

Thèse d'habilitation à diriger des recherches de
l'Université Louis Pasteur de Strasbourg

Modélisation de petits systèmes
biophysiques -
Dynamique d'algorithmes de
recherche

Simona Cocco

8 octobre 2004

présentée le 22 novembre 2004 devant la commission :

Armand Ajdari

Didier Chatenay (garant)

Ulrich Bockelmann

Nicolas Brunel

Jean-Pierre Munch (rapporteur interne)

Henri Orland (rapporteur)

Felix Ritort (rapporteur)

Remerciements

Je remercie tout d'abord D. Chatenay, J. Marko et R. Monasson pour m'avoir encouragée dans mon activité au cours des dernières années et partagé leurs savoirs et passions pour la recherche.

Ensuite, je remercie tous les membres du Laboratoire des Fluides Complexes de Strasbourg et du Laboratoire de Physique Statistique de l'ENS, en particulier les secrétaires J. Haccoun, N. Bouaouina, N. Sadaoui et C. Philippe.

Une grande partie de cette thèse a été rédigée aux Houches pendant l'Ecole d'Eté d'août 2004 et je tiens à remercier J. Dalibard, I. Lelièvre et B. Rousset pour leur affabilité et disponibilité pendant tout le séjour.

Pour finir, je remercie tous les membres du jury, A. Ajdari, U. Bockelmann, N. Brunel, J-P. Munch, H. Orland et F. Ritort pour avoir accepté d'en faire partie.

Table des matières

1	Introduction.	7
1.1	Curriculum Vitae	7
1.2	Thèmes développés	7
2	Modélisation de l'ADN et de l'ARN.	9
2.1	Introduction aux micromanipulations de molécules uniques et à leur théorie	9
2.2	Dégraissage de l'ADN et de l'ARN à force fixée : thermodynamique	11
2.3	Dégraissage à force fixée : nucléation de l'ouverture	14
2.4	Dégraissage à force fixée : propagation de l'ouverture et effets de séquence .	18
2.5	Dégraissage à vitesse d'éloignement des extrémités constante	26
2.6	Elasticité de molécules avec des interactions entre les monomères	34
2.7	Thermodynamique et dynamique de surétirement de l'ADN	37
3	Etude de la dynamique persistante d'un neurone isolé	43
3.1	Système expérimental et persistance de l'activité.	43
3.2	Modèle de Hodgkin-Huxley pour une autapse	45
3.3	Modèle Integrate-and-Fire pour une autapse	47
4	Mécanique statistique et problèmes d'optimisation	51
4.1	Dynamique d'algorithmes de décision.	51
4.2	Procédure de redémarrage aléatoire.	57
4.3	Etats fondamentaux des modèles de verre de spins sur réseaux aléatoires. .	60
5	Travaux en cours et projets	63
5.1	Reconstruction de séquences d'ADN à partir de données de micromanipulation	63
5.2	Formation de boucles dans l'ADN	70
5.3	Distribution de fluorescence dans une population de bactéries monoclonale.	74
6	Autre activités	77
6.1	Encadrement d'étudiants	77
6.2	Organisation de congrès et d'écoles.	77
6.3	Diffusion de l'information scientifique.	78
6.4	Allocations de recherche.	78

7	Liste des publications	79
	Bibliographie	83

Chapitre 1

Introduction.

1.1 Curriculum Vitae

Lieu et date de naissance : Rome (Italie), le 6/5/1970

Titres :

20/5/2000 Doctorat en Physique de L'Ecole Normale Supérieure de Lyon et
Doctorat en Biophysique de L'Université de Rome La Sapienza

Post-doctorats et séjours à l'étranger :

11/99-8/2000 Laboratoire de Physique Théorique de L'Ecole Normale Supérieure, Paris.

9/00-8/01 Physics Department of The University of Illinois at Chicago.

9/01-10/01 International Center for Theoretical Physics, Trieste Italy.

Prise de fonction au CNRS :

10/2001 Section 05, au Laboratoire de Dynamique des Fluides Complexes
UMR 7506, Strasbourg.

2/2004 Changement d'affiliation au Laboratoire de Physique Statistique de l'Ecole Normale Supérieure, Paris.

1.2 Thèmes développés

Cette thèse d'habilitation à diriger des recherches comporte les travaux que j'ai effectués pendant mon premier post-doc, au LPT-ENS à Paris, mon deuxième post-doc à l'UIC à

Chicago et mes trois premières années d'activité au CNRS au sein du LDFC à Strasbourg et, à partir de février 2004, au LPS à Paris.

J'ai développé essentiellement deux thèmes de recherche :

- la modélisation des systèmes biophysiques, qui inclut une partie sur l'ADN, une partie sur la modélisation de systèmes neuronaux et, dans les projets en cours, une partie sur l'étude de variations phénotypiques dans de populations de bactéries.
- l'étude des problèmes d'optimisation centrée autour de la modélisation de la dynamique de résolution du problème de Satisfiabilité (SAT) par un algorithme de recherche en arbre de Davis-Putnam (DP).

Ces deux thèmes sont apparemment très éloignés l'un de l'autre mais sont, en fait, reliés par l'utilisation de techniques analytiques similaires, propres à la mécanique statistique. Parmi ces techniques, mentionnons : le formalisme de l'intégrale de transfert et son analogue quantique à une dimension, la méthode des répliques pour les systèmes désordonnés, les calculs instantoniques à la Langer et le formalisme de Kramers pour les processus activés (passage de barrière), les techniques de fonctions génératrices ... En retour, l'étude de ces systèmes permet parfois de mieux comprendre la validité de ces techniques, comme en témoigne le travail sur les états fondamentaux dans les verres de spins de champ moyen. Dans mes activités de biophysique, je dédie une grande attention à la comparaison avec les données expérimentales qui se trouvent dans des travaux publiés par des équipes qui ne travaillent pas au sein du laboratoire (comme dans les études de la dynamique de dégraffage de l'ADN et l'ARN, de l'élasticité des molécules avec des interactions entre les monomères et de la thermodynamique et dynamique de surétirement de l'ADN) ou être le fruit d'une collaboration plus étroite avec les expérimentateurs du laboratoire (par exemple dans l'étude de la dynamique spontanée de réseaux de neurones en culture et dans l'étude de l'évolution statistique d'une population bactérienne monoclonale).

Chapitre 2

Modélisation de l'ADN et de l'ARN.

2.1 Introduction aux micromanipulations de molécules uniques et à leur théorie

L'article A1 présente une revue centrée avant tout sur les expériences de micromanipulation de molécules uniques, tandis que l'article A2 traite essentiellement des modèles.

Toutes les cellules vivantes contiennent des longues molécules d'ADN (Fig. 2.1). Ces molécules renferment l'information génétique à travers leur séquence. Elles doivent pouvoir être lues afin que les protéines puissent être produites, et doivent aussi être dupliquées et séparées les unes des autres afin que les cellules puissent se diviser. En outre, si l'ADN d'une cellule est endommagé, il doit être réparé. Tous ces processus essentiels sont accomplis par des enzymes, c'est-à-dire des protéines qui catalysent des réactions biochimiques et qui travaillent le long de la molécule. Les expériences de micromanipulation développées par des physiciens à partir des années 80 et portant sur des molécules isolées d'ADN et d'ARN représentent une nouvelle approche visant à la compréhension de la manière dont l'ADN est organisé et traité dans les cellules à l'aide d'une caractérisation biomécanique de ces molécules.

Un des thèmes de recherche que j'ai développés s'inscrit dans le cadre de la modélisation des systèmes biologiques, en rapport direct avec les expériences de micromanipulation. On pourrait se demander quel est le rôle de la physique dans ses expériences de micromanipulation et dans leur théorie. La physique entre en jeu d'abord dans le développement des instruments et de techniques expérimentales. En outre, elle aide à la compréhension et caractérisation mécanique des biomolécules, à commencer par leurs propriétés élastiques et de déformation, et des interactions avec les enzymes. On cherche par exemple à caractériser avant toute chose les forces en jeu dans de telles interactions. Il est important de remarquer que ces forces sont de l'ordre du piconewton et les interactions ont lieu à l'échelle du nanomètre, ce qui donne lieu à des énergies comparables aux fluctuations thermiques ($1 k_B T = 4.1 \text{ pN nm}$). La mécanique statistique est donc un outil fondamental pour l'analyse de ces processus. La théorie des expériences sur molécules uniques est un

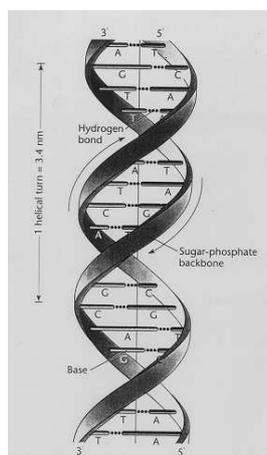


FIG. 2.1 – La double hélice de Watson et Crick de la conformation ADN B. Les deux brins sont composés par un squelette sucre–phosphate auquel les bases sont attachées comme des barreaux d’une échelle; les deux brins sont liés par les liaisons hydrogènes entre bases. L’adénine (A) est appariée avec la Thymine (T) par l’intermédiaire de deux liaisons hydrogène, et la Guanine (G) se lie à la Cytosine (C) par trois liaisons hydrogène.

nouveau développement de la physique des polymères. Elle est stimulée par la comparaison avec les données expérimentales, facilitée par la simplicité et le contrôle du système étudié, qui étaient plus difficiles à obtenir dans le cadre de l’étude classique des systèmes biologiques. Pour citer quelques exemples, les expériences de micromanipulation ont permis la caractérisation des propriétés élastiques, sous l’action d’une force, d’une molécule d’ADN double brin (voir les travaux des groupes de Bustamante [1, 3] et Chatenay [2]), d’ADN simple brin (Bustamante [5, 3], Bensimon et Croquette[6]) ou d’ARN (Bustamante [7], Chatenay [9]), et également quand la molécule est soumise à un couple de torsion (Bensimon et Croquette [10], Chatenay [11]). D’autres manipulations visent à comprendre l’action de certains enzymes : l’ouverture mécanique des deux brins de la double hélice de l’ADN suite à l’interaction avec, par exemple, l’ARN-polymérase qui copie l’ADN en ARN pendant la transcription ou l’ADN-polymérase [12] qui reconstruit un double brin d’ADN à partir d’un simple brin et des nucléotides éparpillés en solution. La première étape de la théorie est celle qui consiste à décrire les expériences d’étirement de molécules uniques d’ADN faites à faible vitesse, de manière quasi-statique. Elle est basée sur les modèles d’élasticité des polymères à l’équilibre[13]. Mais la richesse de la structure des biopolymères doit souvent être prise en compte dans la théorie à cause des transformations des structures secondaires (exemples en sections 2.2, 2.5 et 2.7), des inhomogénéités dues à la séquence (2.4), des interactions entre monomères qui dépendent de la séquence (2.6), ... Les expériences d’étirement d’un double brin d’ADN témoignent à la fois de la présence de comportements élastiques réguliers et de transitions structurelles brutales dues à des transformations de la structure secondaire (2.7). Les expériences d’étirement d’un simple brin d’ADN ou d’un ARN

reflètent, d'autre part, les interactions d'appariement entre les bases (2.6). La deuxième étape consiste en l'étude des propriétés d'équilibre qui induisent des modifications dans les molécules de base, par exemple des interactions avec des protéines qui agissent sur l'ADN sous contrôle mécanique. Ainsi les protéines autour desquelles l'ADN s'enroule (histones) [14], ou qui permettent à l'ADN de former des boucles (LAC)[15] peuvent être modélisées en prenant en compte l'énergie d'interaction ADN-protéine et le rétrécissement de la partie étirable de l'ADN [16] (2.6).

La théorie fait appel à de nombreuses techniques de physique statistique et de mécanique quantique. Une molécule unique d'ADN est composée d'un nombre très élevé d'unités élémentaires (monomères). L'existence d'interaction entre premiers voisins (monomères contigus) amène souvent à des fonctions de partition ayant la forme d'intégrales de chemin, où l'abscisse curviligne joue le rôle du temps, qui peuvent être analysées avec les méthodes de mécanique quantique. Les idées de la théorie des transitions de phases sont aussi utilisées intensivement, par exemple pour décrire la transition brutale observée dans l'étirement d'une seule molécule d'ADN. La troisième phase consiste dans l'étude de phénomènes dynamiques, par exemple la dynamique du dégraissage à haute vitesse (2.5,2.7) ou sous taux de charge élevé (2.3), ou encore les interactions avec des enzymes qui consomment de l'ATP.

