoblanchie :

$$\begin{aligned} \mu_P = & -\tilde{J}_{BB}l_m^2 \nabla^2 (B_F + B_P) - \tilde{J}_{BB}q(B_F + B_P) + \tilde{\epsilon}(\mathbf{r}) \\ & + \ln \left(\frac{B_P}{1 - B_F - B_P} \right) \end{aligned} \quad (3.92)$$

Les conditions aux bords de Neumann liées à ces équations couplées sont :

$$\mathbf{n} \cdot \nabla \mu_F(\mathbf{r}, t)|_{\text{bords}} = \mathbf{n} \cdot \nabla \mu_P(\mathbf{r}, t)|_{\text{bords}} = 0 \quad (3.93)$$

$$\mathbf{n} \cdot \nabla B_F(\mathbf{r}, t)|_{\text{bords}} = \mathbf{n} \cdot \nabla B_P(\mathbf{r}, t)|_{\text{bords}} = 0 \quad (3.94)$$

Il est difficile d’obtenir une forme analytique des coefficients d’Onsager lorsque deux espèces, ou plus, évoluent sur le même réseau. Néanmoins, le coefficient $\tilde{L}_B(B_F + B_P)$ possède les limites attendues pour les deux sous-espèces des protéines ParB et leur comportement collectif.

Dans la partie suivant, nous allons constater que ce modèle est incomplet en l’état et ne permet pas d’interpréter les résultats des expériences de FRAP.

Résultats

En appliquant la méthode de photoblanchiment *in silico* exposée précédemment, nous parvenons à reproduire les courbes de récupération ou de perte de la

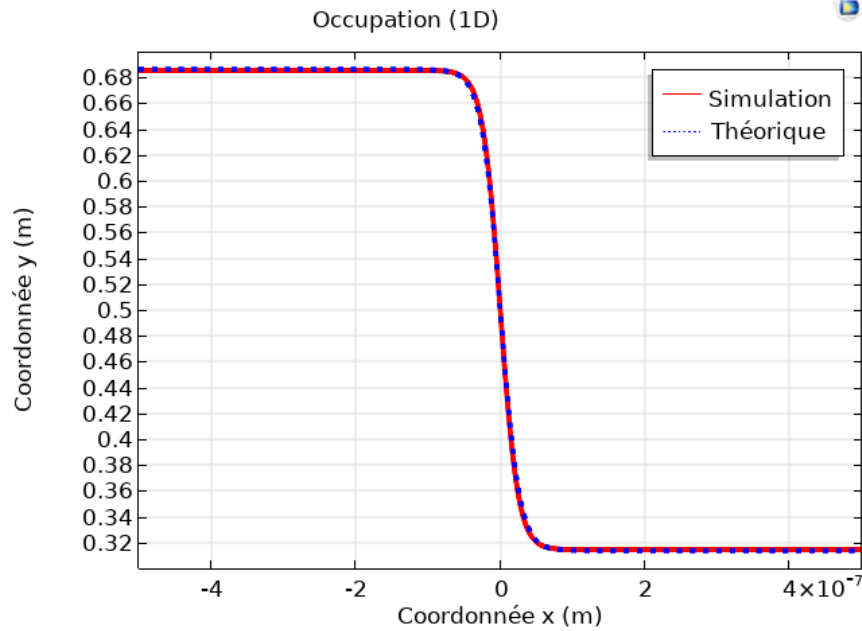


FIGURE 3.17 – **Profil d’une interface unidimensionnelle** entre états de haute (à gauche) et basse occupation (à droite). En rouge et trait plein, les résultats d’une simulation ; en bleu et pointillé, les résultats théoriques.

fluorescence à une dimension (voir figure 3.18) pour l’ensemble des paramètres indiqués dans le tableau 3.1. Nous devons choisir une énergie spécifique d’adsorption ε de $4 k_B T$ pour éviter que l’occupation locale autour de $parS$ n’excède numériquement la valeur 1 par instabilité numérique. Remarquons que les protéines ParB existent dans un régime de coexistence métastable, conformément aux conclusions du chapitre 1. Cet ensemble de paramètres, et notamment \tilde{J}_{BB} et l’occupation moyenne des protéines ParB $\Phi_m = 0,0815$, vérifie les conditions déjà établies dans le cadre des équations de Smoluchowski, c’est-à-dire que 90% des protéines existent dans deux foyers de diamètre $l_f = 0,037 \mu m$. La valeur du coefficient de diffusion de la phase dense doit être plus faible de deux ordres de grandeurs que la valeur mesurée $D_B^L = 0,02 \mu m^2/s$. Les mêmes équations doivent être intégrées à des dimensions supérieures en guise de confirmation. A deux dimensions, nous observons un type de résultats très différent : la création de deux complexes stables dans la phase d’initialisation ou leur survie dans la phase de récupération sont empêchées par un mécanisme de mûrissement d’Ostwald [126]. Il est clair que nos équations, seules, mènent à la formation d’un unique complexe, de préférence à la périphérie du système, comme on peut l’observer dans les séparations de phase classiques (voir sous-section Résultats du modèle à un champ). Dans le scénario métastable, s’il existe deux séquences $parS$ à différentes positions dans l’espace intracellulaire, l’équilibre thermodynamique correspond à un complexe le plus grand

Paramètre	Symbole	Valeur
Coordination du réseau	q	2
Longueur d'un monomère	l_m	16 pb ($\sim 5,4$ nm)
Coefficient de la phase basse occupation	D_V^B	$0,5\mu m^2 s^{-1}$
Coefficient de la phase haute occupation	D_L^B	$7,5 \cdot 10^{-5}\mu m^2 s^{-1}$
Couplage entre premiers voisins	J_{BB}	$4,809k_B T_r$
Température physiologique	T_r	300K
Longueur du réseau unidimensionnelle	L	$1\mu m$
Energie d'adsorption spécifique maximale	ε	$-4k_B T_r$
Occupation moyenne	Φ_m	0,0815

TABLE 3.1 – Paramètres utilisés pour la photoblanchiment *in silico* sans les protéines ParA.

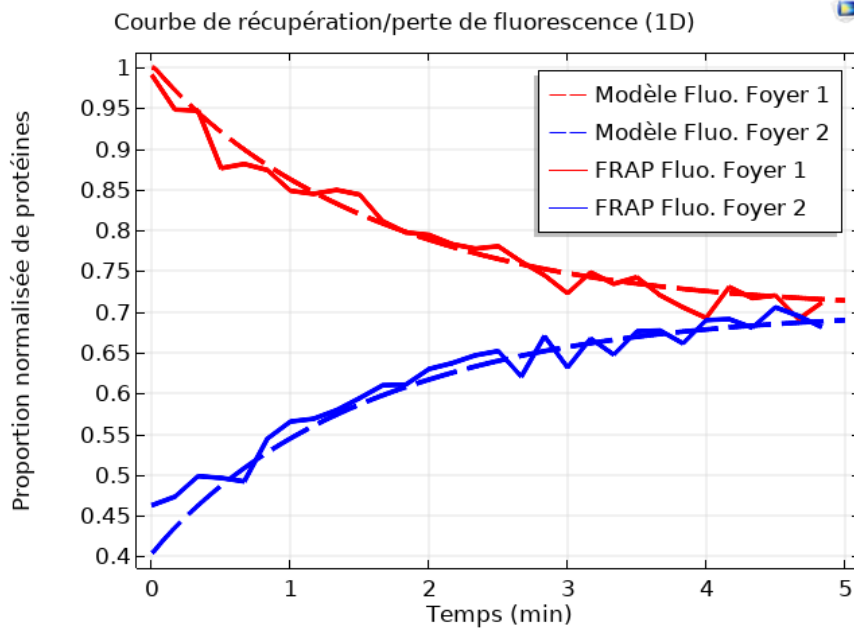


FIGURE 3.18 – Courbe de recouvrement *in silico* pour l'approche fondamentale sans protéines ParA, à une dimension. En trait plein : courbes de recouvrement en bleu et perte en rouge issues des expériences. En trait discontinu : courbes de recouvrement en bleu et perte en rouge issues des simulations.

possible autour d'une première séquence *parS* tandis que la deuxième recrute quelques protéines ParB de manière très minoritaire. Ce résultat implique deux conclusions :

1. Que la dimension 1 puisse nous permettre de créer deux complexes ParBS stables doit nous convaincre que cette dimension échoue à décrire correctement notre système biologique. Elle peut être utilisée en guise d'illustration en gardant à l'esprit qu'elle sous-estime très gravement l'import-

tance de la diffusion et donc des échanges en protéines ParB entre les complexes et l'espace cytoplasmique.

2. L'existence expérimentale de deux complexes ParBS indique que le système biologique se maintient hors de l'équilibre thermodynamique. Les protéines ParA, troisième élément du système de partition ParABS, pouvant hydrolyser les molécules d'ATP, pourraient être l'ingrédient permettant au système biologique de se maintenir hors de l'équilibre. Un modèle simple a déjà pu montrer le rôle crucial de ParA dans le positionnement des complexes ParBS.

Néanmoins, nous pouvons extraire un temps typique d'existence d'un complexe ParBS sans ParA de nos simulations pour plusieurs valeurs de paramètre dans le but de discriminer certaines tendances. Pour les mêmes paramètres que dans le tableau 3.1 et le même potentiel ϵ aux positions 1/4-3/4, nous observons dans un premier temps la formation de deux gouttes équilibrées. Elles se déséquilibrent par la suite et l'état stationnaire correspond à une goutte, dite majoritaire, qui contient la majorité des protéines ParB, et une goutte minoritaire qui persiste grâce à l'adsorption spécifique provoquée par ϵ . Nous définissons le temps typique τ_s de survie d'un complexe ParBS sans ParA comme le temps entre le début du déséquilibre et cet état final. Ce temps typique correspond à un temps d'échange de protéines ParB du foyer minoritaire au foyer majoritaire et donne un indice sur le temps typique d'échange des protéines fluorescentes dans les expériences de photoblanchiment. Nous rappelons que le temps de recouvrement de la fluorescence dans le cas des complexes ParBS est de l'ordre de 300 s.

- Le temps τ_s augmente avec l'occupation moyenne Φ_m utilisée en tant qu'occupation homogène initiale. Plus le nombre de protéine à transférer est important, plus le temps de transfert est long (voir figure 3.19).
- Il augmente aussi avec le couplage \tilde{J}_{BB} . Il est plus difficile pour une protéine ParB de quitter le foyer minoritaire quand \tilde{J}_{BB} est grand car les liaisons entre ParB sont plus fortes (voir figure 3.20).
- Il diminue par contre avec le coefficient de diffusion de la phase condensée. Plus le phénomène collectif d'emprisonnement des protéines ParB est important, plus le coefficient de diffusion de la phase dense est faible. Il faut ainsi plus de temps à un ParB pour se libérer du complexe minoritaire (voir figure 3.21).

Notons que, sur les gammes de valeurs testées, un temps typique de l'ordre de quelques minutes comparable au temps de recouvrement de la fluorescence en photoblanchiment n'a été observé qu'avec un coefficient de diffusion de phase condensée faible. Les tendances pour \tilde{J}_{BB} et Φ_m montrent qu'augmenter les valeurs de ces paramètres ne suffisent pas pour obtenir des temps typiques de plusieurs minutes, d'autant plus que \tilde{J}_{BB} et Φ_m doivent être relativement faibles, conformément aux faisceaux d'indices explicités en chapitre 1.

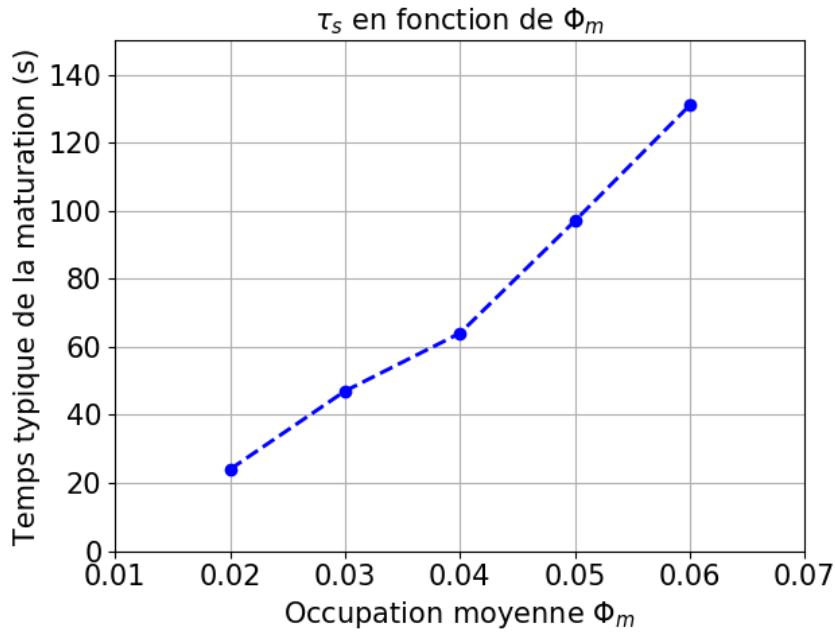


FIGURE 3.19 – Temps de survie d’une goutte en fonction de l’occupation moyenne.

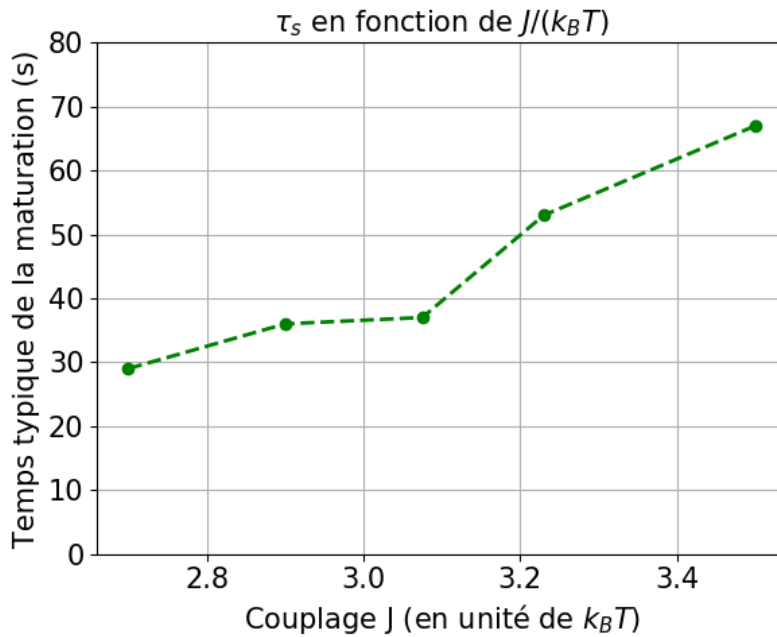


FIGURE 3.20 – Temps de survie d’une goutte en fonction du couplage entre premiers voisins.