2.2 Dégraissage de l'ADN et de l'ARN à force fixée : thermodynamique

En collaboration avec R. Monasson (LPT-ENS Paris) et J. Marko (UIC Chicago) (Publications A3, A4, A5), nous avons étudié la dynamique de dégraissage de l'ADN et l'ARN sous une force tendant à éloigner les deux brins complémentaires de la double hélice (voir Fig.2.2). Nous avons d'abord considéré le cas d'une force fixée aux deux extrémités de la molécule. Nous avons étudié plusieurs phénomènes d'équilibre et hors équilibre reliés à ces expériences : dans cette section, j'introduis la thermodynamique du dégraissage quand la molécule est soumise à une force de dégraissage et un couple de torsion fixés ; la section 2.3 est consacrée à l'étude du temps de nucléation de l'ouverture et de la dynamique d'ouverture à taux de charge fixé pour des séquences homogènes. En section 2.4, j'introduis l'étude des effets dûs à la séquence sur la dynamique de dégraissage. Le cas de l'ouverture à vitesse d'éloignement des deux bras constante est analysé en section 2.5.

Si, dans la configuration expérimentale de la Fig.2.2, la force est suffisante, on constate qu'un certain nombre de bases (n) s'ouvrent à l'équilibre thermodynamique. Il s'agit, pour commencer, d'obtenir l'expression de la force minimale permettant cette ouverture. En première approximation, la force d'ouverture est obtenue en comparant l'énergie libre gagnée à la suite de l'ouverture $f \times d$, où $d \simeq 1$ nm est la longueur de deux bases ouvertes, avec l'énergie de liaison de ces deux bases qui est de l'ordre de $2.5 k_B T = 10$ pN nm. On obtient ainsi une force de dégraissage de l'ordre de $f = 10$ pN. Pour être plus précis, avec une force f aux extrémités de la molécule, l'énergie libre G de la molécule avec n paires de bases ouvertes est la différence entre l'énergie libre des deux simples brins

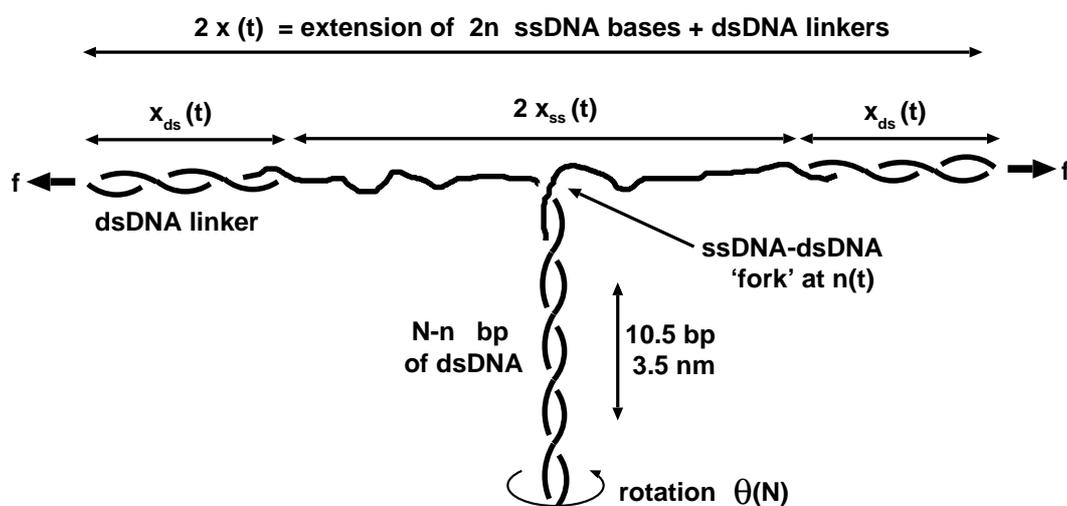


FIG. 2.2 – Configuration des expériences de dégraissage : deux bras de double brins d'ADN sont attachés à la molécule à ouvrir et utilisés pour éloigner les extrémités 5' et 3' adjacentes de la molécule. L'expérience peut être faite en appliquant une force f constante sur les extrémités des bras, ou en écartant les bras à vitesse $v = dx/dt$ constante.

(ssDNA) étendus, chacun comportant n bases, et l'énergie libre perdue lors de l'ouverture des n premières paires de bases ($i = 1, \dots, n$) de l'ADN double brin (dsDNA) :

$$G(f, n) = 2 n \mathcal{F}_{ss}(f) - \sum_{i=1}^n g_{ds}(i) \quad (2.1)$$

L'énergie libre par paire de base de l'ADN étendu, $\mathcal{F}_{ss}(f)$, peut être écrite, pour des forces $f \simeq 10$ pN, en utilisant le modèle de polymère à articulation libres Freely Jointed Chain (FJC, voir A2) modifié (FJCL) par le groupe de Bustamante pour reproduire la courbe expérimentale de force-extension de l'ADN simple brin [1]. A ces grandes forces, les effets d'appariement entre nucléotides du ssDNA sont absents. Une discussion plus approfondie des modèles d'élasticité du ssDNA est présentée en section 2.6. Nous commençons par considérer une séquence homogène, où toutes les paires de bases ont une énergie libre d'appariement $g_{ds} = -g_0$. La force critique de dégraissage, f_u , est simplement donnée par la condition :

$$G(f_u, n) \equiv n g(f_u) = [2 \mathcal{F}_{ss}(f_u) + g_0] n = 0 \quad (2.2)$$

Pour $f < f_u$, l'ADN double brin est thermodynamiquement stable ; si $f > f_u$, la double hélice s'ouvre de manière similaire à ce qu'il advient lors d'une transition du premier ordre. Nous nous sommes efforcés d'abord à obtenir une expression simple de la force de dégraissage en fonction de l'énergie libre d'appariement de bases. L'expansion quadratique de $\mathcal{F}_{ss}(f)$ autour de $f = 10$ pN donne l'énergie libre d'un polymère gaussien, $\mathcal{F}_{ss}^{GP}(f) = -f^2 b^2 / (6k_B T)$ avec une longueur de Kuhn effective $b = 7 \text{Å}$. En utilisant ce modèle gaussien pour le ssDNA, nous obtenons $f_u^{GP} = \sqrt{3k_B T g_0} / b$. La force d'ouverture d'une séquence homogène peut donc permettre d'obtenir une mesure directe de l'énergie libre de liaison des paires de bases ou des énergies libres relatives de l'ADN sous formes ss et ds à température fixée.

Rief *et al* ont mesuré la force de dégraissage f_u pour des ADN ayant différentes séquences répétées [17]. Ils ont trouvé $f_u(\text{poly dA-dT}) = 9 \pm 3$ pN et $f_u(\text{poly dG-dC}) = 20 \pm 3$ pN, ce qui donne $g_0^{GC}(A-T) = 0.8 k_B T$, $g_0^{FJCL}(A-T) = 1.1 k_B T$, $g_0^{GC}(G-C) = 4.2 k_B T$, et $g_0^{FJCL}(G-C) = 3.5 k_B T$ respectivement. Ces valeurs pour les énergies libres de dénaturation sont comparables avec les données thermodynamiques issues de la dénaturation de l'ADN [18].

Considerons maintenant le cas d'une molécule d'ADN soumise à un couple de torsion Γ à l'extrémité opposée à celle qui est ouverte. Le travail de déroulement contribue à l'énergie libre de dégraissage (par paire de base) :

$$g(f, \Gamma) = 2\mathcal{F}_{ss}(f) + g_0 + \Theta_0 \Gamma \quad (2.3)$$

$\Theta_0 = 2\pi/10.4$ correspond à la variation de l'angle de twist pendant la conversion du dsDNA en ssDNA. Le diagramme de phase pour la transition de dégraissage dans le plan force-torsion pour une séquence uniforme avec $g_0 = 2.5k_B T$ est montré en Fig. 2.3. La force critique que l'on trouve à torsion nulle autour de $f_u = 15$ pN est en bon accord avec les résultats obtenus par micromanipulation d'une molécule unique notamment dans le groupe

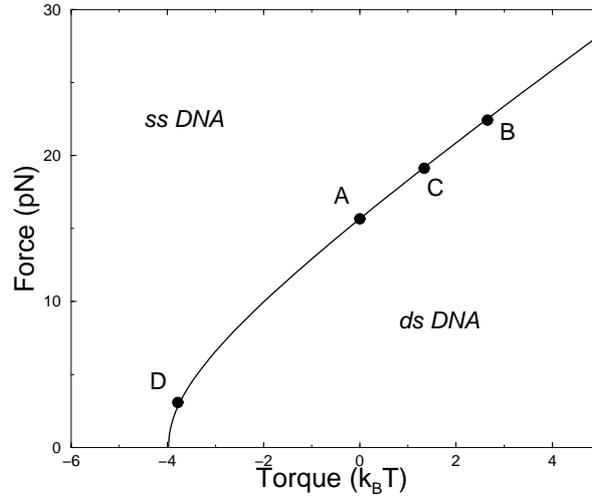


FIG. 2.3 – Diagramme de phase du dégraissage de l'ADN en fonction du couple Γ en unité $k_B T$ et de la force f en pN pour une séquence homogène avec une énergie libre d'appariement de $g_0 = 2.5k_B T$. A couple nul, la force de dégraissage vaut $f_u \simeq 15$ pN. En présence d'un couple positif, la force f_u augmente; elle diminue lorsque le couple est négatif, jusqu'à s'annuler lorsque $\Gamma \simeq -4k_B T$. Les points A,B,C,D se referrent aux figure 2.16, 2.17 .

de Bockelmann et Heslot à Paris. Cette transition a été aussi étudiée d'un point de vue théorique, sans couple de torsion mais sur un ADN avec une séquence désordonnée, par Lubensky et Nelson [19]. D'autre part, le couple critique à force nulle est, pour $g_0 = 2.5k_B T$, de l'ordre de $\Gamma \sim -4k_B T$. Pour la séquence riche en AT que nous avons considérée dans les Publications A3 et A4, $g_0 = 1.4k_B T$, et nous obtenons la valeur $\Gamma \sim -2k_B T$ en bon accord avec les expériences du groupe de Bensimon et Croquette [10] sur l'ouverture par un couple de torsion de bulles correspondant aux zones riches en AT dans la séquence du λ -phage et les travaux théoriques que j'ai effectués pendant la thèse [20] et ceux du groupe de Marko [21].

Dans le cas de séquences hétérogènes, l'énergie libre pour ouvrir les n premières paires de bases, $G(f, n)$ (voir Article A5), peut être calculée en utilisant les énergies d'appariement qui dépendent de la séquence, $g_{ds}(i)$ (par exemple grâce au serveur Mfold [22]).

Passons maintenant à l'étude de la nucléation de l'ouverture et la propagation de cette ouverture le long de la molécule.

2.3 Dégraissage à force fixée : nucléation de l'ouverture

Voir les Publications A3 et A4. En première approximation, en absence de force d'étirement, le temps nécessaire à l'ouverture d'une séquence homogène est proportionnel à l'exponentielle de l'énergie libre d'ouverture $G = \sum_{i=1}^n g_0(i)$ où n est le nombre de bases de la molécule

et g_0 varie entre 1 et $4 k_B T$ selon la séquence. Cela signifie que des molécules de plus de 30 bases ne sont pas susceptibles de se séparer par fluctuation thermique parce que cela nécessiterait des fluctuations survenant sur des échelles de temps déraisonnables. Une approche plus appropriée pour étudier la dissociation doit tenir compte de la manière avec laquelle la double hélice s'ouvre. Quand les liaisons hydrogènes se cassent, les bases sont encore empilées dans la double hélice ; c'est seulement en poursuivant l'ouverture que les bases sortent de l'empilement et les deux simples brins se détachent, en conférant aux bases une plus grande mobilité par rapport à leur configuration empilée. La configuration avec les bases ouvertes mais empilées est très coûteuse en énergie libre et représente donc une barrière à franchir lors de l'ouverture. En outre, cette barrière doit être traversée par la chaîne des bases, c'est-à-dire une séquence de particules liées entre elles. Il faut donc étudier la configuration optimale de la chaîne polymérique qui amène d'une base ouverte à l'extrémité de la molécule à une base fermée après un certain nombre de bases. Cette configuration définit la fourche d'ouverture. Les deux questions que nous nous sommes posées dans l'article A3 sont les suivantes : combien de bases composent une telle configuration ? quelle est son énergie libre ? La réponse à cette dernière question donne accès à l'énergie d'activation et donc le temps de formation de la fourche d'ouverture. Ce temps est à multiplier par l'exponentielle de la différence d'énergie libre entre le ssDNA et le dsDNA, $\Delta G = n g_0$, pour obtenir le temps d'ouverture. Nous avons ainsi calculé les constantes de dissociation cinétiques pour une séquence homogène en fonction de la force de dégraissage et de la longueur de la molécule. Une fois obtenu le temps de dissociation, ou son inverse, c'est-à-dire la probabilité par unité de temps de l'ouverture nous avons également considéré la situation expérimentale, rencontrée en particulier dans les expériences de microscopie à force atomique (AFM), dans laquelle la force est incrémentée de façon graduelle (à taux de charge constant). La rupture intervient alors à une certaine force, dont la valeur typique dépend de la longueur de la molécule et du taux de charge. Nous avons calculé analytiquement cette force de rupture. Pour calculer la forme optimale de la fourche d'ouverture, un modèle semi-microscopique où l'on tient en compte le potentiel d'empilement, doit être utilisé. Nous avons fait appel au modèle que j'avais étudié en collaboration avec M. Barbi pendant ma thèse de doctorat. Ce modèle incorporait la structure en double hélice du dsDNA (forme B) au modèle de Peyrard-Bsichop (PB) [23] et il avait permis d'étudier le couplage entre ouverture et déroulement pour la dynamique d'ouverture et la transition de dénaturation [20].

Comme on le voit sur la Figure 2.4, dans ce modèle l'axe de la molécule est droit. On introduit trois degrés de liberté par paire de bases : le rayon de la paire r_j , son angle de rotation φ_j dans le plan orthogonal à l'axe de l'hélice et sa hauteur z_j le long de l'axe. Une contrainte pour chaque paire de bases est aussi introduite : la longueur le long d'un brin entre deux bases adjacentes, c'est-à-dire la distance entre les deux phosphates sur le squelette, est fixée à $L = 7\text{\AA}$. De cette manière on peut éliminer un degré de liberté en écrivant la distance entre deux paires de bases le long de l'axe $h_j = z_j - z_{j-1}$ comme $h_j(r_j, r_{j-1}, \varphi_j - \varphi_{j-1}) = \sqrt{L^2 - r_j^2 - r_{j-1}^2 + 2 r_j r_{j-1} \cos(\varphi_j - \varphi_{j-1})}$. Précisons que les configurations des rayons et angles qui donnent des valeurs imaginaires de h_j ne respectent pas

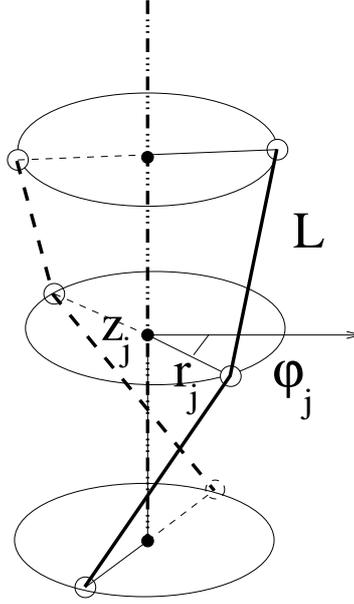


FIG. 2.4 – Représentation schématique du modèle de la double hélice d'ADN. Chaque paire de bases est prise en compte par son rayon r_j , son angle φ_j (par rapport à une direction arbitraire fixe) et sa hauteur z_j le long de l'axe de la molécule. La distance axiale entre deux bases successives et l'angle de twist sont respectivement égaux à $h_j = z_j - z_{j-1}$ et $\theta_j = \varphi_j - \varphi_{j-1}$. La distance L entre deux bases adjacentes le long d'un brin est fixée à 7\AA .

la contrainte rigide et doivent être éliminées en posant que l'énergie potentielle est infinie dans ce cas. Pour des configurations physiquement acceptables, l'énergie potentielle s'écrit

$$\begin{aligned}
 V[\{r_j, \varphi_j\}] &= \sum_{j=1}^N D \left(e^{-a(r_j-R)} - 1 \right)^2 + \sum_{j=2}^N K \left(h_j(r_j, r_{j-1}, \varphi_j - \varphi_{j-1}) - H \right)^2 \\
 &+ \sum_{j=2}^N E e^{-b(r_j-r_{j-1}-2R)} (r_j - r_{j-1})^2
 \end{aligned} \tag{2.4}$$

où $R = 10\text{\AA}$ est le rayon de repos des paires de base. Comme dans le modèle PB, le premier terme est le puits de potentiel dû aux liaisons hydrogènes, le quatrième terme est le potentiel d'empilement entre bases adjacentes le long de la séquence. En plus, un potentiel harmonique dans la distance axiale $h_j(r_j, r_{j-1}, \varphi_j - \varphi_{j-1})$ reproduit les interactions hydrophobiques qui empaquettent les paires de bases les unes sur les autres avec une distance axiale moyenne $H \simeq 3.4\text{\AA}$ et donne sa structure hélicoïdale à l'ADN.

La contrainte de distance fixe entre deux bases le long de chaque brin est le seul élément essentiel pour la description de la phase dénaturée. Chaque brin, dans cette phase, est donc représenté par une suite de bâtonnets rigides de longueur L , libres de s'articuler les

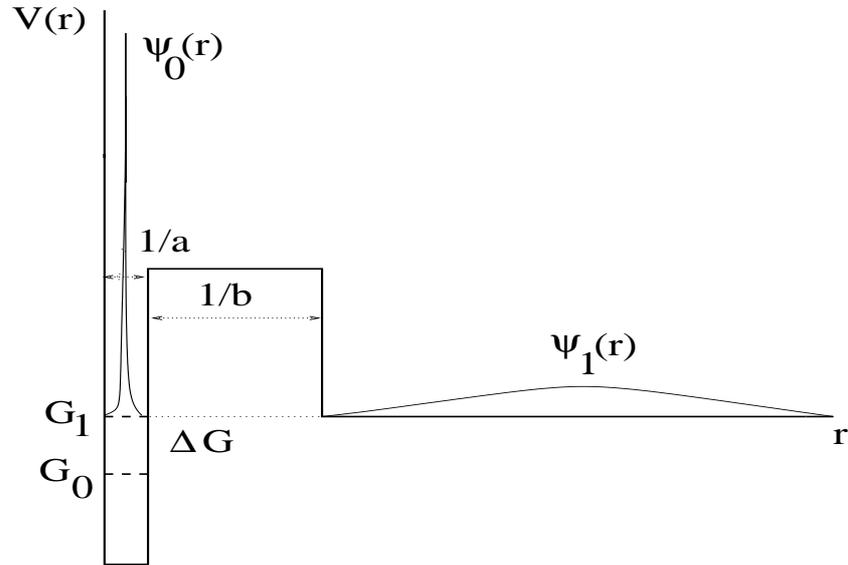


FIG. 2.5 – Allure qualitative du potentiel unidimensionnel $V(r)$ effectif déduit de la diagonalisation de la matrice de transfert en fonction du rayon de la paire de bases r . Le puits de potentiel à gauche représente la liaison hydrogène avec une largeur $1/a$ (voir l'énergie du modèle donnée par l'équation (2.4)), la barrière dont la largeur dépend du paramètre $1/b$ est due aux effets d'empilements. Pour $r > 14 \text{ \AA}$, le potentiel peut être considéré comme nul (région dénaturée). A température ambiante, la fonction d'onde de l'état lié fondamental $\psi_0(r)$ avec énergie libre G_0 décrit un dsDNA et la fonction d'onde du premier état excité (délocalisé) $\psi_1(r)$, avec énergie libre $G_1 \equiv G_0 + \Delta G$, décrit un ssDNA.

uns par rapports aux autres ; c'est le modèle de la Freely Jointed Chain (FJC). L'énergie libre obtenue à partir du potentiel 2.4 se calcule par la diagonalisation de la matrice de transfert. Les fonctions propres de la matrice de transfert donnent la probabilité que les deux bases d'une même paire soient à distance $2r$. Pour calculer la configuration de la fourche optimale et son énergie libre, nous avons utilisé une version continue de notre modèle, où l'indice discret des paires de bases, j , est traité comme une variable continue. Les fonctions propres de la matrice de transfert correspondent, dans la version continue, aux fonctions d'ondes d'une équation de Schrödinger avec le potentiel $V(r)$ esquissé Figure 2.5. La barrière du potentiel est d'origine entropique : comme la rigidité de la molécule est beaucoup plus grande dans l'état fermé (dsDNA) que dans l'état ouvert (ssDNA), l'entropie des configurations explorables est plus élevée dans cette dernière conformation que dans la première, induisant un potentiel répulsif en r qui favorise les grandes valeurs de r . La superposition de ce potentiel à celui de la liaison hydrogène conduit à la Figure 2.5.

Cette extension du modèle permet d'utiliser commodément des méthodes d'approximation de mécanique quantique (WKB), la théorie de la nucléation de Langer (appliquée ici à une théorie des champs unidimensionnelle) et la théorie d'Evans de la rupture de liaisons

simples.

Les données expérimentales de cinétique de dissociation de petites molécules, obtenues notamment par Pörschke dans les années 70 pour une solution de molécules et plus récemment par Bonnet et collaborateurs sur des molécules uniques, nous ont permis de déterminer la valeur du paramètre $b = 1.75\text{\AA}^{-1}$ du potentiel (2.4) qui contrôle la largeur de la barrière. En particulier avec ce choix nous obtenons la valeur de la dimension de la fourche d'ouverture $n = 4$ en bon accord avec la valeur $n = 3$ mesure par Pörschke [24]. En outre, à force nulle, la dépendance du temps de dissociation de l'ADN en fonction de la longueur de la molécule (N) s'exprime comme $t_- = 10^{0.6N-6.3} \text{sec}$ (voir encart Fig 2.6). Une molécule de 10 paires de bases a donc un temps de vie d'environ 10sec , et une molécule de 30 bases est stable pendant $\approx 10^{12} \text{sec} \approx 30000$ ans. Notre estimation de t_- pour $n = 5$, $t_- \approx 0.3\text{ms}$, est en accord avec les résultats obtenus par le groupe de Bonnet et al. [25]. Dans la Fig 2.6 le temps de dissociation en fonction de la force f pour des molécules de longueurs différentes est représenté. La dépendance de la force la plus probable de dégraissage en fonction du taux de charge et de la longueur de la molécule est montrée en Figure 2.7. La même allure a été observée dans des expériences de séparation faites par le groupe de Güntherodt [26] en AFM en tirant sur les bouts opposés des deux brins (extrémités 5'-5') et pas sur les deux bouts (extrémités 3'-5') de la même extrémité comme dans notre cas. Nos résultats théoriques fournissent donc des prédictions variées pour différents types d'expériences de dégraissage en AFM, certaines déjà faites, d'autres tout-à-fait réalisables. Signalons enfin que notre travail représente une extension de la théorie d'Evans de la rupture des liaisons chimiques à la séparation d'objets unidimensionnels. D'autres extensions sont possibles à savoir la considération de chemins de réactions complexes résultant par exemple de potentiels dans un espace multi-dimensionnel [31].

2.4 Dégraissage à force fixée : propagation de l'ouverture et effets de séquence

Dans la publication A5, nous avons étudié la dynamique d'ouverture de séquences spécifiques. Notre travail a été motivé par les expériences de dégraissage de molécules d'ARN faites par le groupe de Bustamante [7]. La première structure qui a été étudiée est l'ARN P5ab montré en figure 2.9. Si on regarde la figure 2.12 des résultats expérimentaux, on s'aperçoit que, quand une force constante de 14.2pN est appliquée aux extrémités de la molécule, cette dernière passe d'un état fermé à un état ouvert avec un temps caractéristique d'environ 1sec .

La première question qui se posait était de savoir pourquoi la molécule montre ce signal télégraphique entre deux états favoris. En fait, ces deux états favoris s'expliquent à l'aide du paysage d'énergie libre, calculé à partir de la formule 2.1 et des énergies d'appariement données par le programme MFold et montré en Figure 2.10. La séquence P5Ab à la force critique, qui selon la théorie est de 15pN , donne lieu à deux minima absolus à la même hauteur, correspondant à la position du buldge (en position 4 à partir du haut de la

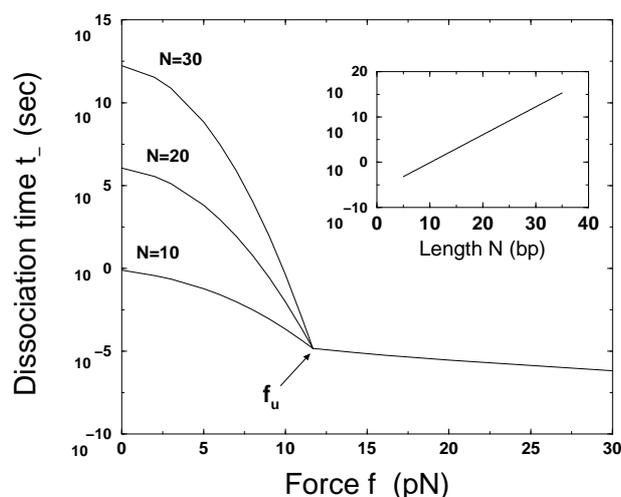


FIG. 2.6 – Temps de dissociation t_- (en sec) en fonction de la force f (en pN). Trois régimes apparaissent. Pour $f < f_u = 12$ pN (pour $g_0 = 1.4k_B T$), les temps de dissociation dépendent de la longueur N de la séquence ($N = 10, 20, 30$ bp de bas en haut). Pour $f_u < f < f_b = 230$ pN, le temps de dissociation est indépendant de la longueur et diminue puisque la barrière à franchir d'abaisse. Pour $f > f_b$, il ne reste plus de barrière et la dissociation est immédiate. Encart : temps de dissociation thermique t_- (à force nulle) en fonction du nombre de paires de bases N .