Pour outrepasser le problème de l’équilibre des complexes $ParBS$, nous proposons d’ajouter à notre modèle deux nouvelles équations pour inclure dans la modélisation $ParA$ et son cycle biochimique. Dans la suite, nous nous restrei-

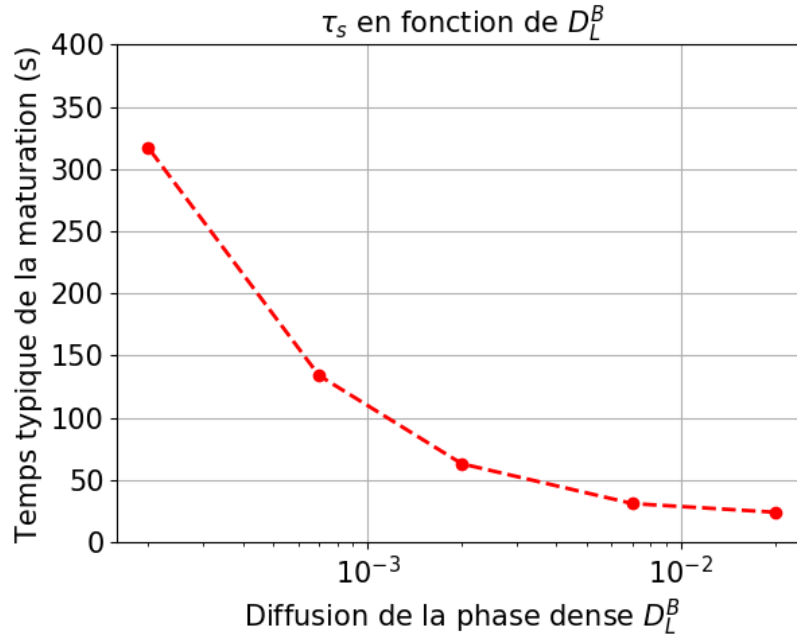


FIGURE 3.21 – Temps de survie d’une goutte en fonction du coefficient de diffusion de la phase condensée.

gnons au cas à deux dimensions. Nous cherchons à illustrer nos propositions en apportant des preuves de faisabilité reposant sur des valeurs de paramètres raisonnables. Le cas à deux dimensions sous-estime aussi les effets de la diffusion par rapport au cas réel à trois dimensions. Ainsi, une discussion sur les valeurs des paramètres ne saurait être rigoureuse que dans le cas à trois dimensions, difficile à réaliser techniquement à cause de temps de calcul longs. Dans le cas du modèle à deux espèces, nous avons déjà quatre équations couplées. Pour les modèles à trois et quatre champs introduits dans la section suivante, nous aurons six puis huit équations couplées à résoudre. Le temps de calcul augmente non-linéairement avec le nombre de champ sur le réseau.

3.4.3 Equilibre des complexes ParBS

Dans cette partie, nous complétons l’approche fondamentale du gaz sur réseau à deux espèces en prenant en compte les protéines ParA sur le réseau d’ADN. Nous utilisons ce nouveau modèle pour montrer que l’hydrolyse des protéines ParA permet d’assurer l’équilibre des complexes ParBS lorsque plusieurs séquences *parS* existent dans l’espace intracellulaire en activant ou inhibant à propos l’hydrolyse de l’ATP. Nous montrons enfin que l’ajout des protéines ParA et de l’hydrolyse de l’ATP permet de retrouver les courbes expérimentales de photoblanchiment.

Méthode

Nous simplifions le cycle biochimique de la protéine ParA [53, 41]. De ses différents états moléculaires, nous en retenons deux principaux : une forme dite ATP, lente car adsorbée sur l'ADN, et une forme dite ADP, rapide, pouvant diffuser librement dans l'espace intracellulaire. Ainsi, nous introduisons deux nouvelles espèces : les protéines ParA sous la forme ATP, modélisées par un champ d'occupation $A_T(\mathbf{r}, t)$, et les ParA sous forme ADP, de champ de concentration normalisée $A_D(\mathbf{r}, t)$. Ces deux espèces possèdent des coefficients de diffusion bien distincts [54]. Le coefficient de diffusion de la forme rapide, ParA-ADP, est $D_V^A = 1 \mu\text{m}^2/\text{S}$ et le coefficient de diffusion de la forme lente, ParA-ATP, est $D_L^A = 10^{-2} \mu\text{m}^2/\text{S}$, c'est-à-dire que la forme lente, liée à l'ADN, diffuse une centaine de fois moins vite que la forme rapide. Notre nouveau système d'équations couplées est :

$$\frac{\partial B}{\partial t} = \nabla \left(\tilde{L}_B(B) \nabla \mu_B \right) \quad (3.95)$$

$$\frac{\partial A_T}{\partial t} = \nabla \left(\tilde{L}_A(A_T) \nabla \mu_A \right) + k_1 A_D - k_2 A_T B \quad (3.96)$$

$$\frac{\partial A_D}{\partial t} = D_V^A \nabla^2 A_D - k_1 A_D + k_2 A_T B \quad (3.97)$$

avec μ_A le potentiel chimique lié aux protéines ParA-ATP sur l'ADN :

$$\begin{aligned} \mu_A = & -\tilde{J}_{AA} l_m^2 \nabla^2 A_T - \tilde{J}_{AA} q A_T - \tilde{J}_{AB} l_m^2 \nabla^2 B - \tilde{J}_{AB} q B \\ & + \ln \left(\frac{A_T}{1 - A_T - B} \right) \end{aligned} \quad (3.98)$$

et μ_B , le nouveau potentiel chimique lié au protéines ParB :

$$\begin{aligned} \mu_B = & -\tilde{J}_{BB} l_m^2 \nabla^2 B - \tilde{J}_{BB} q B - \tilde{J}_{AB} l_m^2 \nabla^2 A_T - \tilde{J}_{AB} q A_T \\ & + \tilde{\epsilon}(\mathbf{r}) + \ln \left(\frac{B}{1 - A_T - B} \right) \end{aligned} \quad (3.99)$$

Notons que l'équation (3.97) pour l'espèce ParA-ADP est une simple équation de réaction-diffusion avec un terme de diffusion en $\nabla^2 A_D$ provenant d'une loi de Fick de diffusion dans l'espace intracellulaire, car cette espèce diffuse directement dans le cytoplasme et non pas sur l'ADN. Nous prenons en compte les interactions entre protéines ParA avec un couplage J_{AA} et les interactions entre protéines ParA et ParB d'amplitude J_{AB} . Nous considérons ainsi que les protéines ParA sont aussi capable de séparation de phase, ce qui pourrait donner l'origine des patches de ParA observés expérimentalement [54]. Le cycle biochimique simplifié des protéines ParA entre dans la modélisation au travers des termes d'échange suivants [53] :

- le terme d'hydrolyse $k_2 A_T B$. Les protéines ParB sont des catalyseurs des réactions d'hydrolyse de l'ATP par les protéines ParA. La production de ParA-ADP ou l'hydrolyse des ParA-ATP a donc lieu principalement autour des complexes ParB*S* avec un taux k_2 , estimé à $k_2 = 68,5 \text{ s}^{-1}$ [127]. Les protéines Par-ATP sont liées à l'ADN : l'hydrolyse de l'ATP provoque un changement de conformation de la protéine qui se libère de l'ADN.
- le terme de recharge de l'ATP $k_1 A_T$. Le recharge en ATP des protéines ParA a lieu n'importe où dans l'espace cytoplasmique avec un taux k_1 , estimé à $k_1 = 0,02 \text{ s}^{-1}$ [53].

De manière rigoureuse, la mobilité des protéines ParA doit dépendre de l'occupation de protéines ParB, et réciproquement. Néanmoins, nous ne possédons en l'état d'aucune information sur la diffusion d'un type de protéine dans une phase d'une protéine d'un autre type. Ainsi, par souci de simplification, nous découplons dans un premier temps les mobilités des espèces ParA et ParB. Le coefficient $\tilde{L}_A(A_T)$ prend en compte les deux limites du coefficient collectif de diffusion des protéines ParA sous forme ATP sur le réseau d'ADN, basse occupation D_V^A et haute occupation D_L^A :

$$\tilde{L}_A(A_T) = \frac{D_V^A D_L^A A_T (1 - A_T)}{D_V^A A_T + D_L^A (1 - A_T)} \quad (3.100)$$

On considère que les protéines ParA sous forme ADP diffuse avec le même coefficient de diffusion que les protéines ParA sur le réseau dans la limite basse occupation.

Dans nos simulations, nous pouvons activer ou désactiver à des instants prédéfinis des propriétés chimiques particulières du système biologique. Cette possibilité s'apparente à la création de mutant dans les expériences de biologie. Nous utilisons cette astuce pour montrer dans la sous-section suivante que les protéines ParA peuvent être à l'origine de l'équilibre des complexes ParB*S*.

Résultats

Les valeurs de paramètres utilisées sont indiquées dans le tableau 3.2. La valeurs J_{BB} des interactions en protéines ParB peut sembler élevée : notre précédente estimation (voir chapitre 1) était de l'ordre de $1 k_B T$. Or, nous avons vu que le couplage critique en champ moyen dépend de la dimensionalité du système. A deux dimensions, il faut des valeurs de couplage plus élevées qu'à trois dimensions pour exister dans le régime de coexistence métastable et obtenir les mêmes occupations à l'intérieur et à l'extérieur des complexes ParB*S*. Ainsi, la valeur relativement grande de J_{BB} compense le plus faible nombre d'interactions possibles entre premiers voisins sur le réseau dû à la basse dimensionalité $D = 2$ ($q = 4$). Une simulation à trois dimensions nécessiterait un

Paramètre	Symbole	Valeur
Coordination du réseau	q	4
Longueur d'un monomère	l_m	16 pb ($\sim 5,4$ nm)
Coefficient de basse occupation des ParB	D_V^B	$0,5\mu m^2 s^{-1}$
Coefficient de haute occupation des ParB	D_L^B	$0,002\mu m^2 s^{-1}$
Coefficient de basse occupation des ParA	D_V^A	$1\mu m^2 s^{-1}$
Coefficient de haute occupation des ParA	D_L^A	$0,01\mu m^2 s^{-1}$
Couplage entre ParB premiers voisins	J_{BB}	$2,5k_B T_r$
Couplage entre ParA premiers voisins	J_{AA}	$2k_B T_r$
Couplage entre ParA et ParB	J_{AB}	$2k_B T_r$
Température physiologique	T_r	300K
Longueur principale du système	L	$1\mu m$
Longueur secondaire du système	l	$0,3\mu m$
Energie d'adsorption spécifique maximale	ε	$-4k_B T_r$
Occupation moyenne des ParB	Φ_B	0,02
Occupation moyenne initial des ParA-ATP	Φ_A	$2\Phi_B$
Taux d'hydrolyse de l'ATP	k_2	$100s^{-1}$
Taux de recyclage de l'ADP	k_1	$10s^{-1}$

TABLE 3.2 – Paramètres utilisés pour la photoblanchiment *in silico* avec les protéines ParA.

couplage J_{BB} plus faible. Nous pouvons faire la même remarque pour les valeurs des couplages J_{AA} et J_{AB} . L'occupation moyenne des protéines ParB est fixée à $\Phi_B = 0,02$ en basse occupation et de telle sorte que les protéines ParB existent dans un régime de coexistence métastable. L'occupation moyenne initiale Φ_A des protéines ParA-ATP respectent le ratio entre protéines ParA et ParB [22, 54] :

$$\frac{\Phi_A}{\Phi_B} = 2. \quad (3.101)$$

Le taux k_1 de rechargement des protéines ParA en ATP estimé expérimentalement est considérablement plus faible que celle utilisée dans notre simulation. Cette différence peut s'expliquer par notre faible dimensionalité : la recharge de l'ATP est beaucoup moins efficace sur une surface pure que dans un volume. Les autres paramètres sont dans leur gamme attendue de valeurs.

Nous avons produit une simulation pour laquelle la deuxième séquence *parS* à la position $x = 0,25\mu m$ n'est pas activée avant $t = 10s$ et l'hydrolyse de l'ATP par les protéines ParA, contrôlée par le paramètre k_2 , n'est pas activée avant $t = 50s$. Les résultats sont présentés dans un ordre chronologique et nous représentons les valeurs des champs de protéines selon l'axe principale de la bactérie :

- Figure 3.22 : Avant $t = 0$, les protéines ParA sous forme ATP et ParB

sont réparties de manière homogène sur le réseau tandis qu'il n'y a pas de protéines ParA sous forme ADP, soit $B(x, t < 0) = \Phi_B$, $A_T(x, t < 0) = 2\Phi_B$, et $A_D(x, t < 0) = 0$. A l'instant initial, nous activons une seule séquence $parS$ à la position 1/4 de la bactérie. L'hydrolyse de l'ATP par les protéines ParA est désactivée ($k_2 = 0$).

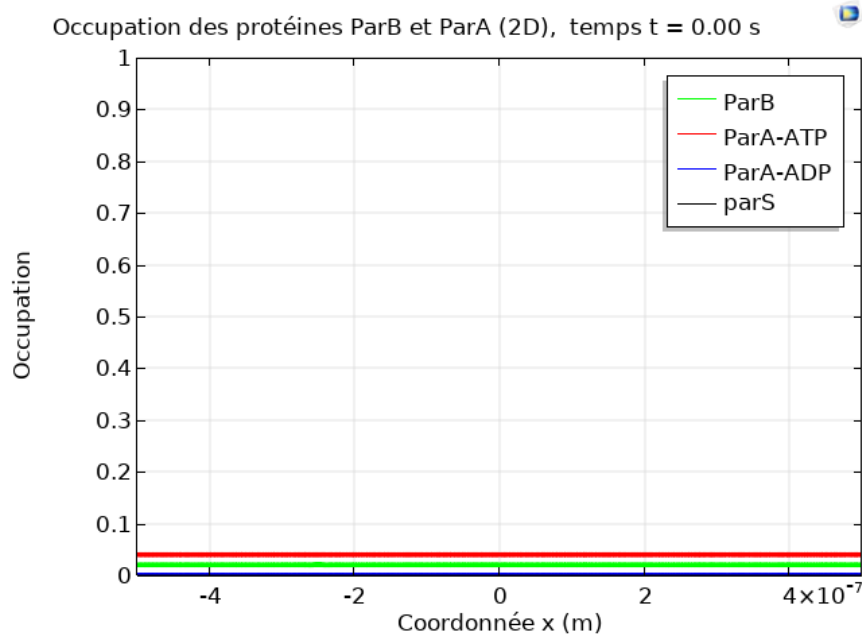


FIGURE 3.22 – **Profil des occupations pour $t < 0$ s** des protéines ParB (en vert) et protéines ParA sous forme ATP (en rouge) et protéines ParA-ADP (en bleu) selon l'axe principale de la bactérie. Les séquences $parS$ ne sont pas activées. L'hydrolyse est désactivée.

- Figure 3.23 : Dès les premiers instants $t = 0^+s$, les sites d'adsorption spécifique ($parS$) recrutent des protéines ParB : un complexe ParBS commence à se former. Les interactions attractives entre protéines ParA-ATP et ParB provoquent le recrutement des protéines ParA-ATP par le complexe ParBS. Les protéines ParA subissent, comme les protéines ParB, une séparation de phase. Ces recrutements successifs aboutissent à la formation d'un complexe où se mélangent les protéines ParA-ATP et ParB autour de la séquence $parS$.
- Figure 3.24 : A l'instant $t = 10s$, nous activons une seconde séquence $parS$ à la position 3/4 de l'espace d'intégration. Ce second exemplaire de ϵ , de même intensité que le premier, recrute de manière minoritaire quelques protéines ParB et, par conséquent, ParA-ATP. A l'issue de ce recrutement, la grande majorité des protéines ParB sont localisées autour de la première séquence $parS$. Nous montrons ainsi que l'ajout des protéines ParA-ATP ne suffit pas à provoquer l'équilibrage des complexes ParBS : l'équilibre thermodynamique correspond toujours à un

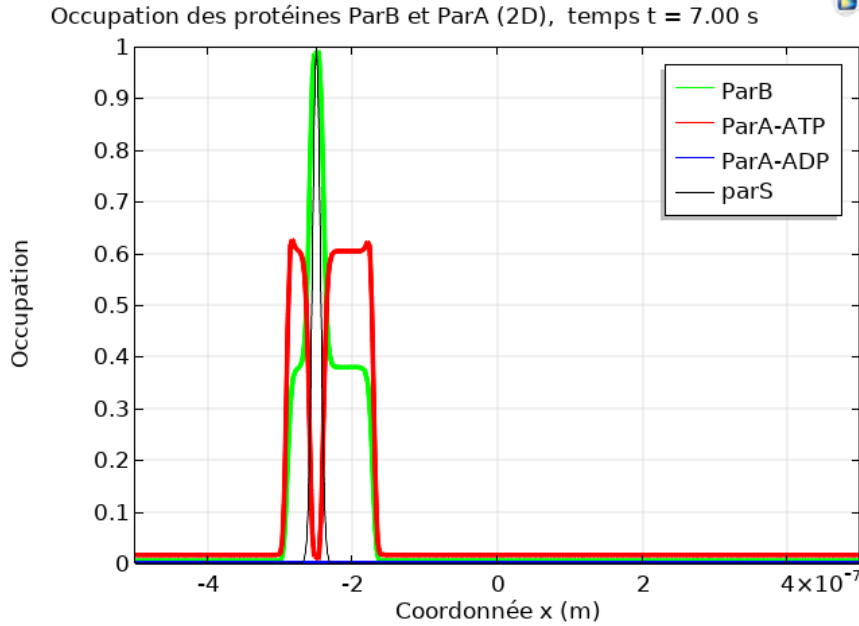


FIGURE 3.23 – **Profil des occupations pour $t = 0^+$** des protéines ParB (en vert) et protéines ParA sous forme ATP (en rouge) et protéines ParA-ADP (en bleu) selon l'axe principale de la bactérie. Les séquences *parS* activées sont en trait noir et fin : une seule séquence est active. L'hydrolyse de l'ATP n'est pas activée. On observe la formation d'un complexe ParBS colocalisé avec un complexe de protéines ParA-ATP, formant un complexe ParBS+ParA-ATP.

seul complexe ParBS+ParA-ATP.

- Figure 3.25 : A $t = 50s$, nous activons l'hydrolyse de l'ATP ($k_2 > 0$). En premier lieu, nous observons une dynamique complexe interne au complexe ParBS+ParA-ATP : les protéines ParA, capturées dans la structure du complexe ParBS grâce aux interactions ParA-ParB, hydrolysent leur ATP, se décroche de l'ADN et peuvent diffuser librement dans le cytoplasme. Nous observons la formation de protéines ParA sous forme ADP autour du complexe ParBS majoritaire. Ces protéines ParA-ADP diffusent ensuite progressivement dans tout l'espace d'intégration où elles sont reconverties sous forme ATP.
- Figure 3.26 : A partir de $t = 50s$, les protéines ParA-ATP sont recrutées par le complexe ParBS minoritaire. Ce deuxième complexe, étant plus petit, hydrolyse moins efficacement l'ATP des protéines ParA sur l'ADN. Il est donc plus favorable à un ParA-ATP d'exister dans le second complexe. Dans les faits, les protéines ParA-ATP enveloppent le complexe ParBS à cause de l'hydrolyse maximale au centre du complexe et de l'exclusion locale entre champs prévues par nos équations.
- Figure 3.27 : Après l'instant $t = 51s$, les protéines ParA-ATP localisées

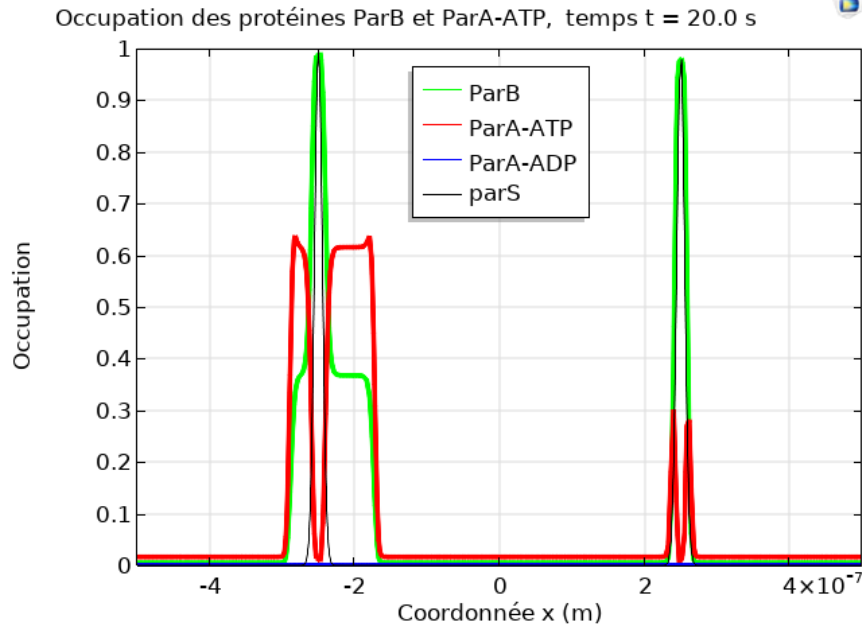


FIGURE 3.24 – **Profil des occupations pour $t = 20$ s** des protéines ParB (en vert) et protéines ParA sous forme ATP (en rouge) et protéines ParA-ADP (en bleu) selon l'axe principale de la bactérie. Les séquences *parS* activées sont en trait noir et fin : deux séquences *parS* sont activées. L'hydrolyse de l'ATP n'est pas encore activée. Un deuxième complexe ParB+ParA-ATP minoritaire se forme autour de la deuxième séquence *parS*.

dans le second complexe recrutent davantage de protéines ParB. On observe alors un transfert des protéines ParB du complexe majoritaire vers le complexe minoritaire alors que les protéines ParA y sont davantage hydrolysées. Le système tend progressivement vers un équilibre en terme de nombre de protéines ParA et ParB dans les deux complexes ParB.

Ces résultats sont résumés dans la figure 3.28. Il apparaît que les interactions entre protéines ParA-ATP et ParB alliées à l'hydrolyse de l'ATP autour des complexes ParB permettent d'équilibrer les complexes ParB. Les complexes ParB entretiennent des gradients de protéines ParA-ATP autour d'eux grâce à une forte hydrolyse localisée au même endroit, et à la diffusion rapide des protéines ParA-ADP. Ces gradients forment des réservoirs de protéines ParA-ATP capables d'initier un mécanisme de recrutement si une nouvelle séquence *parS* est générée dans l'espace cytoplasmique (voir figure 3.29). Nous pouvons en déduire que les complexes ParB s'entretiennent entre eux. L'hydrolyse de l'ATP dans le cadre du système ParABS permet donc d'inverser la maturation d'Ostwald [113, 114]. L'idée que des gradients d'une espèce chimique régule la formation de gouttes de protéines rapproche le système de partition ParABS du système d'orientation des cellules de l'espèce *C. elegans* [18, 19]. Dans ce

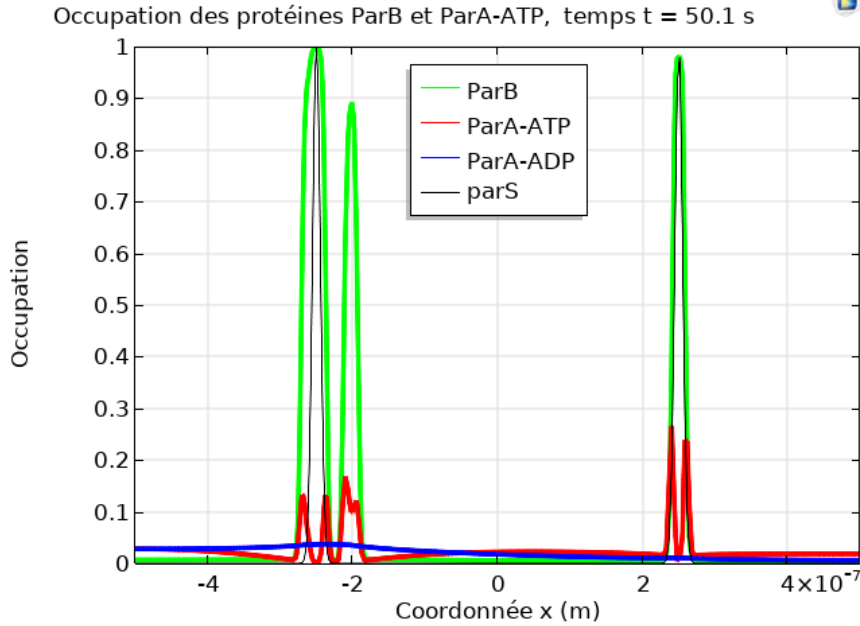


FIGURE 3.25 – **Profil des occupations pour $t = 50,1$ s** des protéines ParB (en vert) et protéines ParA sous forme ATP (en rouge) et protéines ParA-ADP (en bleu) selon l’axe principale de la bactérie. Les séquences *parS* activées sont en trait noir et fin : deux séquences *parS* sont actives. L’hydrolyse de l’ATP par les protéines ParA autour des complexes ParBS est désormais activée. Une dynamique complexe au complexe ParBS a lieu. La forme ADP de la protéine ParA apparaît dans l’espace cytoplasmique.

système, les granules P sont démantelés par les protéines Mex-5 réparties dans l’espace intracellulaire suivant un gradient dans le sens de l’axe principale des cellules. Les granules P ne continuent ainsi d’exister qu’à un seul pôle de la cellule.

3.4.4 Photoblanchiment *in silico* avec protéine ParA

Nous sommes désormais en position de produire un photoblanchiment *in silico*. Un profil stationnaire contenant deux complexes ParBS stables est généré en utilisant le système formé des équations (3.95), (3.96) et (3.97). Les protéines ParB existent sous deux formes : la forme photoblanchie $B_P(\mathbf{r}, t)$ et la forme fluorescente $B_F(\mathbf{r}, t)$ dont les profils initiaux sont définis dans la précédente méthode de photoblanchiment. Le système d’équations pour l’étape

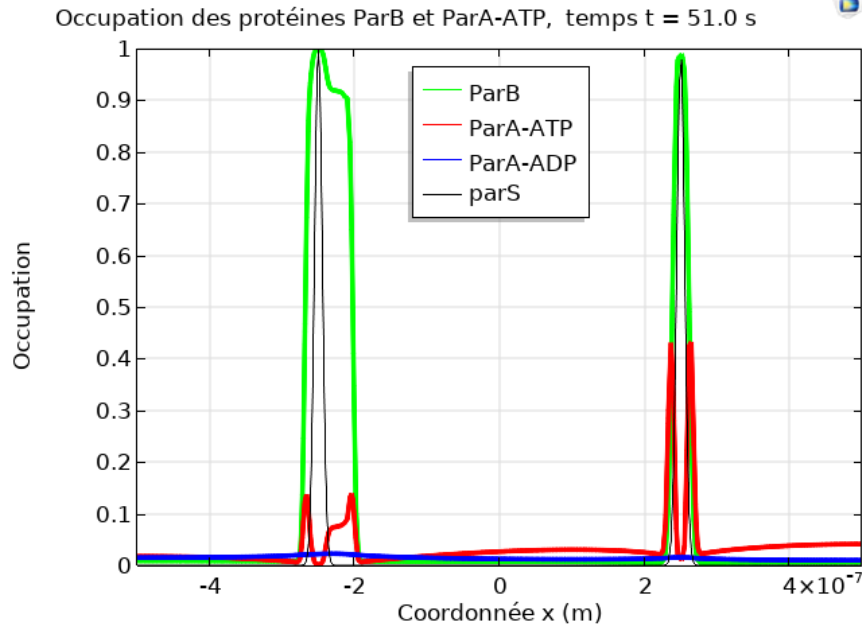


FIGURE 3.26 – **Profil des occupations pour $t = 51$ s** des protéines ParB (en vert) et protéines ParA sous forme ATP (en rouge) et protéines ParA-ADP (en bleu) selon l'axe principale de la bactérie. Les séquences *parS* activées sont en trait noir et fin : deux séquences *parS* sont actives. L'hydrolyse de l'ATP est aussi activée. La forme ADP des protéines ParA est recyclée en forme ATP sur l'ADN. Les nouvelles protéines ParA-ATP sont recrutées par le second complexe ParB*S*.

de recouvrement de la fluorescence est :

$$\frac{\partial B_F}{\partial t} = \nabla \left[\tilde{L}_B (B_F + B_P) \nabla \mu_F \right] \quad (3.102)$$

$$\frac{\partial B_P}{\partial t} = \nabla \left[\tilde{L}_B (B_F + B_P) \nabla \mu_P \right] \quad (3.103)$$

$$\frac{\partial A_T}{\partial t} = \nabla \left[\tilde{L}_A (A_T) \nabla \mu_A \right] + k_1 A_D - k_2 A_T (B_F + B_P) \quad (3.104)$$

$$\frac{\partial A_D}{\partial t} = D_V^A \nabla^2 A_D - k_1 A_D + k_2 A_T (B_F + B_P), \quad (3.105)$$

avec les potentiels chimiques liés aux protéines ParB fluorescentes :

$$\begin{aligned} \mu_F = & -\tilde{J}_{BB} l_m^2 \nabla^2 (B_F + B_P) - \tilde{J}_{BB} q (B_F + B_P) - \tilde{J}_{AB} l_m^2 \nabla^2 A_T - \tilde{J}_{AB} q A_T \\ & + \tilde{\epsilon}(x) + \ln \left(\frac{B_F}{1 - A_T - B_F - B_P} \right), \end{aligned} \quad (3.106)$$

aux protéines ParB photoblanchies :

$$\begin{aligned} \mu_P = & -\tilde{J}_{BB} l_m^2 \nabla^2 (B_F + B_P) - \tilde{J}_{BB} q (B_F + B_P) - \tilde{J}_{AB} l_m^2 \nabla^2 A_T - \tilde{J}_{AB} q A_T \\ & + \tilde{\epsilon}(x) + \ln \left(\frac{B_P}{1 - A_T - B_F - B_P} \right), \end{aligned} \quad (3.107)$$