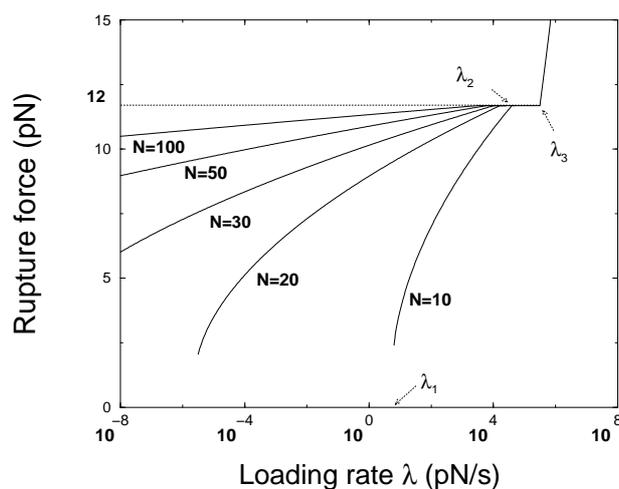


FIG. 2.7 – Force de rupture (pN) en fonction du taux de charge λ (pN/sec) pour cinq longueurs de molécules $N=10, 20, 30, 50$ and 100 . Les flèches indiquent les différents taux de charge pour $N = 10$. En-dessous de λ_1 ($=10^{0.8}$ pour $N = 10$), la rupture survient à force essentiellement nulle par l'intermédiaire d'une dissociation d'origine thermique. Pour des taux de charge allant de λ_1 à λ_2 ($=10^{4.6}$ pour $N = 10$), la force de rupture est finie, et l'activation thermique est à l'origine de la forte dépendance en N , jusqu'à ce que la force atteigne la valeur d'équilibre $f_u = 12$ pN. Pour des taux de charge plus élevés, la force de rupture ne dépend plus de la longueur. Elle augmente à nouveau lorsque $\lambda > \lambda_3 = 10^{5.5}$ pN/sec, puisque la molécule est incapable de répondre à la force avant que celle-ci ne devienne très grande.

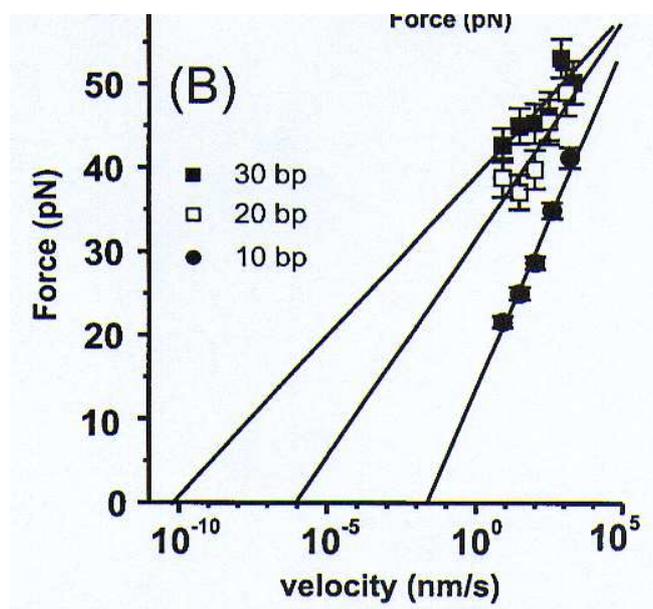


FIG. 2.8 – Force de rupture la plus probable en fonction du taux de charge dans les expériences de séparation faites par le groupe de Güntherodt et al. en AFM en tirant sur les bouts opposés des deux brins (extrémités 5'-5'),

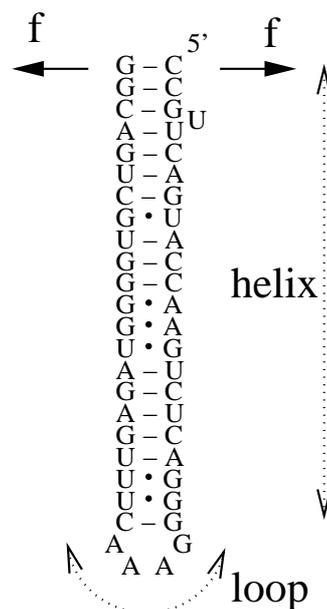


FIG. 2.9 – Molécule P5ab RNA : une seule structure hélice-boucle présente dans le domaine P4-P6 d'un intron auto-épissant (self-splicing) de groupe I du *Tetrahymena thermophila*. La structure montrée est prédite par Mfold à l'exception de l'appariement faible G-A et du buldge U translaté.

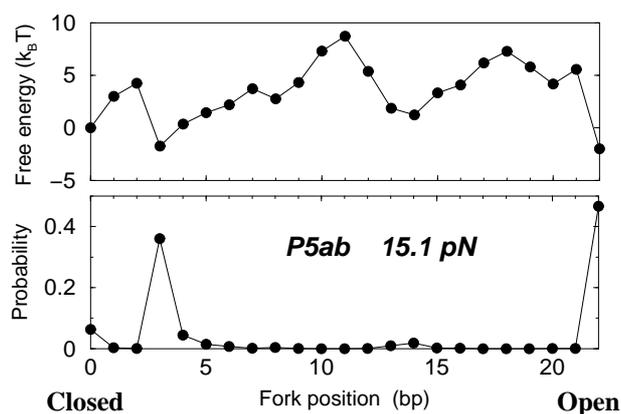


FIG. 2.10 – Paysage d'énergie libre et distribution de probabilité pour la position de la fourche d'ouverture à la force critique (15.1 pN) d'ouverture de la molécule de la Fig. 2.9.

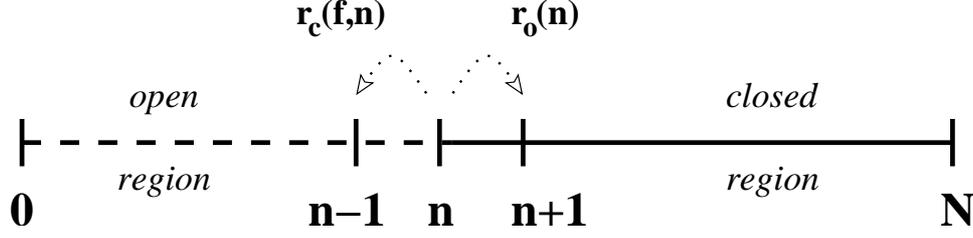


FIG. 2.11 – Modèle dynamique pour l’ouverture et la fermeture d’une seule base à force constante. La frontière entre la portion ouverte et fermée de la molécule se situe à la paire de base numéro n . Les taux élémentaires de fermeture ($n \rightarrow n - 1$) et d’ouverture ($n \rightarrow n + 1$) sont $r_c(f, n)$ et $r_o(n)$ respectivement.

molécule) et à celle de la boucle (position 22) en fin de la molécule.

La deuxième question était de savoir pourquoi l’on observe cette échelle de temps très grande alors que l’objet à ouvrir est de taille microscopique (quelques nanomètres seulement). Cette longue échelle de temps provient du franchissement dynamique d’une barrière élevée $\approx 10k_B T$. Nous avons analysé avec un modèle simple la dynamique d’ouverture, Comme il est montré en Fig. 2.11, le modèle décrit le mouvement de l’interface entre les portions ouverte et fermée de la molécule à travers les taux élémentaires d’ouverture et fermeture de la paire de bases n

$$r_o(n) = r e^{-g_0(n)/k_B T}, \quad r_c(f, n) = r e^{2\mathcal{F}_{ss}(f, n)} \quad (2.5)$$

Notons que selon la notation de la formule 2.1, \mathcal{F}_{ss} est négative et $g_0(n)$ est positive. Le taux d’ouverture est choisi de manière à dépendre seulement de l’énergie d’appariement entre les bases. Ceci est raisonnable du fait de la séparation des échelles d’espace entre le régime où les bases interagissent et celui où elles sont étendues dans la direction de la force : les liaisons hydrogène et d’empilement sont cassées dès que les bases s’éloignent d’une fraction d’Å avant que le travail de la force sur une plus longue distance (allant jusqu’à ≈ 0.7 nm) ne soit appréciable. D’autre part, pour fermer la paire de bases, il faut d’abord travailler contre la force appliquée, et donc le taux de fermeture est choisi de manière à dépendre seulement de la force. Les taux élémentaires 2.5 donnent lieu à équation maîtresse pour la probabilité $\rho_n(t)$ que la fourche soit au site n au temps t qui, dans la limite du temps continu, s’écrit :

$$\frac{d\rho_n(t)}{dt} = - \sum_{m=0}^N T_{n,m} \rho_m(t) \quad (2.6)$$

La matrice $T_{n,m}$, $(N + 1) \times (N + 1)$ est tridiagonale avec, comme éléments non nuls, $T_{m-1,m} = -r_c(f, m)$, $T_{m+1,m} = -r_o(m)$ et $T_{m,m} = r_c(f, m) + r_o(m)$. Nous avons diagonalisé cette matrice dans le cas de P5ab (Fig. 2.9).

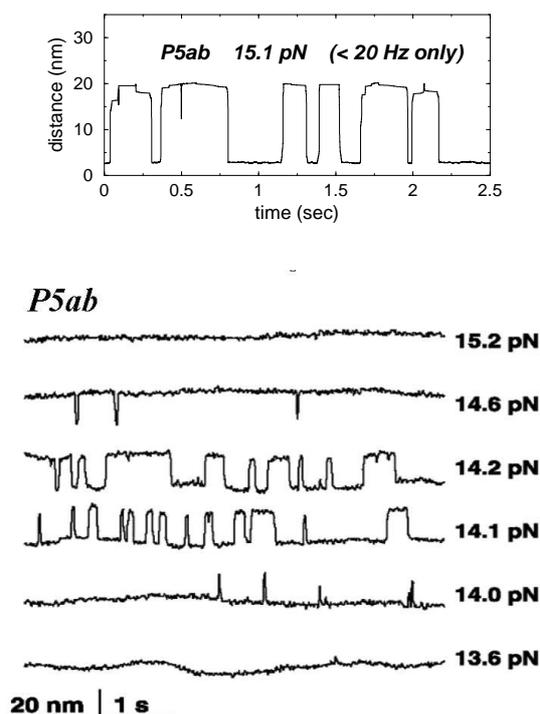


FIG. 2.12 – Cinétique de dégraissage à la force critique : la distance entre les extrémités de la molécule en fonction du temps est montrée pour la molécule P5ab. Haut (théorie) : il y a des basculements lents entre les configurations $n = 3$ et $n = 22$, ainsi que des transitions rapides entre leurs configurations voisines ; pour comparer avec les données expérimentales, nous avons convolué la trace théorique en moyennant les oscillations de fréquence supérieure à 20 Hz. Bas : résultats expérimentaux obtenus par le groupe de Bustamante ; différentes traces sont montrées pour des forces variant entre 13.6 et 15.2 pN.

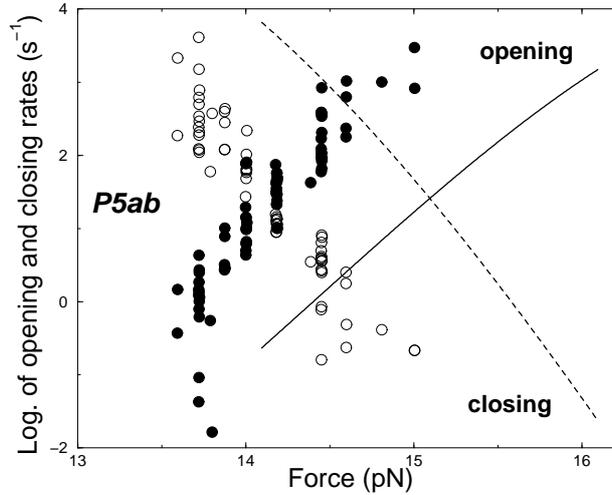


FIG. 2.13 – Logarithme du taux d’ouverture et fermeture pour P5ab en fonction de la force (cercles pleins : ouverture, cercles vides : fermeture) comparé à la théorie (ligne pleine : ouverture, ligne avec tirets : fermeture). La théorie reproduit les données expérimentales quand on translate la force critique à laquelle les taux d’ouverture et fermeture concident de $f^* = 15pN$ à $f^* = 14.3pN$. Les pentes de $\ln k_o, \ln k_c$ ne dépend pas de la valeur de r obtenue à l’aide du fit. Elles donnent la position relative ($n_o = 8$), ($n_c = 11$) de l’état de transition entre les états fermé ($n=3$) et ouvert ($n=24$) avec une position absolue en $n^* = 12$.

La valeur propre la plus petite est nulle ; le vecteur propre correspondant est la distribution d’équilibre de Boltzmann . A la force critique, la valeur propre non nulle la plus petite vaut $\lambda_1 = 2.1 \times 10^{-6}r$. Elle correspond au taux du mode de fluctuation le plus lent, c’est-à-dire le basculement de la frontière de $n \approx 3$ à $n \approx 22$. Les 21 valeurs propres restantes sont bien séparées des deux valeurs propres les plus petites ; ainsi, la deuxième valeur propre (non nulle) la plus petite vaut $\lambda_2 = 0.9 \times 10^{-4}r$ (voir la table de l’article A5). Donc la dynamique théorique de P5ab comporte une transition lente d’ouverture-fermeture, combinée

avec beaucoup d’autres transitions qui sont au moins 50 fois plus rapides.