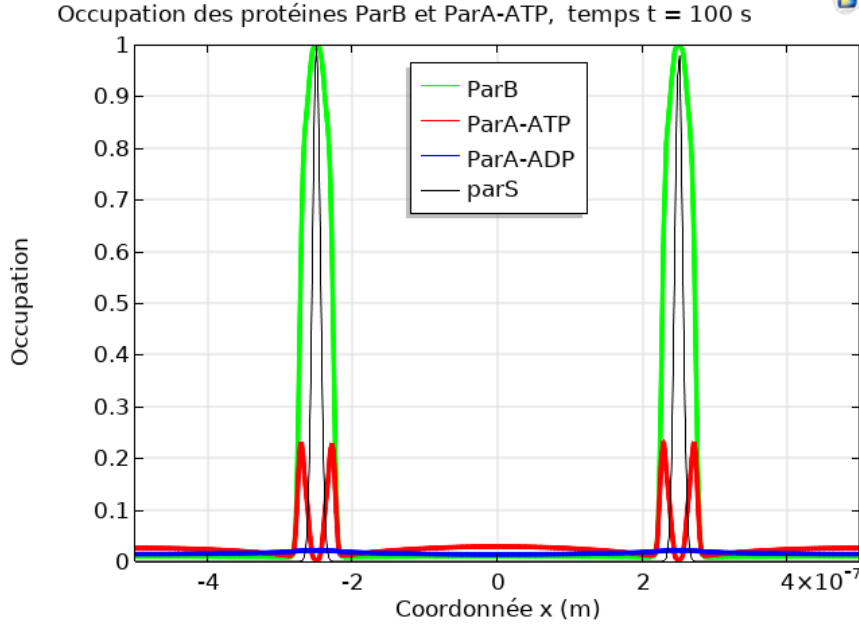


FIGURE 3.27 – **Profil des occupations pour $t = 100s$** des protéines ParB (en vert) et protéines ParA sous forme ATP (en rouge) et protéines ParA-ADP (en bleu) selon l'axe principale de la bactérie. Les séquences *parS* activées sont en trait noir et fin : deux séquences *parS* sont actives. L'hydrolyse de l'ATP est aussi activée. Au temps long, on observe un équilibrage des complexes ParBS grâce à l'hydrolyse de l'ATP localisée autour des complexes par les protéines ParA.

aux protéines ParA sous forme ATP :

$$\begin{aligned} \mu_A = & -\tilde{J}_{AA}l_m^2 \nabla^2 A_T - \tilde{J}_{AA}qA_T - \tilde{J}_{AB}l_m^2 \nabla^2 (B_F + B_P) - \tilde{J}_{AB}q(B_F + B_P) \\ & + \ln \left(\frac{A_T}{1 - A_T - B_F - B_P} \right). \end{aligned} \quad (3.108)$$

Nous utilisons les mêmes paramètres du tableau 3.2 pour obtenir les courbes de récupération de la figure 3.30, si ce n'est la valeur de D_L^B , le coefficient de diffusion des protéines ParB en phase condensée $D_L^B = 9 \times 10^{-5} \mu m^2 \cdot s^{-1}$, valeur 20 fois plus faible que la valeur utilisée précédemment. Nous constatons à nouveau la nécessité d'un coefficient de diffusion très faible dans la phase condensée pour expliquer les temps de recouvrement de fluorescence de l'ordre de plusieurs minutes. Il faut cependant noter que dans le processus d'échange des protéines ParB, les deux complexes se déséquilibrent légèrement. Bien que l'état final utilisé pour créer les profils des protéines ParB photoblanchies et fluorescentes soit un équilibre du système formé par les équations du modèle à 3 champs (équations (3.95) à (3.97)), les profils initiaux des protéines ParB fluorescentes $B_F(t = 0)$ et photoblanchies $B_P(t = 0)$ ne sont pas des états d'équilibres du système formé par les équations du modèle à 4 champs (équations (3.102) à (3.105)). Ce léger déséquilibre se traduit dans la figure 3.30

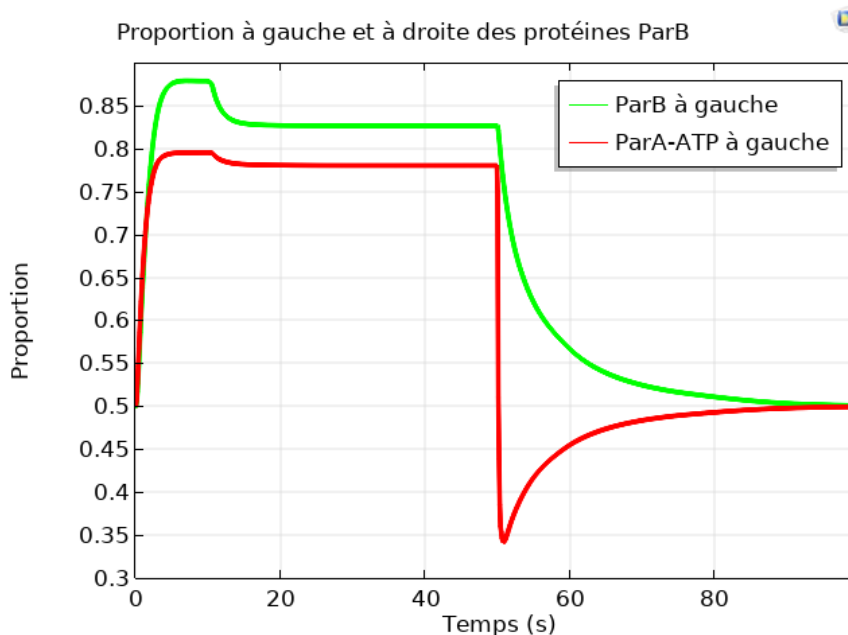


FIGURE 3.28 – **Evolution des proportions des protéines ParB et ParA-ATP** dans la partie gauche de l’espace d’intégration. La proportion des protéines ParB à gauche est en vert et celle des protéines ParA-ATP en rouge. On observe un équilibrage synchronisé des protéines ParA et ParB dans les parties gauche et droite de l’espace intracellulaire.

par une augmentation du nombre de protéines ParB dans le complexe 1 dans les premières secondes de la simulation. La récupération observée dans cet ultime photoblanchiment *in silico* est à la fois un transfert lent des protéines d’un complexe vers l’autre, semblable au mécanisme développé dans notre approche phénoménologique, mais aussi un processus d’équilibrage des gouttes. Le photoblanchiment n’est pourtant pas considéré comme une technique invasive. Nous touchons probablement aux limites de cette dernière approche appliquée aux expériences de photoblanchiment.

Nous souhaitons désormais expliciter les liens entre la cinétique des protéines du système de partition et la ségrégation des complexes ParBS. Notre approche fondamentale permet aussi d’étudier la partie dynamique du système ParABS. Dans la prochaine et dernière section, nous montrerons comment notre modèle à trois champs explique la ségrégation du matériel génétique chez les bactéries.

3.5 Lien entre cinétique et dynamique

Dans toutes les approches présentées, le système physique constitué du réseau d’ADN et des protéines ParA et ParB n’était pas totalement libre. Nous avons en effet supprimé un degré de liberté fondamental du système de parti-

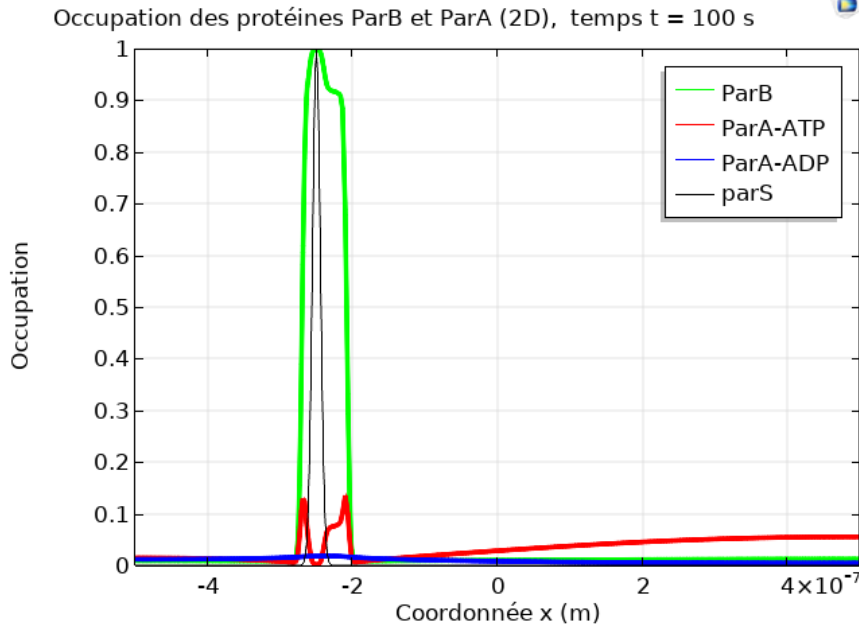


FIGURE 3.29 – **Profil des occupations avec une seule séquence $parS$** des protéines ParB (en vert) et protéines ParA sous forme ATP (en rouge) et protéines ParA-ADP (en bleu). Les séquences $parS$ activées sont en trait noir et fin. Le champ de protéines ParA-ATP forment un gradient vers la droite d’espace intracellulaire capable de générer un complexe ParBS équilibré si une nouvelle séquence $parS$ apparaît.

tion, les positions des séquences $parS$, considérées comme immobiles dans les sections précédentes. Or, un ensemble d’équation dont l’ambition est de décrire le fonctionnement du système de partition ParABS doit comporter une équation dynamique pour chaque séquence $parS$ dans l’espace intracellulaire. Cette équation dynamique contient, outre un terme d’amortissement, la force exercée par le champ des protéines ParB sur la séquence $parS$ c’est-à-dire la force réciproque de l’attraction spécifique des protéines ParB sur la séquence $parS$ induite par le terme en $\tilde{\epsilon}(\mathbf{r})$. Nous exposons d’abord une première formulation du problème du positionnement des complexes ParBS que nous adaptons ensuite à notre propre formulation en terme de protéines sur réseau.

3.5.1 Force de protéophorèse

Les expériences de 3D-SIM montrent que la ségrégation et le positionnement des complexes ParBS a lieu selon l’axe principal, longitudinal, de la bactérie [54]. Dans cette première approche [41], le système est ainsi réduit à l’axe principale de la bactérie de longueur L et nous utilisons un formalisme unidimensionnel. Les concentrations des espèces ParA-ATP $v(x, t)$ et ParA-ADP

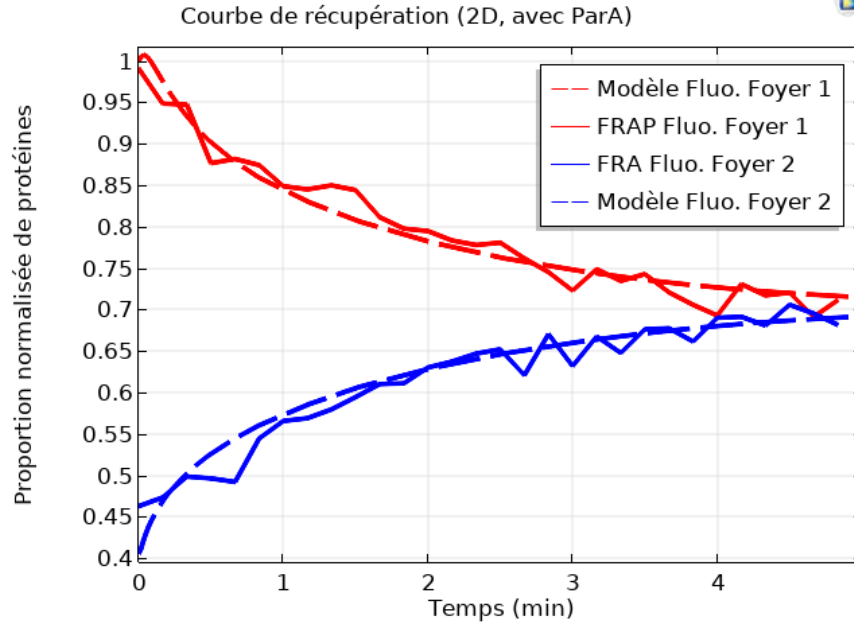


FIGURE 3.30 – **Courbe de recouvrement** *in silico* pour l’approche fondamentale avec protéines ParA, à une dimension. En trait plein : courbes de recouvrement en bleu et perte en rouge issues des expériences. En trait discontinu : courbes de recouvrement en bleu et perte en rouge issues des simulations. Ces courbes ont été produites pour un coefficient de diffusion de la phase condensé de $D_L^D = 9 \times 10^{-5} \mu\text{m}^2 \cdot \text{s}^{-1}$.

$u(x, t)$ sont régies par de simples équations de réaction-diffusion [35] :

$$\text{ParA} - \text{ADP} : \quad \frac{\partial u}{\partial t} = D_A^V \frac{\partial^2 u}{\partial x^2} - k_1 u + k_2 v \sum_i S(x - x_i) \quad (3.109)$$

$$\text{ParA} - \text{ATP} : \quad \frac{\partial v}{\partial t} = D_A^L \frac{\partial^2 v}{\partial x^2} + k_1 u - k_2 v \sum_i S(x - x_i) \quad (3.110)$$

où k_2 et k_1 sont les taux de la réaction d’hydrolyse près des complexes ParBS et de recyclage de l’ATP dans le cytoplasme définis dans la précédente section. Ces équations sont liées à des conditions aux bords qui annulent tout flux de particules en dehors de l’espace de taille L considéré :

$$\mathbf{n} \cdot \partial_x u|_{\text{bords}} = \mathbf{n} \cdot \partial_x v|_{\text{bords}} = 0 \quad (3.111)$$

où \mathbf{n} est un vecteur normal aux bords du système. Dans cette définition, u correspond à notre champ A_D à un facteur de normalisation près. Notons que la forme de l’équation pour l’espèce ParA-ATP correspond à l’équation (3.96) dans une limite basse occupation, sans protéines ParB et sans interactions entre protéines. Nos équations de cinétique thermodynamique peuvent ainsi être vues comme une généralisation des équations de réaction-diffusion. Un complexe

ParBS à une position x_i est considéré comme un objet en soi, décrit par la fonction $S(x - x_i)$: il s'agit, dans ce modèle, d'une probabilité d'interaction et d'hydrolyse de l'ATP par les protéines ParB. Ces équations sont couplées avec les équations dynamiques suivantes pour les position x_i des complexes ParBS :

$$m\gamma \frac{dx_i}{dt} = \epsilon_h \int_L \frac{\partial v(x, t)}{\partial x} S(x - x_i) dx \quad (3.112)$$

Nous négligeons la force brownienne, qui ne semble pas jouer un rôle fondamental dans le mécanisme de positionnement, et la force de confinement dans l'espace intracellulaire qui agissent sur le complexe, l'espace intracellulaire étant assimilé directement au volume du nucléoïde. Le terme de gauche correspond à la force de frottement qui agit sur le complexe à cause de la viscosité du milieu. Cette force dépend d'un coefficient massique $m\gamma$ que nous pouvons réécrire en fonction du coefficient de diffusion effectif du complexe de partition D_{BS} avec la relation d'Einstein $m\gamma = k_B T / D_{BS}$. Le terme de droite correspond à la force exercée par les protéines ParA sur le complexe ParBS et dépend de l'énergie $\epsilon_h \approx 10k_B T$ libérée par l'hydrolyse de l'ATP par les protéines ParB. Il s'agit d'une force phénoménologique, dite de protéophorèse, c'est-à-dire qu'elle dépend des interactions entre les protéines ParA et ParB dans le volume du complexe ParBS, en opposition à une force de chimiophorèse pour laquelle les interactions n'auraient lieu qu'en surface. Le complexes ParBS est considéré comme suffisamment poreux pour que les protéines ParA puissent pénétrer et diffuser à l'intérieur. La force de protéophorèse mène le complexe ParBS à remonter le gradient de concentration de protéines ParA-ATP le plus important. En supposant que le complexe ParBS est fixe ($\dot{x}_i = 0$), nous pouvons déterminer l'expression des profils stationnaires de concentration des protéines sur l'axe principal de la bactérie pour une probabilité $S(x - x_i)$ de la forme d'une fonction de Dirac. En réinjectant ces expressions dans l'équation dynamique (3.112), il est possible de montrer que :

1. pour un seul complexe ParBS, la position d'équilibre de la force de protéophorèse correspond à l'abscisse $x = 0$;
2. pour deux complexes ParBS, les positions d'équilibre sont $x = L/2$ et $x = -L/2$ qui correspondent aux positions 1/4 et 3/4 de l'axe principale de la bactérie ;
3. pour trois complexes, et nous pouvons facilement généraliser au-delà, les positions d'équilibre sont $x = -L/3$, $x = 0$ et $x = L/3$.

Ces positions sont telles que les gradients de protéines ParA-ATP sont symétriques autour des sources $S(x - x_i)$, ce qui correspond à un équilibre des forces à gauche et à droite des complexes ParBS. Nous pouvons aussi montrer que, dans un espace infini ou très grand devant la longueur de corrélation $\sqrt{D_A^V/k_1}$, le système supporte des solutions en onde progressive, du type

$v(x, t) = v(x - c_{\text{OP}}t)$ dont l'existence dépend de la valeur critique α_c du paramètre sans dimension :

$$\alpha = \frac{\epsilon D_{\text{BS}}}{k_B T D_A^L}. \quad (3.113)$$

La valeur α_c peut s'écrire simplement :

$$\alpha_c \sim (\sigma_{\text{BS}} c_A)^{-1} \quad (3.114)$$

pour une fonction source S de la forme d'un Dirac pondérée d'une largeur $\sigma_{\text{BS}} = 35$ nm, égale à la taille d'un complexe ParBS [63], et une concentration totale en protéines ParA $c_A = 1200 \mu\text{m}^{-1}$ [54]. De manière plus générale, la forme du couplage critique α_c dépend non-linéairement de la forme de la fonction S . Ainsi, pour un seul complexe de partition :

1. si $\alpha < \alpha_c$, la vitesse des ondes progressives c_{OP} de concentration est nulle. Les positions d'équilibre de la force de protéophorèse sont alors stables et les séquences $S(x - x_i(t))$ convergent vers ces positions.
2. si $\alpha > \alpha_c$, la vitesse c_{OP} est non nulle. Les positions d'équilibre des forces deviennent instables. Dans un système fini, cette instabilité se traduit par des oscillations autour de ces positions.

Par conséquent, le critère d'existence des ondes progressives est aussi un critère de stabilité des positions d'équilibre de la force de protéophorèse. Les figures 3.31 et 3.32 illustrent la phénoménologie de ce modèle pour un seul et deux complexes respectivement dans un système fini de taille $L = 1 \mu\text{m}$ et pour un coefficient de diffusion d'un complexe ParBS de l'ordre de $D_{\text{BS}} = 5 \times 10^{-5} \mu\text{m}^2/\text{s}$. La fonction source S a la forme d'une gaussienne de largeur $10^{-2} \mu\text{m}$. Nous montrons les types de trajectoires obtenues selon si le couplage α est supérieur ou inférieur au couplage critique α_c . Le couplage critique effectif pour deux termes de source S , correspondant au cas avec deux complexes ParBS, est supérieur du couplage critique α_c défini pour un seul complexe : nous devons ainsi prendre des valeurs bien supérieures à α_c pour retrouver l'instabilité dynamique. Pour les paramètres biologiques, la valeur biologique α_{bio} du paramètre α est légèrement inférieure à la valeur α_c de l'équation (3.114) : les complexes ParBS convergent vers les positions 1/4 et 3/4 de la bactérie. Cette approche phénoménologique explique ainsi le positionnement des complexes ParBS.

3.5.2 Force dérivée de l'énergie libre

Dans l'approche basée sur le modèle de gaz sur réseau à plusieurs espèces, nous pouvons aussi définir une équation dynamique dérivant de l'énergie libre

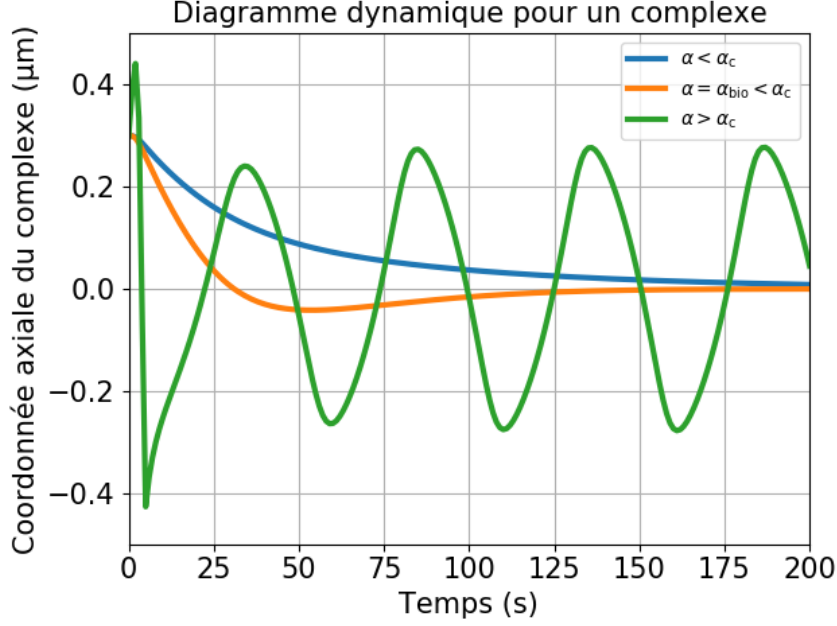


FIGURE 3.31 – **Diagramme dynamique pour un seul complexe** de partition et pour trois valeurs de α plaçant le système dans un régime stable de positionnement ($\alpha < \alpha_c$, en bleu et en orange) et un régime instable d'oscillations ($\alpha > \alpha_c$ en vert). La source S est initialement positionnée en $x = 0,3 \mu\text{m}$. La courbe en orange correspond à la valeur biologique $\alpha = \alpha_{\text{bio}}$ du couplage α : le positionnement est dans ce cas plus rapide que dans le cas sur-amorti en bleu.

\mathcal{F}_{ABS} du système ParABS :

$$\begin{aligned} \mathcal{F}_{\text{ABS}} = & \int_V \frac{J_{AA}}{2} \{l_m^2 [\nabla A_T]^2 - q A_T^2\} + \frac{J_{BB}}{2} \{l_m^2 [\nabla B]^2 - q B^2\} \\ & + J_{AB} \{m^2 [\nabla A_T \nabla B] - q A_T B\} + B\epsilon + k_B T A_T \ln A_T \\ & + k_B T [B \ln B + (1 - A_T - B) \ln(1 - A_T - B)] \frac{d^D \mathbf{r}'}{l_m^D} \end{aligned} \quad (3.115)$$

Néanmoins, il ne s'agit pas d'équations dynamiques pour les complexes ParBS. Dans notre approche, un complexe ParBS est le résultat de la cinétique des protéines ParB sur l'ADN et n'est pas un objet que nous pouvons simplement repérer et contraindre dans l'espace comme dans l'approche précédente. Ces nouvelles équations dynamiques concernent les séquences *parS*. Les positions \mathbf{r}_i des séquences *parS* deviennent dépendantes en temps et sont chacune gouvernées par une équation très similaire en forme à (3.112) :

$$\frac{d\mathbf{r}_i}{dt} = \frac{D_S}{k_B T} \mathbf{F}_{B/S}^i \quad (3.116)$$

Par souci de simplicité, nous négligeons la force brownienne agissant sur le complexe. Le terme de gauche et le rapport $D_S/(k_B T)$ proviennent de la force

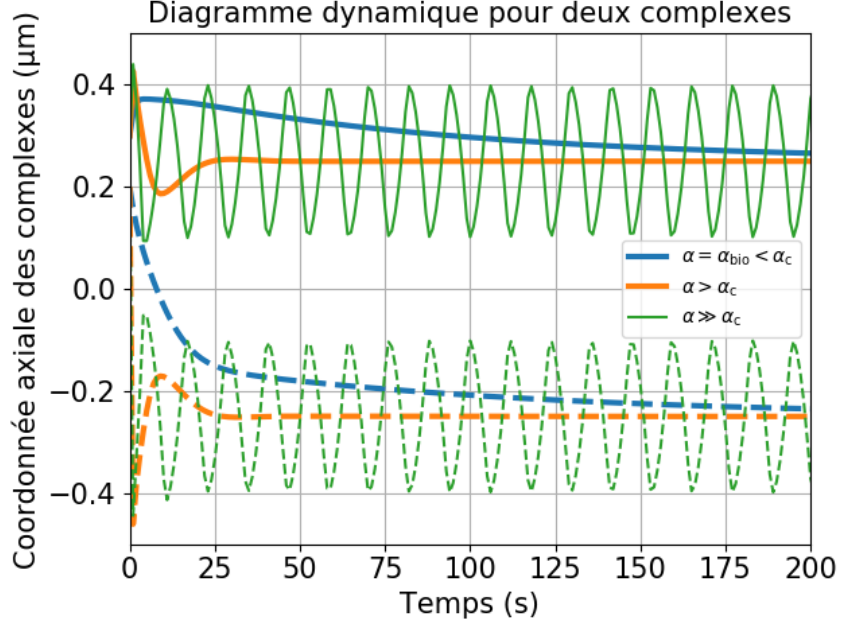


FIGURE 3.32 – **Diagramme dynamique pour deux complexes** de partition et pour trois valeurs de α plaçant le système dans un régime stable de positionnement (en bleu et en orange) et un régime instable d'oscillations (en vert). Les courbes pour la première source S , initialement positionnée en $x = 0,3 \mu\text{m}$, sont en trait plein et en trait discontinu pour la deuxième initialement située en $x = 0,2 \mu\text{m}$. La courbe en bleu correspond à la valeur biologique $\alpha = \alpha_{\text{bio}}$ du couplage α : pour deux complexes, cette valeur correspond à une convergence sur-amortie vers les positions $1/4$ et $3/4$.

visqueuse agissant sur la séquence $parS$ dans le milieu intracellulaire et de la relation d'Einstein. La force $\mathbf{F}_{B/S}$ correspond à la force exercée par les protéines ParB sur la séquence $parS$. Il s'agit de l'attraction réciproque à celle des séquences $parS$ sur les protéines ParB modélisée par l'énergie spécifique d'adsorption $\epsilon(\mathbf{r} - \mathbf{r}_i)$. Cette force dérive de l'énergie libre \mathcal{F}_{2E} , d'où :

$$\mathbf{F}_{B/S}^i = -\frac{\delta \mathcal{F}_{ABS}}{\delta \mathbf{r}_i} = \frac{1}{l_m^D} \int_V B(\mathbf{r}, t) \nabla \epsilon(\mathbf{r} - \mathbf{r}_i) d^D \mathbf{r} \quad (3.117)$$

Bien que similaire en forme à la force de protéophorèse introduite dans la sous-section précédente, elle ne possède pas la même fonction. Si la force de protéophorèse agit sur le complexe $ParBS$, notre force issue de l'énergie libre agit sur les séquences $parS$. De plus, la force de protéophorèse, alliée à l'hydrolyse de l'ATP, amène les complexes $ParBS$ à suivre le gradient le plus élevé en protéines ParA-ATP. Dans notre cas, la force $\mathbf{F}_{B/S}$ est liée à la formation de phase haute occupation autour des séquences $parS$. Ainsi, contrairement à la force de protéophorèse pour laquelle le complexe $ParBS$ creuse des gradients de protéines ParA-ATP, la force $\mathbf{F}_{B/S}$ mène à la formation de phases localement homogène. Si la taille effective de la séquence $parS$ ou fonction $\epsilon(\mathbf{r} - \mathbf{r}_i)$ est

inférieure à la taille de la phase haute occupation générée, alors il est simple de montrer que :

$$\mathbf{F}_{B/S}^i = \vec{0} \quad (3.118)$$

puisque ϵ , et par conséquent $B(\mathbf{r}, t)$ localement, sont pairs. Cette force n'a ainsi d'importance que lorsque le système est hors-équilibre thermodynamique, c'est-à-dire lorsque les phases haute occupation ou complexes ParBS sont perturbés.

Nous avons résolu les équations (3.95), (3.96) et (3.97) liées à deux équations dynamiques de la forme (3.117) pour deux séquences *parS* dans une géométrie d'intégration à deux dimensions identique à celle des dernières expériences de photoblanchiment *in silico* et des paramètres identiques à ceux indiqués dans le tableau 3.2 si ce n'est l'occupation moyenne initiale $\Phi_A = 4\Phi_B$ des protéines ParA, ce qui revient à augmenter la force exercée par les protéines ParA sur le complexe ParBS. En prenant la valeur $D_S = 1\mu\text{m}^2\text{s}^{-1}$ du coefficient de diffusion d'une séquence *parS*, nous pouvons tracer le diagramme dynamique de deux séquences *parS* initialement positionnées en $x = \pm 0,1\mu\text{m}$ (voir figure 3.33). Les deux complexes ParBS convergent vers les positions

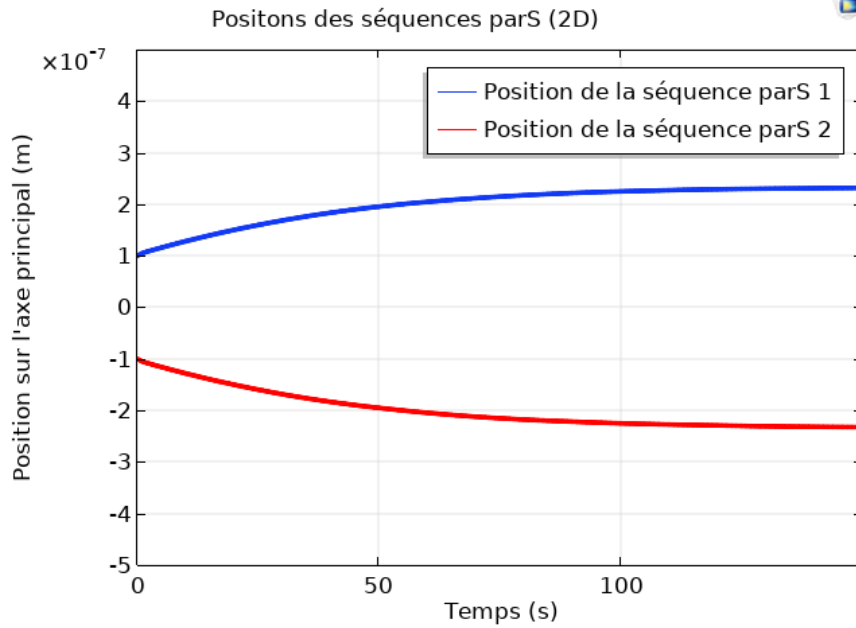


FIGURE 3.33 – **Diagramme dynamique** de deux séquences *parS*. Les deux séquences sont initialement positionnées en $x = \pm 0,1\mu\text{m}$. Elles convergent vers les positions $x = \pm 0,25\mu\text{m}$ ou $1/4$ et $3/4$ de l'espace d'intégration.