Les taux nets d’ouverture (k_o) et fermeture (k_c) peuvent être calculés à partir de $\lambda_1 = (k_o + k_c)$ et du rapport des probabilités d’équilibre égal à k_c/k_o . Pour comparer nos résultats théoriques avec les expériences de [7], nous avons fitté nos résultats pour $k^* \equiv k_o = k_c$ à la force f^* où les états ouverts et fermés ont les mêmes probabilités avec les données expérimentales , ce qui permet d’obtenir $r = 3.6 \times 10^6 \text{ sec}^{-1}$. La figure 2.13 montre les résultats pour les logarithmes des taux d’ouverture et fermeture en fonction de la force, comparés aux données expérimentales. Les pentes de ces courbes donne la position de la barrière.

Nous nous sommes par la suite demandé comment dépendent les barrières de la séquence

le long de l'hélice et de la dimension de la boucle finale. Pour une séquence homogène sans boucle à la fin, le mouvement de la fourche d'ouverture à la force critique est essentiellement une diffusion dans un paysage d'énergie libre plat et ne montre donc pas le caractère à deux états de la molécule P5ab. Le temps pour passer de l'état fermé à l'état ouvert est le temps de diffusion de 2×10^{-4} sec. La présence d'une boucle à la fin est suffisante pour générer une grande barrière, comme on le montre avec la séquence homogène Poly(GC) de 24 bases, qui se termine par une boucle de 4 bases. Notre théorie prédit un comportement de basculement entre deux états avec la même échelle de temps que P5ab. Le temps de transition est 50 fois plus élevé pour une boucle de 8 bases. La barrière d'énergie libre à la force critique, G^* , pour une hélice de S paires de bases (avec une énergie libre d'appariement uniforme g_0) suivie par une boucle de L bases (avec une énergie libre de fermeture $g_{loop}(L)$ à force nulle) peut être estimée simplement. La force critique f^* est donnée par la condition que l'énergie libre de la molécule ouverte est égale à l'énergie libre de la molécule fermée $G(0, f^*) = G(S, f^*)$, c'est-à-dire, $0 = S g_0 - (2S + L) g_{ss}(f^*) - g_{loop}(L)$.

La hauteur de la barrière $G^*(S, L) \equiv G(S - 1, f^*)$ est donc

$$G^*(S, L) = (S - 1)(g_0 - 2 g_{ss}(f^*)) = \frac{(S - 1)(g_0 L + 2 g_{loop}(L))}{L + 2S} . \quad (2.7)$$

La Table 1 de l'article A5 montre que, pour une longueur fixée de l'hélice S , la force critique diminue avec le longueur L de la boucle, tandis que la barrière d'énergie libre G^* et le temps de transition t^* augmentent. Pour des séquences petites et non aléatoires, (2.7) peut être utilisée en substituant g_0 par une énergie d'appariement moyennée; elle permet d'estimer comment le temps de transition dépend de S et L . Il est à remarquer que la barrière à la force critique dépend essentiellement de la plus petite des deux longueurs S, L . Nous avons aussi proposé des séquences qui pourraient montrer des dynamiques ouverture-fermeture avec des échelles de temps différentes, ou avec trois ou plus états. Des modèles plus sophistiqués des taux de transition demanderont des expériences ultérieures pour déterminer leur forme. En outre, dans notre modèle, nous avons considéré que la force était constante. Or, si la force est réellement fixée dans les micromanipulations avec des pinces magnétiques, elle ne l'est dans le dispositif avec des pinces optiques utilisées par le groupe de Bustamante que grâce à un mécanisme de rétroaction. La force change avec la distance entre les extrémités de la molécule, parce qu'elle dépend de la position de la bille dans le piège optique. Pour rendre constante la force, il faut donc ajuster la position du piège optique à chaque mouvement. Pour une modélisation précise de cette expérience il faudrait donc inclure dans le modèle dynamique ce mécanisme de rétroaction.

2.5 Dégraissage à vitesse d'éloignement des extrémités constante

En collaboration avec J. Marko, R. Monasson (Publication A6).

Bockelmann, Heslot et leurs collaborateurs ont mesuré la force de dégraissage de l'ADN du λ phage quand on éloigne les deux extrémités de la molécule à vitesse constante dans

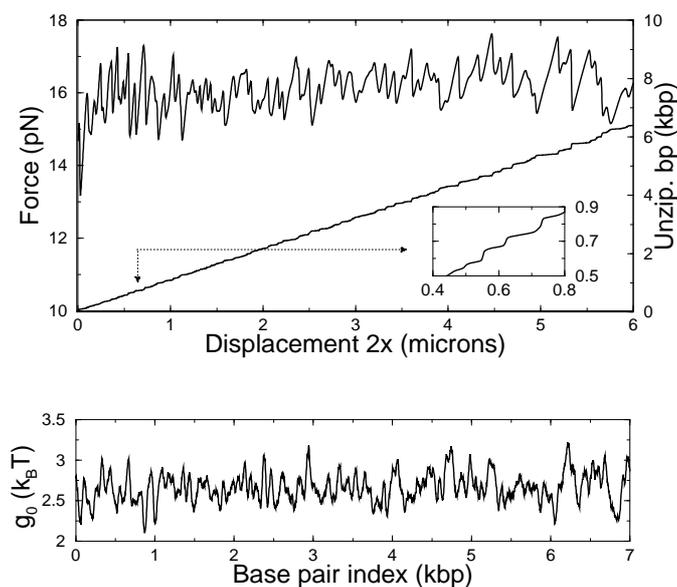


FIG. 2.14 – Haut : Force et nombre de bases dégraftées à l'équilibre pour la molécule de λ DNA, en fonction du déplacement $2x$ après l'initiation de l'ouverture. Le nombre de bases dégraftées augmente linéairement avec x avec des marches caractéristiques d'une dynamique du type stick-slip. Bas : énergie libre d'appariement en fonction du numéro de la paire de bases, moyennée avec un poids gaussien sur 20 bp.

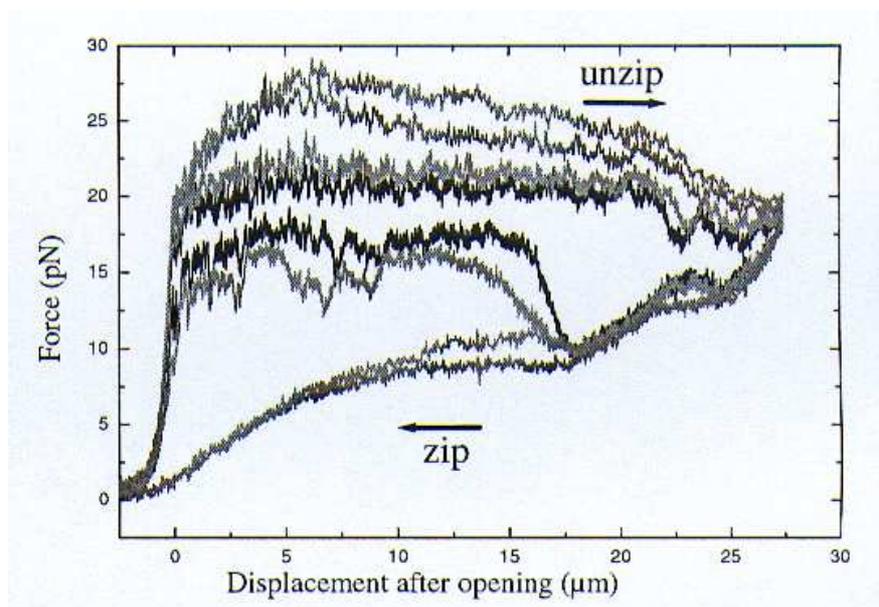


FIG. 2.15 – Force en fonction du déplacement mesurée dans les expériences du groupe de Bockelmann et Heslot pendant le dégraissage et la fermeture de l'ADN λ phage à des vitesses de 4, 8, 16 et 20 $\mu\text{m}/\text{s}$.

la configuration montrée en Fig. 2.2. Si on tire à basse vitesse, on observe une force de dégraissage d'environ 15 pN. Les oscillations de la force reproduisent la séquence des bases fortes ou faibles de l'ADN. Le signal d'ouverture à l'équilibre a été théoriquement analysé par Bockelmann, Essevaz-Roulet, Heslot[33] qui ont calculé numériquement les valeurs moyennes (à l'équilibre thermique) de la force et du nombre de paires de bases ouvertes en considérant l'énergie élastique du piège optique, de l'ADN double brin et de l'ADN simple brin, et l'énergie d'appariement des bases sur la fourche (voir Fig. 2.2). Le signal ainsi obtenu reproduit bien le signal qu'on observe expérimentalement. Il est du type 'stick-slip' comme montré dans la Figure 2.14. Même si le nombre de bases ouvertes augmente en moyenne linéairement avec l'ouverture, l'encart sur une plus petite échelle montre qu'il croît par sauts et plateaux successifs. Dans la phase 'stick', la fourche d'ouverture se bloque sur une région de la séquence riche en GC, en donnant un plateau; durant ce blocage, la force augmente parce qu'on continue à tirer jusqu'à ouvrir les bases riches en GC et un nombre élevé des bases à la suite. Dans la phase 'slip', le nombre de base ouvertes augmente considérablement et la force décroît brutalement.

Avec un dispositif expérimental plus récent [34], Bockelmann et Heslot ont été capables d'observer le contenu en GC de la séquence sur des longues séquences (de l'ordre de 10 kb) avec une résolution d'environ 10 paires de bases. La variance dans le numéro de la base ouverte augmente avec le nombre de bases ouvertes parce que la raideur totale décroît avec l'augmentation de la longueur du simple brin. En outre, le signal de force est sensible à certaines substitutions d'une seule paire de bases [34]. La Fig.2.14 montre la valeur de la

force et du nombre de bases ouvertes que nous obtenons en minimisant l'énergie libre en maintenant l'éloignement entre les extrémités fixé. Nous avons pré-moyenné l'énergie libre de dénaturation avec un poids gaussien correspondant à une déviation standard de 20 paires de bases. La méthode que l'on utilise est approchée par rapport à celle de Bockelmann, Essevaz-Roulet, Heslot mais présente l'avantage d'être plus rapide : à la place d'une simulation Monte Carlo, on calcule numériquement le minimum de l'énergie libre à chaque extension donnée.

Dans un travail successif, Thomen, Bockelmann, Heslot [8] ont observé des effets hors équilibre dans l'ouverture à haute vitesse, $dx/dt > 10\mu$ m/s c'est-à-dire 10 kbases/s. La force de dégraissage à haute vitesse augmente jusqu'à 40% par rapport à celle d'équilibre surtout au début de l'ouverture. Puisque le dégraissage implique la mise en rotation de la portion de la double hélice qui reste à ouvrir (Fig.2.2), on s'attend à ce que le couple de rotation visqueuse produise une contribution à la force de dégraissage [8]. Ayant auparavant étudié la thermodynamique de l'ouverture de la molécule d'ADN sous une force de dégraissage appliquée sur une extrémité de la molécule et un couple de torsion sur l'autre extrémité, nous nous sommes particulièrement intéressés à la modélisation de cette expérience. Nous avons d'abord analysé le cas d'une séquence uniforme. Nous avons négligé les fluctuations d'extension du piège optique, parce que la raideur de ce dernier est supérieure d'un facteur dix à celle du double brin et est supérieure à celle du simple brin dès que l'on a ouvert quelques centaines de bases [34] ; un modèle qui tient aussi compte des fluctuations d'extension du piège optique a été récemment introduit par Mañosas et Ritort [35] pour étudier la thermodynamique et la cinétique du dégraissage de l'ARN.

L'énergie libre à déplacement x fixé (voir Fig 2.2) s'écrit en fonction du nombre de bases ouvertes n et de l'extension du double brin x_{ds} :

$$F_x(n, x_{ds}) = 2 W_{ds}(x_{ds}) + 2W_{ss}(x - x_{ds}, n) + n [g_0 + \Gamma\theta_0] \quad (2.8)$$

ou $W_{ds}(x_{ds})$ et $W_{ss}(x_{ds})$ sont les travaux effectués par la force pendant l'extension des simple et double brins. Ils sont liés aux énergies libres à force fixée, $\mathcal{F}_{ds}(f)$ et $\mathcal{F}_{ss}(f)$ que l'on a dans la modélisation de l'ouverture à force fixée (équation 2.1) par une intégration par partie. La minimisation par rapport à x_{ds} donne simplement la condition que la force qui agit sur le double brin doit être la même que celle qui agit sur le simple brin. La valeur de la force en fonction du nombre de bases ouvertes se calcule par la condition $x_{ds}(f) + x_{ss}(f) = x$. La minimisation par rapport à n de $F_x(n, x_{ds}(f(n)))$ donne la valeur d'équilibre du nombre de bases ouvertes.