$1/4$ et $3/4$ de l'espace d'intégration entraînant les séquences *parS* avec eux. Le mécanisme de positionnement est contrôlé par les protéines ParA et l'hydrolyse de l'ATP. A cause des interactions entre protéines ParA et ParB, les

3.5. LIEN ENTRE CINÉTIQUE ET DYNAMIQUE

protéines ParA-ATP occupent de manière relativement importante l'ADN autour des complexes ParBS. Or, la distribution des protéines ParA autour des complexes ParBS est asymétrique lorsque les complexes n'occupent pas leurs positions d'équilibres grâce à l'hydrolyse de l'ATP. Dans la figure 3.30, avec une séquence *parS* immobile, nous pouvions déjà voir une asymétrie des protéines ParA-ATP autour des phases haute occupation en protéines ParB. Nous pouvons constater la même asymétrie en figure 3.34 pour les deux séquences *parS*. Les protéines ParA déforment les complexes ParBS de manière asymétrique,

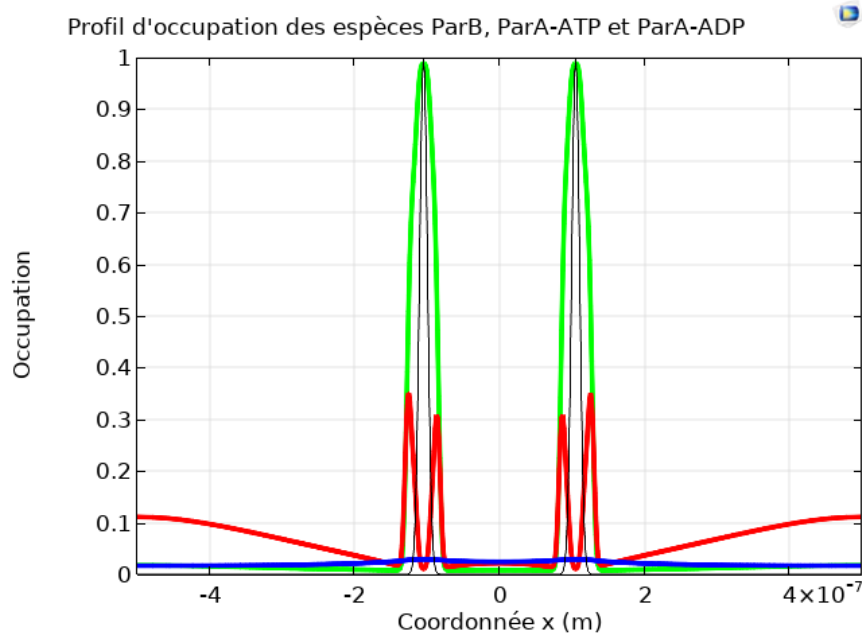


FIGURE 3.34 – **Profil des occupations avec deux séquences *parS* hors équilibre dynamique** des protéines ParB (en vert) et protéines ParA sous forme ATP (en rouge) et protéines ParA-ADP (en bleu) pour $t = 2s$, dans une situation dynamiquement instable. Les protéines ParA-ATP encerclent les complexes ParBS de manière asymétrique. Cette situation correspond à un déséquilibre des forces à gauche et à droite de chaque complexe et favorise un mouvement centripète des complexes.

dans le sens du plus grand gradient en protéines ParA-ATP. La déformation du complexe ParBS contraint la séquence *parS* à se repositionner grâce à la force $\mathbf{F}_{B/S}^i$ dérivée de l'énergie libre. Ce mécanisme mène au positionnement des complexes ParBS et donc des séquences *parS* aux positions 1/4 et 3/4 de l'espace d'intégration puisque ces positions correspondent à des distributions en protéines ParA-ATP symétriques autour des complexes ParBS. A trois dimensions, l'interface entre protéines ParA et ParB étant plus importante, ce mécanisme devrait être plus performant et devrait nécessiter une occupation en protéines ParA plus faible que celle utilisée dans notre simulation à deux dimensions. Ainsi, notre approche fondamentale nous permet d'expliquer l'ori-

gine de la force de protéophorèse effective du précédent modèle et de formuler un scénario pour la ségrégation des complexes ParBS.

3.6 Conclusion

A partir des résultats des expériences de photoblanchiment des complexes ParBS, nous avons présenté un modèle de comptage dans l'espace intracellulaire pour déterminer les temps typiques d'existence τ_{in} et τ_{out} des protéines ParB à l'intérieur et à l'extérieur des complexes. En comparant avec le temps typique de diffusion des protéines ParB entre deux complexes, nous avons montré qu'un mécanisme de pure diffusion n'expliquait pas les temps de récupération de la fluorescence de plusieurs minutes observés expérimentalement. Deux mécanismes pouvaient expliquer ces temps typiques : un coefficient de diffusion faible dans la phase à haute densité en protéines ParB ou des barrières d'énergie à la frontière des complexes. Nous avons illustré ces deux scénarios grâce à un modèle phénoménologique basé sur la résolution d'une équation de Smoluchowski permettant de mimer *in silico* l'expérience de photoblanchiment. Une modélisation plus audacieuse nous a mené à considérer un modèle de gaz sur réseau à plusieurs espèces. Cette approche de cinétique thermodynamique hors équilibre permet de montrer que l'existence de deux complexes ParBS stables était impossible sans l'activité d'hydrolyse de l'ATP des protéines ParA, sous l'hypothèse que les complexes ParBS résulte d'une séparation de phase. Nos simulations montrent un mécanisme d'équilibrage des complexes ParBS lorsque l'hydrolyse de l'ATP par les protéines ParA est activée. L'intégration de ce mécanisme de maintien de deux complexes ParBS nous a permis de retrouver les courbes expérimentales des transferts de fluorescence pour un coefficient de diffusion de la phase haute occupation faible. Nous avons finalement explicité les liens entre la cinétique des protéines ParA et ParB et la dynamique du système de partition, c'est-à-dire la ségrégation des séquences *parS*, et proposé un scénario pour cette ségrégation. Les protéines ParA et l'hydrolyse de l'ATP sont les ingrédients essentiels de l'existence de deux complexes ParBS stables et leur ségrégation dans l'espace. Les mécanismes d'inversion de la maturation d'Ostwald et du positionnement des complexes sont consubstantiels. Ils sont les résultats des interactions entre protéines ParA et ParB et de l'hydrolyse de l'ATP par les protéines ParA et accéléré par les protéines ParB.

Notre première approche, très simple, permet de juger si des temps de recouvrement de fluorescence sont expliqués par de la pure diffusion. Les expériences de photoblanchiment pourraient ainsi être un test pertinent pour caractériser les séparations de phases. La deuxième approche est une approche intermédiaire : facile à résoudre numériquement, elle possède un nombre minimal de paramètres. Notre protocole de photoblanchiment *in silico*, correctement calibré à l'aide de données expérimentales, doit permettre de vérifier un

3.6. CONCLUSION

nombre conséquent d'hypothèse théorique. Notre dernière approche offre l'opportunité d'étudier la cinétique des protéines sur l'ADN, qu'elles subissent des séparations de phase ou non. Nous avons montré que ces équations cinétiques sont une généralisation des équations de réaction-diffusion : des effets de séparation de phase et de réaction-diffusion peuvent être intégrés dans une seule et même équation. Cette approche théorique, très détaillée, permet d'explorer complètement l'espace des paramètres d'un système biologique et de prédire les rôles des éléments impliqués. Ce type d'approche analytique résolue numériquement permet de prédire de le comportement des systèmes mutants en modifiant les valeurs de paramètres bien choisis. Notre étude à deux dimensions pourrait s'avérer pertinente pour les systèmes de protéines membranaires tels que le système MinCDE. Enfin, ces travaux, loin de se restreindre à la biologie, pourraient s'avérer utiles dans l'étude des systèmes colloïdaux synthétiques.

3.7 Annexe 1 : Logiciel COMSOL Multiphysics

Nous effectuons nos simulations à l'aide du logiciel COMSOL Multiphysics développé par l'entreprise suédoise COMSOL¹. Ce logiciel, en principe orienté pour répondre aux besoins des ingénieurs, permet entre autre de résoudre des équations différentielles². Le logiciel permet de générer simplement des maillages d'intégration numérique et ainsi de résoudre des équations aux dérivées partielles couplées dans des géométries complexes à l'aide de méthodes aux éléments finis adaptées. Le logiciel offre aussi une large gamme d'outils pour exploiter les résultats des simulations. Il s'agit d'un logiciel exigeant qui requiert un investissement conséquent pour manipuler des équations complexes. Dans notre cas, nous définissons deux équations couplées dotées de leur condition aux bords (flux normal nul) : une équation pour l'occupation et une équation pour le potentiel chimique. Nous sommes contraints à utiliser cette formulation à cause de la gestion des dérivées d'ordre supérieur (3 ou 4) par le logiciel. Ce sont les valeurs interdites introduites par les termes entropiques en logarithme qui peuvent poser problème dans notre cas, dans la mesure où les séparations de phase engendrées poussent précisément les valeurs locales d'occupation proches des bornes 0 et 1.

Nous listons ici une procédure simplifiée de construction d'application COMSOL.

Dimensionnalité : Une première étape consiste à définir la dimension du système. A noter qu'une certaine dimension donne accès aux outils des dimensions inférieures. Il est ainsi possible de faire des intégrales de surface ou de courbe dans un environnement 2D ou 3D.

Paramètres : Les paramètres, définis dans "Définitions globales" sont les variables globales de l'application. Ils sont donc valables dans toutes ses composantes. Dans COMSOL, les variables de "Définition" sont réévaluées à chaque pas de temps, contrairement aux paramètres.

Fonctions : Une fois les paramètres définis, il est aisé de déclarer les fonctions pertinentes de la simulation. Le logiciel permet de déclarer certaines fonctions types (gaussienne normalisée, bruit uniforme ou gaussien etc) ou d'écrire directement une expression analytique. Il est aussi possible d'indiquer manuellement les dérivées partielles de la forme analytique déclarée.

Géométrie : La puissance de COMSOL réside peut-être dans cet aspect. Le logiciel permet à l'utilisateur d'intégrer ses équations dans des géométries audacieuse proche de cas réels. De nombreux objets géométriques

1. Voir <https://www.comsol.fr/>.

2. Il s'agit du module Mathématiques. D'autres modules permettent de simuler du transport de particules, du transfert du chaleur, des écoulements etc.

basiques sont disponibles (pavé, ellipse, point, etc), combinables grâce à des opérations d'union ou d'intersection. La forme finale peut être complexe : on trouve dans la documentation du logiciel un exemple de simulation de la propagation de la chaleur dans une pièce mécanique tridimensionnelle dotée de multiples courbures. Nous avons utilisé une géométrie qui reproduit le rapport d'aspect d'une bactérie.

Matériaux : Une fois la géométrie créée, l'utilisateur choisit la composition chimique de l'objet dans une large bibliothèque de matériaux gérée par le logiciel. Ce choix confère automatiquement les propriétés pertinentes (viscosité, conductivité, etc) à la géométrie ou à certaines de ses parties. Nous n'avons pas utilisé cette fonctionnalité dans nos simulations.

Physique : Cette partie est le coeur de l'application. Pour le module Mathématiques, elle consiste à déclarer les équations régissant le système physique étudié. Il faut déclarer l'unité de la variable à intégrer. Une fonctionnalité du logiciel consiste à pouvoir coupler plusieurs équations de physiques différentes (couplage d'un écoulement et d'une diffusion de particules pour prendre en compte des effets de diffusiphorèse par exemple). L'adimensionnalisation permet d'augmenter la stabilité de la résolution.

Maillage : Pour appliquer ses méthodes de résolution aux éléments finis, COMSOL découpe l'espace en sous-espaces finis dont la taille est contrôlée par l'utilisateur. La forme des éléments est également contrôlée et il est même possible de modifier localement la densité du maillage autour d'un point d'intérêt, cette déformation pouvant elle-même dépendre du temps d'intégration. Lors de nos simulations à deux dimensions, nous avons utilisé un maillage triangulaire généré par une méthode de Delaunay et affiné autour des complexes de partitions pour améliorer la simulation.

Étude : COMSOL Multiphysics permet d'effectuer plusieurs types d'études simultanément : fréquentielle, temporelle et stationnaire. Dans notre cas, il s'agit d'étude temporelle. Nous devons ainsi indiquer au logiciel la durée de l'intégration, ainsi qu'un pas d'enregistrement des données. Le pas d'intégration est parfois plus faible car COMSOL gère par défaut en direct le pas de temps réel de la simulation : ce pas de temps adaptatif permet d'augmenter les chances de converger vers la solution numérique sans erreur numérique fatale. Ces options d'étude sont très vastes. Elles permettent entre autre de contrôler la méthode aux éléments finis utilisée ou la tolérance de stabilité.

Résultats : La section Résultats permet de visualiser les données récoltées lors de la résolution numérique et d'effectuer des opérations, de transformer ces données. Il existe différents modes de représentation des données

selon la dimension de l'étude. Il est toujours possible d'extraire les résultats et les quantités dérivées sous forme de tableau, d'image ou d'animation.

Chapitre 4

Conclusion

Nous sommes en mesure de décrire le fonctionnement du système de partition ParABS pour la ségrégation du plasmide F chez *E. Coli*. Les protéines ParB se lient sur l'ADN et préférentiellement sur la séquence *parS*. Les protéines ParB sur l'ADN, à cause des interactions de *spreading* et de *bridging* entre elles, existent dans un régime thermodynamique de coexistence métastable. Une fois la séquence *parS* occupée, les protéines ParB se recrutent sur l'ADN. Nous avons en effet montré que les interactions de *bridging* jouent le rôle d'interactions à longue portée le long de la chaîne d'ADN. Ces interactions effectives sont telles que les protéines ParB subissent une séparation de phase sur l'ADN. Les interactions de *bridging*, à longue portée dans l'espace, conduisent à la formation d'une structure tridimensionnelle auto-assemblée sur l'ADN comme le montrent les expériences de SPT-PALM. La cinétique d'attachement et de détachement des protéines ParB aboutit à la distribution moyenne déterminée par séquençage ChIP autour de la séquence *parS*. Les protéines ParB capturent la séquence *parS* et forment un objet localisable dans l'espace intracellulaire pouvant interagir plus efficacement avec d'autres espèces chimiques. Les complexes ParBS doivent être vus comme les vaisseaux des séquences *parS*. Lorsqu'il n'y a qu'un seul complexe ParBS, il se trouve localisé au milieu de l'espace intracellulaire. Ce sont les protéines ParA et l'hydrolyse de l'ATP qui permettent ce positionnement. Les protéines ParA possèdent une interaction attractive avec les protéines ParB. Cette interaction est compensée par l'hydrolyse de l'ATP, très accélérée autour des complexes ParBS puisque la protéine ParB est un catalyseur de l'hydrolyse de l'ATP par la protéine ParA. Après une étape sous forme ADP diffusive dans le cytoplasme, la forme ATP de la protéine ParA, liée à l'ADN, occupe ainsi principalement les bords des complexes de partition ou des zones qui en sont éloignées. Les protéines ParA tirent de manière symétrique autour du complexe lorsqu'il occupe le centre de la cellule : il s'agit donc d'un état d'équilibre stable. La réplication du matériel génétique a lieu et deux séquences *parS* coexistent dans le cytoplasme. Elles diffusent et se séparent localement. Les protéines ParA interviennent à nou-

veau pour que chaque séquence *parS* possède son cortège de protéines ParB. En cas de déséquilibre des complexes ParBS, les protéines ParA réfugiées dans les zones éloignées des complexes ParBS sont recrutées par le complexe minoritaire puisque l'hydrolyse de l'ATP y est plus faible. Les protéines ParA-ATP recrutent à leur tour des protéines ParB. Alors que ce complexe minoritaire voit son taux d'hydrolyse augmenter, un mécanisme d'échange réciproque entre les deux complexes a lieu : ce que l'un perd en protéines ParA-ATP, l'autre le gagne en proportion de protéines ParB. Cet échange, qui correspond à une inversion de la maturation d'Ostwald, aboutit à l'équilibrage et à l'existence de deux complexes ParBS de taille similaire. La position centrale devient une position d'équilibre instable au profit des positions 1/4 et 3/4 de l'espace intracellulaire qui deviennent des positions d'équilibre stable. Les protéines ParA-ATP tirent sur les deux complexes ParBS dans le sens des gradients les plus importants : elles déforment en réalité les complexes, obligeant les séquences *parS* à se repositionner à l'intérieur des gouttes de protéines ParB, jusqu'à la stabilisation. Nous pouvons résumer ce scénario : les deux complexes ParBS, vaisseau des séquences *parS*, interagissent entre eux au travers du champ de protéines ParA menant à leur équilibre et à ségrégation aux positions 1/4 et 3/4.

Dans cette thèse, nous avons montré que le système de partition ParABS est un système hors de l'état d'équilibre thermodynamique. Nous avons caractérisé cet état d'équilibre : il s'agit d'un seul complexe ParBS, localisé de préférence dans un bord du système, conformément à la thermodynamique des transitions de phase. Nous avons aussi montré comment ce système est maintenu hors de l'équilibre pour former deux complexes ParBS de tailles identiques grâce à un mécanisme hybride mêlant séparation de phase et réaction-diffusion des entités chimiques impliquées. Au regard de la théorie des séparations de phase biologiques sur l'ADN, nous avons démontré un critère d'existence de ces transitions dans la limite diluée. Nous avons développé des outils analytiques pour tracer les diagrammes de phase de ces séparations de phase. En ce qui concerne la cinétique des espèces chimiques sur l'ADN ou dans le cytoplasme, nous avons proposé une modélisation fondamentale ambitieuse qui apparaît comme la généralisation des équations de réaction-diffusion, avec un terme de diffusion sur la chaîne d'ADN pouvant introduire des propriétés de séparation de phase. Ce modèle semble indiqué pour étudier les processus hybrides de séparation de phase et de réaction-diffusion réunis, dans lesquels une inversion du mûrissement d'Ostwald intervient. Ces outils puissants soulignent la non-pertinence de modèles linéaires et de relations *ad hoc*, certes utiles en première instance, pour comprendre complètement les systèmes biologiques. Ils possèdent de plus des applications intéressantes en matière molle et active.

Ces outils théoriques nous semblent pertinents pour comprendre comment la matière vivante s'organise durant le cycle d'une cellule. L'organisation intra-

cellulaire ne saurait être résumée par nos équations et par notre cas d'étude : elle est trop riche et trop complexe pour être cernée dans sa totalité. Néanmoins, nous pensons avoir discerné des processus particuliers permettant l'organisation spatio-temporelle d'entités auto-assemblées. A ce titre, le système de partition *ParABS* pourrait être perçu comme une opportunité théorique et un exemple paradigmatique.

Bibliographie

- [1] R. Horwitz, Cellular Biophysics, *Biophysical Journal* 110, 993-996 (2016).
- [2] C. P. Satori, M. M. Henderson, E. A. Krautkramer, V. Kostal, M. M. Distefano, E. A. Arriaga, *Bioanalysis of eukaryotic organelles*, *Chemical Reviews* 113, 2733-2811 (2013).
- [3] I. Surovtsev, C. Jacobs-Wagner, *Subcellular Organization : A Critical Feature of Bacterial Cell Replication*, *Cell* 172, 1271-1293 (2018).
- [4] L. Shapira, H. H. McAdams, R. Losick, *Why and How Bacteria Localize Proteins* 326, 1225-1228 (2009).
- [5] D. M. Mitrea, R. W. Kriwacki, *Phase separation in biology ; functional organization of higher order*, *Cell Communication and Signaling* 14 (2016).
- [6] J. D. Halley, D. A. Winkler, *Consistent Concepts of Self-organization and Self-assembly*, *Wiley InterScience* 14(2), 15-17 (2008).
- [7] E. Schrödinger, *What is Life ?*, Cambridge University Press (1967).
- [8] T. Yamaguchi, N. Suematsu, H. Mahara, *Self-Organization of Hierarchy : Dissipative-Structure Assisted Self-Assembly of Metal Nanoparticles in Polymer Matrices*, *Nonlinear Dynamics in Polymeric Systems*, chapitre 2 (2003).
- [9] Y. Wang, A. Lomakin, R. F. Latypov, J. P. Laubach, T. Hideshima, P. G. Richardson, N. C. Munshi, K. C. Anderson, *Phase transition in human IgG solutions*, *The Journal of Chemical Physics* 139, 121904 (2013).
- [10] V. J. Anderson, H. N. W. Lekkerkerker, *Insights into phase transition kinetics from colloid science*, *Nature* 416, 811-815 (2002).
- [11] P. Li, S. Banjade, H.-C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, P. S. Russo, Q.-X. Jiang, B. Tracy Nixon, M. K. Rosen, *Phase Transitions in the Assembly of Multi-Valent Signaling Proteins*, *Nature* 183, 336-340 (2012).
- [12] A.-F. Bitbol, *Inferring interactions partners from protein sequences using mutual information*, *Public Library Of Science Computational Biology*, 14(11), e1006401 (2018).

- [13] S. Sacquin-Mora, A. Carbone, R. Lavery, Identification of Protein Interaction Partners and Protein-Protein Interaction Sites, *Journal of Molecular Biology* 382, 1276-1289 (2008).
- [14] Y. F. Inclan, E. Nogales, Structural models for the self-assembly and microtubule interactions of γ -, δ - and ϵ -tubulin, *Journal of Cell Science* 114, 413-422 (2000).
- [15] A. A. Hyman, C. A. Weber, F. Jülicher, Liquid-Liquid Phase Separation in Biology, *Annual Review of Cell and Developmental Biology* 30, 39-58 (2014).
- [16] J. Amaya, V. H. Ryan, N. L. Fawzl, The SH3 domain of Fyn interacts with and induces liquid-liquid phase separation of the low-complexity domain of hnRNPA2, *The Journal of Biological Chemistry* 293(51), 19522-19531 (2018).
- [17] J. Guillen-Boixet, V. Buzon, X. Salvatella, R. Mendez, CPEB4 is regulated during cell cycle by ERK2 :Cdk1-mediated phosphorylation and its assembly into liquid-like droplets, *eLife* 5, e19298 (2016).
- [18] C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoegel, J. Gharakhani, F. Jülicher, A. A. Hyman, Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation, *Science* 324, 1729-1732 (2009).
- [19] C. F. Lee, C. P. Brangwynne, J. Gharakhani, A. A. Hyman, F. Jülicher, Spatial Organization of the Cell Cytoplasm by Position-Dependent Phase Separation, *Physical Review Letters* 111, 088101 (2013).
- [20] N. Al-Husini, D. T. Tomares, O. Bitar, W. S. Childers, J. M. Schrader, α -Proteobacterial RNA Degradosomes Assemble Liquid-Liquid Phase-Separated RNP Bodies, *Molecular Cell* 71, 1027-1039 (2018).
- [21] B. Monterroso, S. Zorrilla, M. Sobrinos-Sanguino, M. A. Robles-Ramos, M. Lopez-Alvarez, W. Margolin, C. D. Keating, G. Rivas, Bacterial FtsZ protein forms phase-separated condensates with its nucleoid-associated inhibitor SlmA, *EMBO reports*, e45946 (2018).
- [22] A. Sanchez, D.I. Cattoni, J.-C. Walter, J. Rech, A. Parmeggiani, M. Nollmann, and J.-Y. Bouet, Stochastic self-assembly of ParB proteins builds the bacterial DNA segregation apparatus, *Cell Systems* 1, 163–173 (2015).
- [23] M. A. Schumacher, Structural biology of plasmid partition : uncovering the molecular mechanisms of DNA segregation, *Biochemical Journal* 412, 1-18 (2008).
- [24] J. W. Raff, Phase Separation and the Centrosome : A Fait Accompli ?, *Trends in Cell Biology* 29, 612-622 (2019).
- [25] P. St George-Hyslop, J. Q. Lin, A. Miyashita, E. C. Phillips, S. Qamar, S. J. Randle, G. Z. Wang, The physiological and pathological biophysics

- of phase separation and gelation of RNA binding proteins in amyotrophic lateral sclerosis and fronto-temporal lobar degeneration, *Brain Research* 1693, 11-23 (2018).
- [26] A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausia, M. S. Nissen, R. Reeves, C. H. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, Z. Obradovic, Intrinsically disordered protein, *Journal of Molecular Graphics and Modelling* 19, 26-59 (2001).
- [27] C. R. Berland, G. M. Thurston, M. Kondo, M. L. Broide, J. Pande, O. Ogun, G. B. Benedek, Solid-liquid phase boundaries of lens protein solutions, *Proceedings of the National Academy of Sciences* 89, 1214-1218 (1992).
- [28] G. B. Irvine, O. M. El-Agnaf, G. M. Shankar, D. M. Walsh, Protein Aggregation in the Brain : The Molecular Basis for Alzheimer's and Parkinson's Diseases, *Molecular Medicine* 14, 451-464 (2008).
- [29] S. Wegmann, B. Eftekharzadeh, K. Tepper, K. M. Zoltowska, R. E. Bennett, S. Dujardin, P. R. Lasbowski, D. MacKenzie, T. Kamath, C. Commins, C. Vanderburg, A. D. Roe, Z. Fan, A. M. Molliex, A. Hernandez-Vega, D. Muller, A. A. Hyman, E. Mandelkow, J.-P. Taylor, B. T. Hyman, Tau protein liquid-liquid phase separation can initiate tau aggregation, *EMBO Journal* 37, e98049 (2018).
- [30] P. R. ten Wolde, D. Frenkel, Enhancement of Protein Crystal Nucleation by Critical Density Fluctuations, *Science* 277, 1975-1978 (1997).
- [31] H.-X. Zhou, V. Nguemaha, K. Mazarakos, S. Qin, Why Do Disordered and Structured Proteins Behave Differently in Phase Separation?, *Trends in Biochemical Sciences* 43(7), 499-516 (2018).
- [32] E. E. Dormidontova, A. Y. Grosberg, A. R. Khokhlov, Intramolecular phase separation of copolymer chain with mobile primary structure, *Makromolekular Chemie Theory and simulations* 1, 375-385 (1992).
- [33] R. Wedlich-Söldner, T. Betz, Self-organization : the fundament of cell biology, *Philosophical Transactions of the Royal Society B* 373, 20170103 (2018).
- [34] J. Halatek, F. Brauns, E. Frey, Self-organization principles of intracellular pattern formation, *Philosophical Transactions of the Royal Society B* 373, 20170107 (2018).
- [35] A. M. Turing, The Chemical Basis of Morphogenesis, *Philosophical Transactions of the Royal Society of London Series B Biological Sciences* 237, 37-72 (1952).
- [36] J. Denk, S. Kretschmer, J. Halatek, C. Hartl, P. Schwille, E. Frey, MinE conformational switching confers robustness on self-organized Min protein

- patterns, *Proceedings of the National Academy of Sciences* 115(18), 4553-4558 (2018).
- [37] F. Brauns, J. Halatek, E. Frey, Phase-space geometry of reaction-diffusion dynamics, arXiv :1812.08684v1 (2018).
- [38] M. Loose, E. Fischer-Friedrich, J. Ries, K. Kruse, P. Schwille, Spatial Regulators for Bacterial Cell Division Self-Organize into Surface Waves in Vitro, *Science* 320, 789-792 (2008).
- [39] V. V. Rybenkov, V. Herrera, Z. M. Petrushenko, H. Zhao, MukBEF, a Chromosomal Organizer, *Journal of Molecular Microbiology and Biotechnology* 24, 371-383 (2014).
- [40] S. Murray, V. Sourjik, Self-organization and positioning of bacterial protein clusters, *Nature Physics* 13, 1006-1013 (2017).
- [41] J.-C. Walter, J. Dornig, V. Lorman, J. Rech, J.-Y. Bouet, M. Nollmann, J. Palmeri, A. Parmeggiani, F. Geniet, Surfing on protein waves : proteophoresis as a mechanism for bacterial genome partitioning, *Physics Review Letters* 119, 028101 (2017).
- [42] S. Million-Weaver, M. Camps, Mechanisms of plasmid segregation : have multicopy plasmids been overlooked ?, *Plasmid* 0, 27-36 (2014).
- [43] A. San Millan, Evolution of Plasmid-Mediated Antibiotic Resistance in the Clinical Context, *Trends in Microbiology* 26, 978-985 (2018).
- [44] A. Hofmann, J. Mäkelä, D. J. SHerratt, D. Heermann, S. M. Murray, Self-organised segregation of bacterial chromosomal origins, *eLife* 8, e46564 (2019).
- [45] G. Ebersbach, K. Gerdes, Plasmid Segregation Mechanisms, *The Annual Review of Genetics* 39, 453-479 (2005).
- [46] L. Jindal, E. Emberly, DNA segregation under Par protein control, *Public Library of Science* 14(7), e0218520 (2019).
- [47] J. Livny, Y. Yamaichi, M. K. Waldor, Distribution of Centromere-Like *parS* Sites in Bacteria : Insights from Comparative Genomics, *Journal of Bacteriology* 189, 8693-8703 (2007).
- [48] A. Sanchez, La ségrégation du plasmide F d'Echerichia coli : Etude des spécificités d'interaction du centromère avec la protéine SopB et organisation du complexe de partition étendu, Thèse de l'Université de Toulouse, soutenue en 2014.
- [49] V. Dahirel, F. Paillusson, M. Jardat, M. Barbi, J.-M. Victor, Nonspecific DNA-Protein Interaction : Why Proteins Can Diffuse along DNA, *Physical Review Letters* 102, 228101 (2009).
- [50] J.-Y. Bouet, Y. Ah-Seng, N. Benmeradi, D. Lane, Polymerization of SopA partition ATPase : regulation by DNA binding and SopB, *Molecular Microbiology* 63, 468-481 (2007).