Pour étudier les effets hors équilibre, nous avons décrit la dynamique de relaxation visqueuse de l'ouverture dans la limite du continuum pour n :

$$t_n \frac{dn}{dt} = - \frac{dF_x(n, x_{ds}(f(n)))}{dn} \quad (2.9)$$

qui donne l'équation

$$t_n \frac{dn}{dt} = \frac{1}{k_B T} [2w_{ss}(f) - g_0(n) - \Gamma\theta_0] \quad (2.10)$$

où $w_{ss}(f) = \int_0^f l_{ss}(f') df'$ est l'énergie libre par paire de base à force fixée. $l_{ss}(f')$ est l'extension dans la direction de la force d'une paire de base ; comme dans la modélisation du dégraissage à force constante nous avons utilisée l'expression (voir formule 4 de la publication A6) de la FJCL donné par le groupe de Bustamante [3]. t_n est de l'ordre du temps de diffusion d'une base de dimension d'un nanomètre : $t_n \simeq 10^{-8}$ sec.

Dans l'équation 2.10, le couple Γ peut être en particulier dû à la rotation de la double hélice à ouvrir. En effet celle-ci tourne à chaque paire de base qui s'ouvre d'un angle $\theta_0 = 2\pi/10$. Si on suppose qu'elle a rejoint un état d'équilibre stationnaire dans lequel elle tourne à vitesse angulaire uniforme $\omega = \theta_0 dn/dt$, on peut écrire le couple de torsion visqueux pour un cylindre de rayon r et longueur $\Delta(N - n)$ comme

$$\Gamma(n) = 4\pi\eta r^2 \Delta(N - n)\omega. \quad (2.11)$$

Nous avons considéré un rayon hydrodynamique $r = 2nm$ pour l'ADN, $\Delta = 0.34nm$ et η est la viscosité de l'eau. Le facteur $N - n$ indique que la torsion visqueuse à haute vitesse de dégraissage est importante au debut de l'ouverture et décroît au fur et à mesure que l'ouverture progresse, donc la force de dégraissage augmente au debut de l'ouverture pour décroître vers la valeur d'équilibre au cours du dégraissage, comme l'on observe expérimentalement en Fig. 2.15. En inserant ω dans l'équation 2.11 et $\Gamma(n)$ dans l'équation 2.10, nous écrivons l'équation de propagation de la fourche d'ouverture :

$$\frac{dn}{dt} = \frac{1}{k_B T} \frac{[2w_{ss}(f) - g_0(n)]}{t_n + (N - n)t_\Gamma} \quad (2.12)$$

que l'on a résolu par une intégration numérique. Note que $t_\Gamma = 4\pi\eta r^2 \Delta \theta_0^2 / (k_B T) = 2 \times 10^{-9}s$ est comparable à la valeur attendue pour t_n mais il est multiplié au début de l'ouverture par un facteur $N = 5 \cdot 10^4$, la dynamique rotationnelle limite donc la vitesse d'équilibration dans le dégraissage de longues molécules d'ADN. Dans les expériences de la Fig. 2.15, les extrémités de la molécule d'ADN sont éloignées jusqu'à une séparation de $25nm$, qui correspond environ à 25 kbp ouvertes, et elles sont par la suite rapproché lors de la fermeture à la même vitesse. Dans la fermeture, le signe de la vitesse dx/dt doit être renversé de même que ceux de la vitesse angulaire ω et de la torsion visqueuse Γ . Si on soumet l'extrémité inférieure de la molécule à un couple de torsion négatif, qui déroule la double hélice, cela facilite l'ouverture comme il est montré dans le diagramme de phase de la Fig.2.3. La force de dégraissage diminue sous l'effet du couple de torsion visqueuse lors de la fermeture à haute vitesse. La diminution de la force de dégraissage augmente avec la fermeture parce que la torsion visqueuse est proportionnelle à la longueur de la double hélice en aval de la fourche d'ouverture.

La force de dégraissage et le couple de torsion que l'on obtient en résolvant l'équation (2.12) pour une séquence uniforme $g_0(n) = 2.5k_B T$ sont montrées en Fig. 2.16, 2.17 pour des vitesses de dégraissage allant de 4 à 20 $\mu m/s$. On observe l'augmentation initiale de la force de dégraissage et du couple de torsion à grande vitesse et leur diminution progressive pendant le dégraissage. Pendant la fermeture, la force et le couple décroissent et sont inférieurs à leurs valeurs à l'équilibre. La force et le couple sont reliés par le diagramme de phase de Fig.2.3, comme les points A, B, C, D sur les figures 2.16, 2.17, 2.3 le montrent.

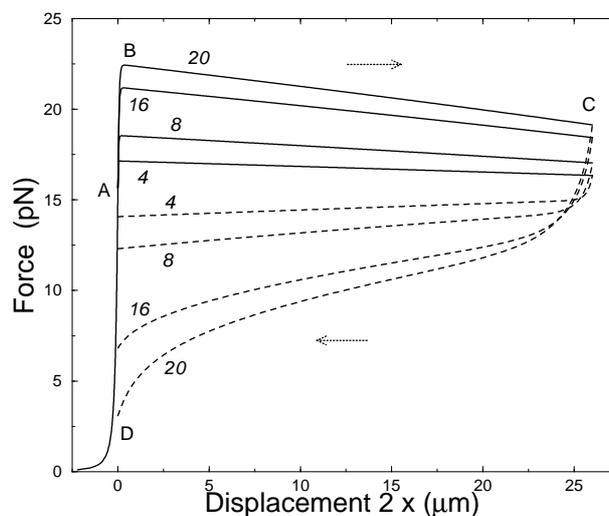


FIG. 2.16 – Force en fonction du déplacement pour un ADN homogène (avec une énergie libre d'appariement $g_0 = 2.5k_B T$) de 50 kb et des bras de 7 kb. Les courbes pleines correspondent au dégraissage et les courbes en tirets à la fermeture à vitesse $2v = 4, 8, 16$ et $20 \mu\text{m/s}$.

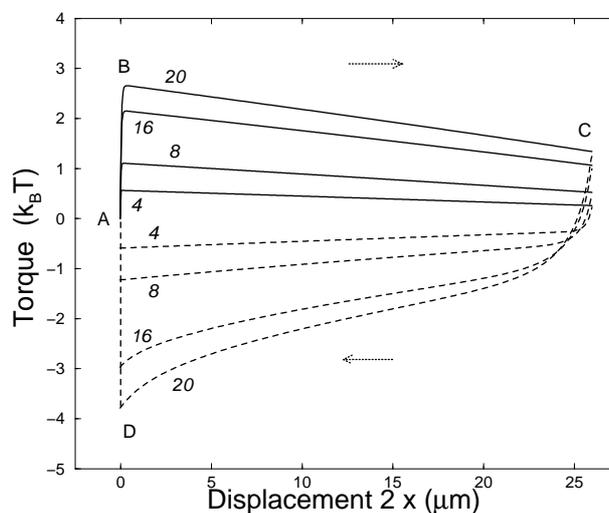


FIG. 2.17 – Couple de torsion en fonction du déplacement pour un ADN homogène (avec une énergie libre d'appariement $g_0 = 2.5k_B T$) de 50 kb et des bras de 7 kbp. Les courbes pleines correspondent au dégraissage et les courbes en tirets à la fermeture à vitesse $2v = 4, 8, 16$ et $20 \mu\text{m/s}$.

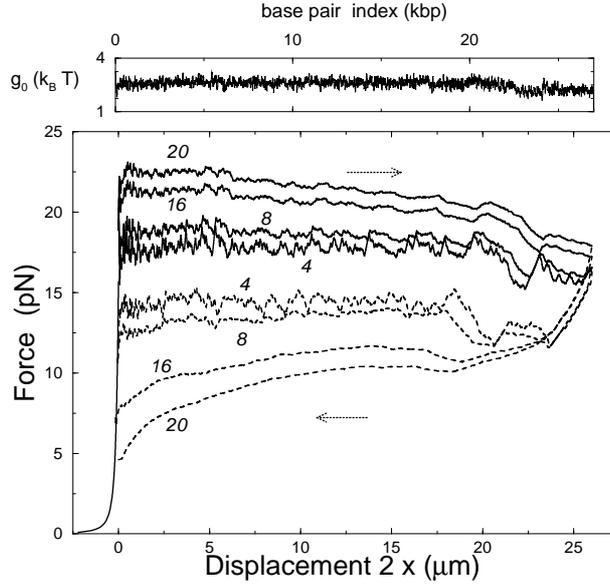


FIG. 2.18 – Force en fonction du déplacement pour un ADN de λ phage (avec une énergie libre d'appariement $g_0 = 2.5k_B T$) de 50 kb et des bras de 7 kbp. Les courbes pleines correspondent au dégraissage et les courbes en tirets à la fermeture à vitesse $2v = 4, 8, 16$ et $20 \mu\text{m/s}$.

Ces mêmes caractéristiques apparaissent dans la figure expérimentale 2.15. Dans le cas du λ phage se superposent aux courbes de la fig. 2.16 les fluctuations de la force de dégraissage dues à la séquence, comme montré en Fig. 2.18. On remarque en particulier la brusque diminution en contenu de GC sur la séquence qui a lieu après 20 kb. On remarque aussi que les fluctuations diminuent avec le simple brin dégraissé parce que, comme mentionné au début du paragraphe, la raideur totale de la partie en amont de la fourche d'ouverture décroît lorsque augmente la longueur du simple brin. Pour finir, les fluctuations diminuent avec la vitesse de dégraissage parce que les détails du paysage d'énergie libre, qui dépendent de la séquence, sont de moins en moins importants.

Jusqu'ici, nous avons supposé que le ssDNA, le dsDNA et la torsion de la molécule s'équilibrent plus rapidement que la propagation de l'ouverture. Cette hypothèse est vérifiée pour la dynamique d'étirement du ssDNA et du dsDNA. Si on calcule le plus grand temps de relaxation selon le modèle de Zimm en considérant des élasticités par monomère de $k_{ds} = 1400\text{pN/nm}$ et $k_{ss} = 140\text{pN/nm}$ (obtenues en linéarisant le courbe de force-extension autour de 15 pN), et une élasticité torsionnelle par monomère $k_{tw} = k_B T C / \Delta^2$ où $C = 80 \pm 20\text{nm}$ est la rigidité élastique de torsion, on obtient (voir équations (9), (10) et (11) de la Publication A6) $t_{ds} = 3 \times 10^{-4}\text{s}$, $t_{ss} = (2n)^2 1.4 \times 10^{-11}\text{s}$, $t_{tw} = (N - n)^2 2 \times 10^{-12}\text{s}$. Si on compare ces temps avec le temps de dégraissage, $t_u \simeq n/v$, on a pour $v = 20\mu\text{m/s}$: $t_{ds}/t_u \simeq 100/n$ donc le double brin est à l'équilibre dès que $n > 100$; $t_{ss}/t_u \simeq n/10^6$ donc le simple brin est à l'équilibre si $n < 10^6$, condition qui est vérifiée pendant tout le dégraissage de

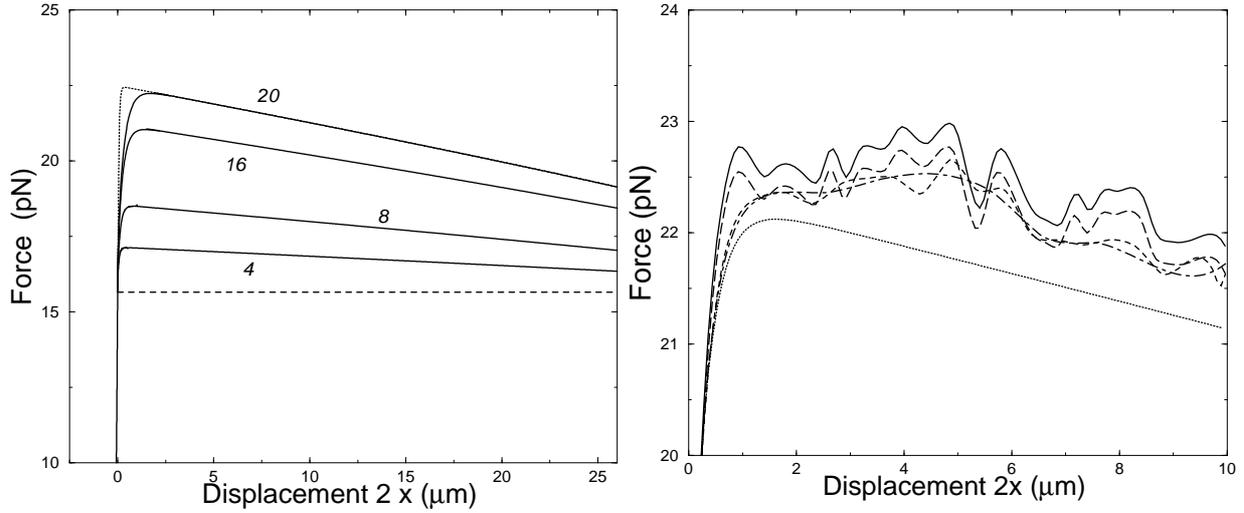


FIG. 2.19 – Force en fonction du déplacement obtenue en considérant aussi la propagation de la torsion. Haut : résultats du dégraissage d’une séquence homogène à vitesse $2v = 4, 8, 16, 20 \mu\text{m/s}$. Bas : résultats pour une séquence de λ phage moyennée sur des fenêtres de longueur 1000 (courbe en traits mixtes), 500 (courbe tiretée), 250 (courbe avec de longs tirets), 100 (courbe pleine) pour $2v = 20 \mu\text{m/s}$, comparés au cas homogène (courbe en pointillés).

la séquence du λ phage qui comporte $5 \cdot 10^4$ bases ; $t_{tw}/t_u = (N-n)^2 2 \cdot 10^{-12}/(n 5 \cdot 10^{-5})$, donc au début du dégraissage ($t_{tw} = 5 \cdot 10^{-3} \text{s}$), la double hélice n’a pas encore atteint sa vitesse de rotation stationnaire. Le couple visqueux initial est plus petit de ce qu’on obtient en considérant toute la double hélice en rotation, ce qui explique la montée plus graduelle de la force dans la figure 2.15 que dans les figures 2.16, 2.18. Nous avons donc considéré explicitement dans la modélisation l’équation pour la dynamique de relaxation torsionnelle (voir équation 23 dans la publication A6) qui équivaut à l’équation de propagation de la chaleur avec des conditions aux bords particulières pour l’angle de rotation $\theta(m, t)$ ($m = 0, \dots, N$ est l’indice de la paire de base). Initialement, $\theta(m, 0) = 0$ parce que la double hélice est fermée. Au bout de la double hélice, la rotation est libre : $d\theta/dm(N, t) = 0$ et l’angle de twist sur la fourche d’ouverture $n(t)$ correspond à l’angle total qui a été déroulé $\theta(n(t), t) = -\theta_0 n(t)$. La résolution de la dynamique couplant ouverture et propagation de la torsion conduit en effet à une augmentation initiale de la force plus graduelle, voir Fig 2.5a. Cet effet est ultérieurement accentué en Fig 2.5b quand l’on considère le λ phage qui est caractérisée par un gradient positif du contenu en bases GC au début de la séquence. On arrive ainsi à un pic de force autour de $4 \mu\text{m}$ (Fig 2.5) plutôt en bon accord avec le pic autour de $6 \mu\text{m}$ observé expérimentalement (Fig 2.15). Il faut pour finir remarquer que nous n’avons pas considéré certains effets comme la dynamique de translation de la fourche d’ouverture. Cette dernière, dans le dispositif expérimental où une extrémité de l’ADN est attachée à une paroi et l’autre mobile, s’éloigne par rapport à la paroi ; nous n’avons pas

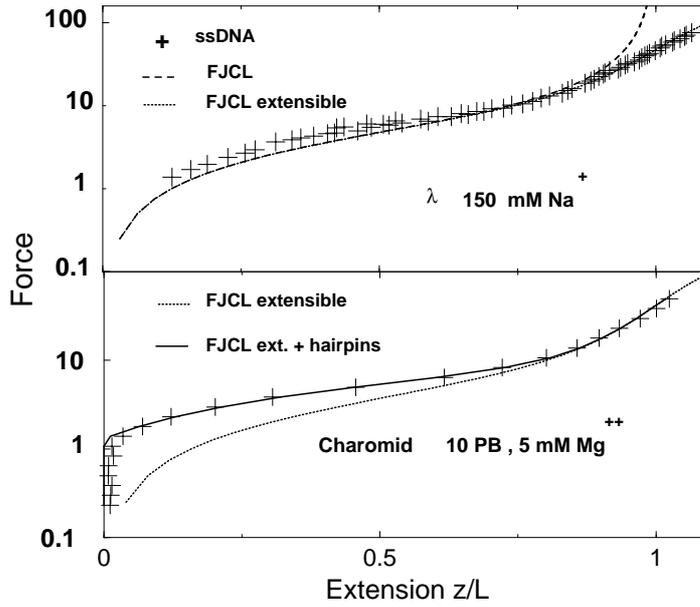


FIG. 2.20 – Haut : courbe de force-extension pour un λ -ssDNA avec une longueur $L_{ss} = l_{ss}N \equiv 0.56 \text{ nm} \times 48502 \text{ bp} = 27 \mu\text{m}$ en 150 mM Na^+ . Données expérimentales : symboles +, courbe en tirets : FJCL avec $b = 1.5 \text{ nm}$, courbe pointillée : FJCL extensible avec $\gamma = 800 \text{ pN}$, à partir de [3]. Bas : courbe de force-extension pour un Charomid ssDNA dans un buffer de phosphate 10 nM , 5 mM Mg^{++} . Les données sont extraites de [6], les fits sont réalisés avec le modèle FJCL avec $b = 1.9 \text{ nm}$ et $\gamma = 800 \text{ pN}$ (courbe pointillée) et avec le modèle à hairpins (courbe pleine) de [27].

considéré non plus le repliement de la double hélice à ouvrir et la possibilité de formation de plectonèmes (enroulements similaires à ceux de la corde du téléphone) dans cette partie.

2.6 Elasticité de molécules avec des interactions entre les monomères

En collaboration avec J. Marko, R. Monasson, A. Sarkar (UIC, Chicago), J. Yan (UIC Chicago) (Publication A7).

L'ADN simple brin (ssDNA) est très flexible et peut atteindre une extension supérieure celle de l'ADN double brin. A première vue, un modèle approprié pour l'ADN simple brin est le modèle de la chaîne à articulations libres (FJC) avec une longueur de Kuhn égale à la longueur du squelette phosphodiester $b = 7 \text{ \AA}$. Par suite de cette grande extensibilité, la force entropique caractéristique, $f_0 \simeq 3k_B T/b = 17 \text{ pN}$, est beaucoup plus grande que celle de l'ADN double brin, $f_0 \simeq 0.1 \text{ pN}$. Cependant, l'élasticité de l'ADN simple brin est compliquée par les interactions entre nucléotides et, par conséquent, un modèle simple de polymère ne décrit pas l'élasticité du ssDNA sur une large gamme de forces et de concentrations ioniques. Nous avons auparavant décrit l'élasticité d'un simple brin d'ADN

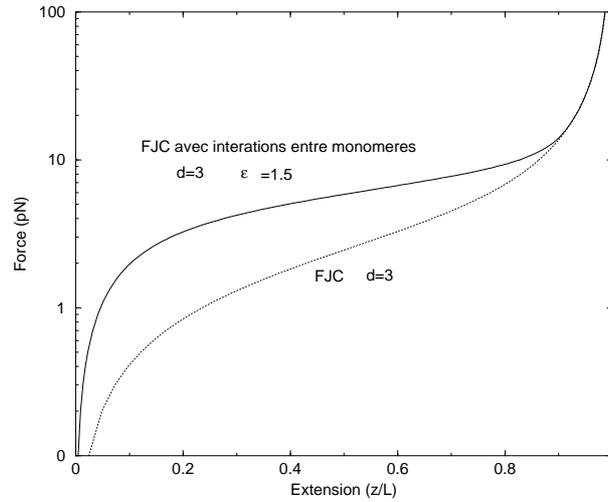


FIG. 2.21 – Courbe force-extension obtenue avec le modèle FJC incluant des interactions entre monomères (courbe pleine) ayant pour paramètres $\epsilon = 1.5k_B T$, $b = 3nm$, comparé à la courbe du modèle FJC simple (courbe en tirets).

à concentration ionique de 250 mM avec un modèle FJC modifié (FJCL), introduit par le groupe de Bustamante [3], possédant trois paramètres effectifs : la longueur de Kuhn, $d = 15 \text{ \AA}$, la longueur par paire de base $l_{ss} = 5.6 \text{ \AA}$, ainsi qu'un module de Young pour décrire l'élasticité à haute force. L'élasticité de l'ADN simple brin dépend fortement de la concentration de sel : à basse concentration ($\approx 1 \text{ mM Na}^+$), des effets d'auto-réplulsion dus à l'interaction électrostatique entre le squelette chargé sont présents ; à haute concentration ($> 100 \text{ mM Na}^+$ ou en présence de Mg^{++}), des liaisons hydrogènes entre nucléotides sont possibles. Les expériences montrent que l'élasticité de l'ADN simple brin dépend du contenu en GC vs. AT et peut être modifiée en ajoutant des agents dénaturants [6]. Des simulations Monte-Carlo qui introduisent la répulsion électrostatique et l'interaction entre monomères [6] reproduisent bien le comportement expérimental. Une analyse théorique de l'élasticité d'un polymère avec structure secondaire en 'hairpin' (épingle à cheveux) avec des appariements non spécifiques entre bases a été développée par Montanari et Mézard [27] dans le cadre de l'approximation de Hartree. Ce modèle n'inclut pas les pseudo-noeuds mais Orland et Zee ont montré comment le résoudre aussi en tenant compte de ces interactions [28]. En Fig.2.20, la courbe force-extension d'un ssDNA en présence de Magnésium est bien fitté par le modèle avec hairpin. En particulier, on observe que pour une longue molécule une transition structurale a lieu entre un état replié (à extension zéro, $f < f_s$) et un état étendu ($f > f_s$) avec $f_s \approx 1 \text{ pN}$. Le travail de la Publication A7 a été d'abord motivé par la description de l'élasticité de l'ADN simple brin, avec une modification simple du modèle de la chaîne à articulations libres pour prendre en compte les interactions de repliement de la chaîne. En effet pour comprendre le comportement élastique du simple brin à haute concentration ionique il suffit de considérer un modèle FJC où chaque unité de Kuhn peut être dans un état replié (caractérisé par une variable

d'Ising $n = 0$) à extension nulle et énergie de répliement 0, ou dans un état étendu ($n = 1$) avec énergie ϵ en unité de $k_B T$ et extension d . ϵ est la différence d'énergie libre entre l'état étendu et l'état replié. La fonction de partition pour un segment s'écrit

$$Z_1 = \frac{1}{2} \int_{-1}^1 d(\cos\theta) \sum_{n=0,1} e^{(-\epsilon + \Delta \cos\theta)n} \quad (2.13)$$

où $\Delta = f d / (k_B T)$. L'énergie libre par segment est donc comme

$$\ln Z_1 = \ln \left(1 + e^{-\epsilon} \frac{\sinh \Delta}{\Delta} \right) \quad (2.14)$$

L'extension est aussi facilement calculée,

$$\frac{z}{L} = \frac{dZ_1}{d\Delta} = \frac{\cosh \Delta - \Delta^{-1} \sinh \Delta}{\Delta e^\epsilon + \sinh \Delta} \quad (2.15)$$

Dans la limite $\epsilon \rightarrow -\infty$ l'état replié n'a jamais lieu et on retrouve le résultat habituel du FJC. Dans la Figure 2.21, j'ai représenté la force en fonction de l'extension pour le choix de paramètres $\epsilon = 1.5k_B T$, $b = 3$ nm, et comparé avec la courbe force–extension du simple FJC avec $b = 3$ nm. On observe la transition de l'état replié à l'état étendu à une force $f_0 \approx \epsilon/b = 2$ pN. Si on développe l'équation (2.15) à faible force, on obtient $Nz = Nd_{eff}^2 f / 3 + O(f^2)$, c'est-à-dire le comportement standard d'un polymère gaussien avec une longueur apparente du segment $d_{eff} = d / (1 + e^\epsilon)$, qui est la longueur de Kuhn qu'on fitterait à partir des données avec un simple modèle FJC. Cette longueur de Kuhn effective peut être donc incohérente et notamment plus petite que la longueur réelle du segment. En outre elle peut ne pas correspondre avec celle que l'on obtient à haute force, limite dans laquelle on réobtient en effet le comportement élastique du FJC avec une longueur de Kuhn d parce que l'état replié ne survit pas. Dans le travail A7, nous avons commencé par introduire un modèle basé sur un polymère du type FJC mais plus complexe de celui que je viens de décrire. Nous considérons que chaque segment statistique de Kuhn est composé par M unités qui peuvent se replier et qui ont une extension d_1 dans l'état étendu et d_0 dans l'état replié. La différence d'énergie libre entre l'état étendu et l'état replié est toujours représentée par ϵ . Nous prenons en compte aussi une élasticité de Young du segment en introduisant un changement additionnel de longueur y , avec une constante élastique f_0 . Nous avons considéré deux interactions possibles entre les monomères qui constituent le segment. Dans le modèle 'zippered', les monomères d'un segment statistique forment des hairpins, l'ouverture d'une unité suit donc celle de l'unité précédente (et la fermeture suit celle de l'unité successive); dans le modèle 'indépendant', le repliement est local et chaque unité peut s'ouvrir et se fermer de façon indépendante. La fonction de partition pour les deux modèles est donnée en formule (3) de la publication A7 et la longueur effective qu'on obtient dans le développement à basse force en équation (6). Pour reproduire l'élasticité de l'ADN simple brin à haute concentration ionique, nous avons utilisé le modèle 'zippered' avec $\epsilon = 0.2$, $d_0 = 0$, $M = 3$, $d_1 = 0.7$ nm, $f_0 = 220$ pN (voir Fig. 3 de la publication A7). Les deux modèles s'adaptent pour reproduire différents résultats expérimentaux. En effet