-
- [51] J.-Y. Bouet, B. E. Funnell, P1 ParA interacts with the P1 partition complex at *parS* and an ATP-ADP switch controls ParA activities, *The EMBO Journal* 18(5), 1415-1424 (1999).
- [52] Y. Ah-Seng, F. Lopez, F. Pasta, D. Lane, J.-Y. Bouet, Dual Role of DNA in Regulating ATP Hydrolysis by the SopA Partition Protein, *The Journal of Biological Chemistry* 284(44), 30067-30075 (2009).
- [53] A. G. Vecchiarelli, Y.-W. Han, X. Tan, M. Mizuuchi, R. Ghirlando, C. Biertümpfel, B. E. Funnell, K. Mizuuchi, ATP control of dynamic P1 ParA-DNA interactions : a key role for the nucleoid in plasmid partition, *Molecular Microbiology* 78, 78-91 (2010).
- [54] A. Le Gall, D. I. Cattoni, B. Guilhas, C. Mathieu-Demazière, L. Oudjedi, J.-B. Fiche, J. Reich, S. Abrahamsson, H. Murray, J.-Y. Bouet, M. Nollmann, Bacterial partition complexes segregate within the volume of the nucleoid, *Nature Communications* 7, 12107 (2016).
- [55] C. P. Broedersz, W. Wang, Y. Meir, J. J. Loparo, D. Z. Rudner, N. S. Wingreen, Condensation and localization of the partitioning protein ParB on the bacterial chromosome, *Proceedings of the National Academy of Sciences* 111(24), 8809-8814 (2014).
- [56] R. E. Debaugny, A. Sanchez, J. Reich, D. Labourdette, J. Dorignac, F. Geniet, J. Palmeri, A. Parmeggiani, F. Boudsocq, V. A. Leberre, J.-C. Walter, J.-Y. Bouet, A conserved mechanism drives partition complex assembly on bacterial chromosomes and plasmids, *Molecular Systems Biology* 14, e8516 (2018).
- [57] O. Rodionov, M. Lobočka, M. Yarmolinsky, Silencing of Genes Flanking the P1 Plasmid Centromere, *Science* 283, 546-549 (1999).
- [58] T. G. W. Graham, X. Wang, D. Song, C. M. Etson, A. M. van Oijen, D. Z. Rudner, J. J. Loparo, ParB spreading requires DNA bridging, *Genes & Development* 28 (2014).
- [59] J.-C. Walter, N.-O. Walliser, G. David, J. Dorignac, F. Geniet, J. Palmeri, A. Parmeggiani, N. S. Wingreen, C. P. Broedersz, Looping and clustering model for the organisation of protein-DNA complexes on the bacterial genome, *New Journal of Physics* 20, 035002 (2018).
- [60] G. David, J.-C. Walter, C. P. Broedersz, J. Dorignac, F. Geniet, A. Parmeggiani, N.-O. Walliser, J. Palmeri, Phase separation of polymer-bound particles induced by loop-mediated 1D effective long-range interactions, *arXiv :1811.09234* (2018).
- [61] C. Manzo, M. F. Garcia-Parajo, A review of progress in single particle tracking : from methods to biophysical insights, *Reports on Progress in Physics* 78(12) (2015).

- [62] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, H. F. Hess, Imaging Intracellular Fluorescent Proteins at Nanometer Resolution, *Science* 313, 1642-1645 (2006).
- [63] B. Guilhas, J.-C. Walter, J. Rech, G. David, N.-O. Walliser, J. Palmeri, C. Mathieu-Demaziere, A. Parmeggiani, J.Y. Bouet, A. Le Gall, M. Nollmann, ATP-driven separation of liquid phase condensates in bacteria, submitted.
- [64] G. L. M. Fisher, C. L. Pastrana, V. A. Higman, A. Koh, J. A. Taylor, A. Butterer, T. Craggs, F. Sobott, H. Murray, M. P. Crump, F. Moreno-Herrero, M. S. Dillingham, The structural basis for dynamic DNA binding and bridging interactions which condense the bacterial centromere, *eLife* 6, e28086 (2017).
- [65] D. S. Johnson, A. Mortazavi, R. M. Myers, B. Wold, Genome-Wide Mapping of in Vivo Protein-DNA Interactions, *Science* 316, 1497-1502 (2007).
- [66] D. Axelrod, D. E. Koppel, J. Schlessinger, E. Elson, W. W. Webb, Mobility measurement by analysis of fluorescence photobleaching recovery kinetics, *Biophysical Journal* 16, 1055-1069 (1976)
- [67] R. Heintzmann, T. Huser, Super-Resolution Structured Illumination Microscopy, *Chemical Reviews* 117, 13890-13908 (2017).
- [68] C. P. Brangwynne, P. Tompa, R. V. Pappu, Polymer physics of intracellular phase transitions, *Nat. Phys.* 11, 899-904 (2015).
- [69] John F. Marko, Biophysics of protein-DNA interactions and chromosome organization, *Physica A* 418, 126-153 (2015).
- [70] F. Paillusson, V. Dahirel, M. Jardat, J.-M. Victor, M. Barbi, Effective interaction between charged nanoparticles and DNA, *Physical Chemistry Chemical Physics* 13, 12603-12613 (2011).
- [71] J. Mozziconacci, C. Lavelle, M. Barbi, A. Lesne, J.-M. Victor, A physical model for the condensation and decondensation of eukaryotic chromosomes, *FEBS Letters* 580, 368-371 (2006).
- [72] N. Haddad, D. Jost, C. Vaillant, Perspectives : using polymer modeling to understand the formation and function of nuclear compartments, *Chromosome Res.* 25, 35-50 (2017).
- [73] D. Jost, C. Vaillant, P. Meister, Coupling 1D modifications and 3D nuclear organization : data, models and function, *Cur. Op. Cell Biol.* 44, 20-27 (2017).
- [74] M. Feric, N. Vaidya, T. S. Harmon, D. M. Mitrea, L. Zhu, T. M. Richardson, R. W. Kriwacki, R. V. Pappu, C. P. Brangwynne, Coexisting Liquid Phases Underlie Nucleolar Subcompartments, *Cell* 165, 1686-1697 (2016).

-
- [75] A. G. Larson, D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding, G. J. Narlikar, Liquid droplet formation by HP1 α suggests a role for phase separation in heterochromatin, *Nature* 547, 236-240 (2017).
- [76] F. Erdel, K. Rippe, Formation of Chromatin Subcompartments by Phase Separation, *Biophysical Journal* 114 (2018).
- [77] I. Junier, O. Martin and F. Képès, Spatial and Topological Organization of DNA Chains Induced by Gene Co-localization, *PLoS Comput. Biol.* 6, e1000678 (2010).
- [78] V. F. Scolari, M. Cosentino Lagomarsino, Combined collapse by bridging and self-adhesion in a prototypical polymer model inspired by the bacterial nucleoid, *Soft Matter* 11, 1677 (2015).
- [79] M. Barbieri, M. Chotalia, J. Fraser, L.-M. Lavitas, J. Dostie, A. Pombo, M. Nicodemi, Complexity of chromatin folding is captured by the strings and binders switch model, *Proc. Nat. Acad. Sci. U.S.A.* 109, 16173-16178 (2012).
- [80] G. Le Treut, F. Képès, H. Orland, Phase Behavior of DNA in the Presence of DNA-Binding Proteins, *Bio. J.* 110, 51-62 (2016).
- [81] J. Johnson, C. A. Brackley, P. R. Cook, D Marenduzzo, A simple model for DNA bridging proteins and bacterial or human genomes : bridging-induced attraction and genome compaction, *J. Phys. : Condens. Matter* 27, 064119 (2015).
- [82] M. Toda, R. Kubo, N. Saito, *Statistical Physics I, Equilibrium Statistical Mechanics*, Springer (1973).
- [83] J.-L. Barrat, J.-P. Hansen, *Basic Concepts for Simple and Complex Liquids*, Cambridge University Press (2003).
- [84] W. Lenz, Beitrag zum Verständnis der magnetischen Erscheinungen in festen Körpern, *Physikalische Zeitschrift* 21, 613-615 (1920).
- [85] T. Ising, R. Folk, R. Kenna, B. Berche, Y. Holovatch, The Fate of Ernst Ising and the Fate of his Model, *arXiv :1706.01764* (2017).
- [86] B. McCoy and T. T. Wu, *The Two-Dimensional Ising Model*, Cambridge, Massachusetts, Harvard UP (1973).
- [87] S. R. A. Salinas, *The Ising Model dans Introduction to Statistical Physics Graduate Texts in Contemporary Physics*, Springer (2001).
- [88] J. Cardy, *Scaling and Renormalization in Statistical Physics*, Cambridge University Press (1996).
- [89] P. Weiss, L'hypothèse du champ moléculaire et la propriété ferromagnétique, *Journal de Physique Théorique* 6, 661-690 (1907).

- [90] L.D. Landau and E.M. Lifshitz, *Statistical Physics, Third Edition, Part 1 : Volume 5 (Course of Theoretical Physics, Volume 5)*, Pergamon Press, Oxford (1980).
- [91] R. P. Feynman, *Statistical mechanics : A set of lectures*. Reading, Mass : W.A. Benjamin (1972).
- [92] L. Onsager, *Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition*, *Physics Review* 65, 117-149 (1944).
- [93] K. Binder, E. Luijten, *Monte Carlo tests of renormalization-group predictions for critical phenomena in Ising models*, *Physics Report* 344(3-6), 179-253 (2001).
- [94] J. E. Mayer and M. G. Mayer, *Statistical mechanics*. 2nd ed. John Wiley & Sons, Inc., New York (1977).
- [95] P. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY (1971)
- [96] P. De Gennes, *Collapse of a polymer chain in poor solvents*, *J. Phys. Lett.* 36, 55 (1975).
- [97] P.-G. de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca (1979).
- [98] R. D. Schram, G. T. Barkema¹ and R. H. Bisseling, *Exact enumeration of self-avoiding walks*, *J. Stat. Mech.* P06019 (2011).
- [99] F. J. Dyson, *Existence of a Phase-Transition in a One-Dimensional Ising Ferromagnet*, *Commun. Math. Phys.* 12, 91-107 (1969).
- [100] P. W. Anderson and G. Yuval, *Some numerical results on the Kondo problem and the inverse square one-dimensional Ising model*, *J. Phys. C* 4, 607 (1971).
- [101] J. M. Kosterlitz, D. J. Thouless, *Ordering, metastability and phase transitions in two-dimensional systems*, *J. Phys. C : Solid State Phys.* 6, 1181-1203 (1973).
- [102] D. Mukamel, *Notes on the Statistical Mechanics of Systems with Long-Range Interactions*, arXiv :0905.1457 (2009).
- [103] H. Schiessel, *Biophysics for Beginners : a journey through the cell nucleus*, Singapour, Pan Stanford Publishing (2014).
- [104] K. G. Petrosyan, C.-K. Hu, *Protein-mediated loops and phase transition in nonthermal denaturation of DNA*, *J. Stat. Mech.*, P01005 (2011)
- [105] L. J. Kijewski and M. P. Kawatha, *One-Dimensional Ising Model with Long-Range Interaction*, *Physics Letters* 31A, 479 (1970).
- [106] E. Luijten, H. W. J. Blöte, *Classical critical behavior of spin models with long-range interactions*, *Phys. Rev. B* 56, 8945 (1997).

-
- [107] A. Hanke, R. Metzler, Entropy loss in long-distance DNA looping, *Biophysical Journal* 85, 167-73 (2003).
- [108] T. Garel, H. Orland, E. Orlandini, Phase diagram of magnetic polymers, *Eur. Phys. J. B* 12, 261-268 (1999).
- [109] J.-L. Barrat, J.-F. Joanny, Theory of polyelectrolyte solutions, *Advances in Chemical Physics : Polymeric Systems* 94, 1-66 (1996).
- [110] R. Everaers, S. Kumar, C. Simm, Unified description of poly- and oligonucleotide DNA melting : Nearest-neighbor, Poland-Sheraga, and lattice models, *Phys. Rev. E* 75, 041918 (2007).
- [111] M. E. Cates, J. Tailleur, Motility-Induced Phase Separation, *Annual Review of Condensed Matter Physics* 6, 219-244 (2015).
- [112] A. P. Solon, J. Stenhammar, M. E. Cates, Y. Kafri, J. Tailleur, Generalized thermodynamics of phase equilibria in scalar active matter, *Physical Review E* 97, 020602 (2018).
- [113] D. Zwicker, A. A. Hyman, F. Jülicher, Suppression of Ostwald ripening in active emulsions, *Physical Review E* 92, 012317 (2015).
- [114] E. Tjhung, C. Nardini, M. E. Cates, Cluster Phases and Bubbly Phase Separation in Active Fluids : Reversal of the Ostwald Process, *Physical Review X* 8, 031080 (2018).
- [115] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, Madan Rao, R. Aditi Simha, Hydrodynamics of soft active matter, *Reviews of Modern Physics* 85, 1143-1189 (2013).
- [116] J. Prost, F. Jülicher, J.-F. Joanny, Active gel physics, *Nature Physics* 11, 111-117 (2015).
- [117] K. Kruse, J. F. Joanny, F. Jülicher, J. Prost, K. Sekimoto, Generic theory of active polar gels : a paradigm for cytoskeletal dynamics, *The European Physical Journal E* 16, 5-16 (2005).
- [118] J. L. Anderson, Transport Mechanisms of Biological Colloids, *Annals of The New York Academy of Sciences* 469, 166-177 (1986).
- [119] A. Zöttl, H. Stark, Emergent behavior in active colloids, *Journal of Physics : Condensed Matter* 28, 253001 (2016).
- [120] S. Thakur, R. Kapral, Dynamics of self-propelled nanomotors in chemically active media, *The Journal of Chemical Physics* 135, 024509 (2011).
- [121] H. Risken, T. Frank, *The Fokker-Planck Equation : Methods of Solution and Applications*, Springer-Verlag Berlin Heidelberg (1996).
- [122] A. Cornish-Bowden, *Fundamentals of Enzyme Kinetics*, Portland Press (2004).
- [123] L. Onsager, Reciprocal Relations in Irreversible Processes I, *Physical Review* 37, 405-426 (1931).

BIBLIOGRAPHIE

- [124] L. Onsager, Reciprocal Relations in Irreversible Processes II, *Physical Review* 38, 2265-2279 (1931).
- [125] J. W. Cahn, J. E. Hilliard, Free Energy of a Nonuniform System. I. Interfacial Free Energy, *Journal of Chemical Physics* 28, 258-267 (1958).
- [126] W. Ostwald, Studien über die Bildung und Umwandlung fester Körper, *Zeitschrift für Physikalische Chemie* 22, 289–330 (1897).
- [127] R. Ietswaart, F. Szardenings, K. Gerdes, M. Howard, Competing ParA Structures Space Bacterial Plasmids Equally over the Nucleoid, *PLOS Computational Biology* 10, e1004009 (2014).
- [128] M. A. Muñoz, Colloquium : Criticality and dynamical scaling in living systems, *Review of Modern Physics* 90, 031001 (2018).