l'ADN dans la cellule est empaqueté à plusieurs niveaux : l'ADN double brin s'enroule autour des histones, les histones se compactent dans la chromatine, la chromatine est empaquetée dans les chromosomes. Nous avons à l'heure actuelle à disposition plusieurs données d'élasticité aux différents niveaux de cet empaquetement. Comme montré en Fig. 2 de la publication A7, l'élasticité des fibres de chromatine est bien reproduite par le modèle indépendant avec paramètres $\epsilon = 4$, $d_0 = 5$ nm, $M = 20$, $d_1 = 10$ nm, $f_0 = 50$ pN. La figure 4 de la publication A7 montre en outre qu'on peut bien reproduire l'élasticité d'un chromosome mitotique entier avec le modèle avec $M=1$, $\epsilon = 2$, $d_0 = 0$, $d_1 = 10$ nm, $f_0 = 50$ pN. Jusqu'à maintenant nous avons considéré le cas d'interactions entre monomères qui ont lieu sur une échelle plus petite que la longueur de Kuhn du polymère. On peut aussi considérer la limite opposée d'interactions entre monomères d'ADN qui forment des grandes boucles, cela a lieu typiquement quand les boucles sont formées grâce aux interactions avec des protéines qui rapprochent des parties éloignées d'ADN. Pour modéliser ce cas, nous sommes partis du modèle du ver (WLC) comme polymère de base et avons resommé la fonction de partition sur les nombres d'unités repliées (chacune réduit la longueur d'un facteur l), voir formule (11) de la publication A7. Nous avons considéré deux cas : les unités qui se replient peuvent être fixées dans le cas d'interactions spécifiques avec les protéines ou glissantes pour des interaction non spécifiques entre l'ADN et les protéines. Le modèle glissant reproduit en particulier les données de force-extension de l'ADN en présence de protéines IHF, voir figure 5 de la publication A7. Pour finir, nous avons montré comment on peut introduire les effets de la torsion à partir de l'expression approchée de l'énergie libre obtenue par Moroz et Nelson [29] pour le modèle WLC avec énergie de torsion. On peut aussi introduire les effets électrostatiques à partir du calcul de Barrat et Joanny [30] de l'énergie libre de la WLC avec un potentiel de Debye- Huckel.

2.7 Thermodynamique et dynamique de surétirement de l'ADN

En collaboration avec D. Chatenay, J.F. Léger (LDFC Strasbourg) et J. Marko (Publication A8).

Les expériences d'élasticité sur une molécule d'ADN ont montré que la molécule subit une transition structurale lorsqu'elle est soumise à une force d'étirement d'environ 60 pN. Cette nouvelle structure (SDNA) a suscité des controverses. Bloomfield et collaborateurs prétendent notamment qu'il ne s'agit pas d'une nouvelle structure mais de deux brins d'ADN n'interagissant pas entre eux (2 ssDNA). Un premier problème avec cette hypothèse est qu'elle est en contradiction avec les résultats expérimentaux du groupe de Chatenay [11], qui ont suggéré une structure bien régulière pour la forme SDNA avec une hélicité bien déterminée de 35 bases par tour d'hélice. Une interprétation simple des courbes force-extension du double brin et du simple brin mesurées expérimentalement est aussi en contradiction avec l'hypothèse d'un ADN dénaturé après la transition. La transition dsDNA/SDNA se termine, pour le λ phage, avec une extension d'environ 0.58 nm/bp

et une force de 68 pN qui concident avec l'extension d'un ssDNA à la même force (voir la Figure 3 de la publication A8). Ceci constituait un argument pour suggérer que la forme S se confondait en réalité avec de l'ADN dénaturé. Si l'on suppose que la structure après la transition est de le ssDNA, les deux filaments parallèles devraient avoir une réponse à la force $x_{2ss}(f) = x_{ss}(f/2)$. En d'autres mots, une force double de celle nécessaire pour un simple brin doit être appliquée pour étirer à la même longueur deux simples brins. D'autre part, l'assimilation de la forme S à de l'ADN simple brin pose également problème. Comme on peut le voir dans la Figure 3 de la Publication A8, les pentes des deux courbes à la fin de la transition ne sont pas les mêmes. La constante élastique du ssDNA (280/0.34 pN/nm) est environ un sixième de celle du SDNA. Ainsi, la constante élastique de deux brins du ssDNA est environ un tiers de celle du SDNA. Enfin, on pourrait émettre l'hypothèse que la phase S est en fait un mélange du ssDNA et du dsDNA mais, dans ce cas, la constante élastique serait comprise entre la constante de le dsDNA et celle de le ssDNA, et serait plus petite que celle du SDNA. Donc, sans aucune analyse thermodynamique ou théorique mais simplement à partir de l'analyse des données expérimentales sur l'ADN de λ phage en conditions standards (pH7, 25 C et 150 mM NaCl), on peut déduire que le SDNA a des propriétés mécaniques différentes du ssDNA (un ou deux brins) ou d'un mélange de ssDNA et dsDNA. Dans le travail A8, nous avons formulé un modèle thermodynamique simple de la séparation des brins résultant de la tension pour démontrer qu'il ne peut pas s'agir de la conversion de la double hélice d'ADN en deux brins séparés. Il peut cependant arriver, pour certaines séquences et conditions expérimentales (pH, température, conditions ioniques), qu'un brin s'ouvre à partir d'une cassure du squelette phosphodiester (nick) souvent présente sur la molécule.

L'analyse thermodynamique avait comme but de comparer les énergies libres des dsDNA, SDNA, ssDNA et 2 ssDNA avec différentes conditions ioniques, température et séquences. Cette analyse a débuté par la définition des formules pour la courbe force-extension de ces différents ADN, $x(f)$. L'énergie libre a été dérivée comme l'intégrale de cette quantité. Pour le dsDNA (aux forces suffisamment élevées auxquelles nous sommes intéressés), nous avons utilisé l'expression d'Odjkin, formule 1 dans la référence A8, qui tient compte de la première correction en puissance de $1/f$ dans le développement autour de la force infinie due aux fluctuations de courbure et de la constante élastique d'étirement. Pour le ssDNA, nous avons considéré un modèle effectif (formule 20, A8) qui tient compte de la dépendance en la concentration ionique et du comportement à basse force dû à la formation de contacts entre monomères. Pour le SDNA, nous avons interpolé les données expérimentales (formule 7) en approximant l'extension du SDNA en réponse à la force sous forme linéaire. A l'énergie libre due à la tension, $w(f)$, doit s'ajouter dans le cas de la cassure des liaisons hydrogènes l'énergie d'appariement $g_0(n)$ qui en général dépend de la séquence, de la température et des conditions ioniques. La valeur de g_0 est prise à partir des énergies libres tabulées par Santa Lucia, cf. serveur Mfold. La différence d'énergie par paire de base entre le ssDNA et le dsDNA, si on considère un ADN homogène est $\Delta G_{unpeeling} = w_{ss}(f) - g_0 - w_B(f)$, celle entre 2 ssDNA et le dsDNA est $\Delta G_{unpairing} = 2w_{ss}(f/2) - g_0 - w_B(f)$ et celle entre le SDNA et le dsDNA est $\Delta G_{BS} = w_S(f) - w_B(f)$. Le diagramme de phase dans le plan concentration ionique-force pour une séquence moyenne est montré en Figure 2.22. On y voit que, à

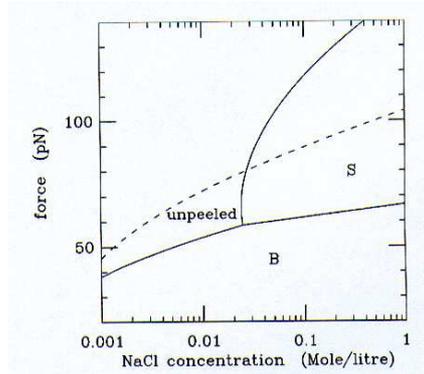


FIG. 2.22 – Diagramme de phase des formes dsDNA, SDNA, ssDNA et 2 ssDNA dans le plan concentration ionique-force, pour un modèle d'ADN avec une séquence d'énergie d'appariement moyenne $g_0 = 2.5 + 0.2 \ln(M/0.15) k_B T$. Pour une concentration élevée de sel, le dsDNA se transforme d'abord en SDNA puis en ssDNA lorsque la force augmente. Pour de basses concentrations de sel, l'ADN se transforme directement en ssDNA. La formation d'ADN à deux filaments parallèles à partir du dsDNA a lieu à des forces plus élevées et n'est jamais favorable par rapport au ssDNA.

des concentrations ioniques plus élevées que 30 mM/litre, il y a d'abord une transition B-S suivie par une transition S-ss (dénotée par unpeeled dans la figure). A des concentrations ioniques inférieures, seule cette dernière transition subsiste. La forme 2ssDNA n'est jamais favorable par rapport au ssDNA et donc peut se rencontrer seulement dans le cas d'une molécule qui n'a pas de nick et donc où un simple brin ne peut pas s'ouvrir.

Dans une deuxième partie du travail, nous avons étudié la dynamique d'étirement pour pouvoir effectuer des comparaisons avec les courbes force-extension mesurées expérimentalement. A force suffisamment élevée et pour une molécule de λ , une deuxième transition est effectivement observable. Elle a été étudiée en détail à l'aide des expériences du groupe de Rief qui consistait à étirer l'ADN avec un microscope à force atomique (voir Fig. 2.24) qui permet d'atteindre des forces de plusieurs centaines de pN. A la différence de la transition B-S qui ne présente aucune dépendance en la vitesse d'étirement sur un régime $100 \text{ nm/s} < v < 1500 \text{ nm/s}$, la deuxième transition présente cette dépendance.

Cela s'explique si l'on considère une transition à la forme ssDNA parce que lors de la transition SDNA/ssDNA, l'ADN doit s'ouvrir. Comme l'énergie d'ouverture dépend de la séquence, dans certaines zones la transition est plus facile que dans d'autres riches en GC. Le paysage d'énergie libre de détachement d'un brin (unpeeling) possède donc des barrières et la dynamique d'ouverture dépend donc de la position du nick. Pour reproduire cette dépendance de la vitesse et les données expérimentales de Rief et al., nous avons considéré un modèle de dynamique équivalente au modèle utilisé pour le dégraissage. La courbe force-extension issue du modèle dynamique (Figure 2.23), et en particulier sa dépendance en la vitesse d'extension après le plateau de la transition à la forme S, reproduit bien les données

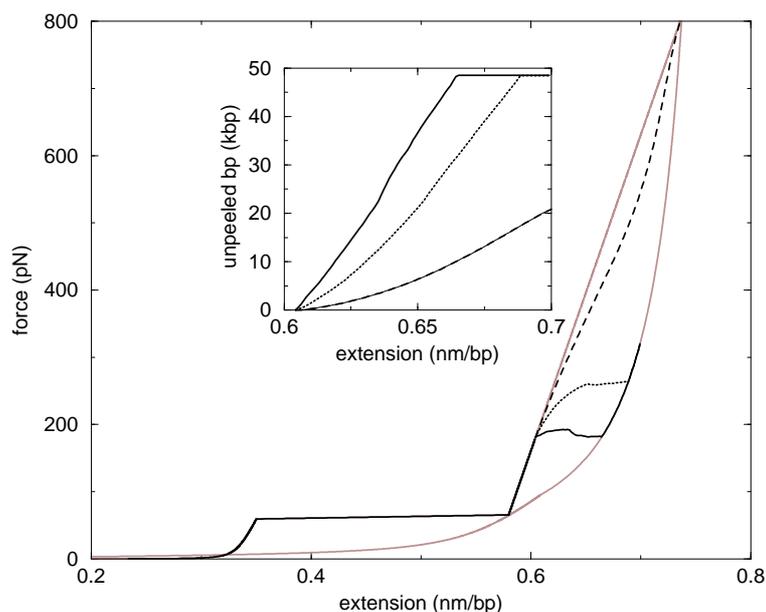


FIG. 2.23 – Courbe de force-extension pour un ADN de λ phage avec un nick au début de la séquence, étiré avec une vitesse d'extension constante. Les interactions d'appariement et l'élasticité de l'ADN simple brin sont calculées pour les conditions suivantes : 150 mM NaCl, pH 2.5, 25 C. Différentes vitesses d'extension sont montrées (courbe en noir : $v = 100\text{nm/s}$ (plein), $v = 300\text{nm/s}$ (pointillés), $v = 1500\text{nm/s}$ (tirets)). La force est montrée en fonction de l'extension par paire de base : pour des extensions lentes l'unpeeling commence vers 170 pN ; une extension rapide provoque l'unpeeling de la molécule à des forces plus élevées. Les courbes en gris montrent la caractéristique force-extension à l'équilibre.

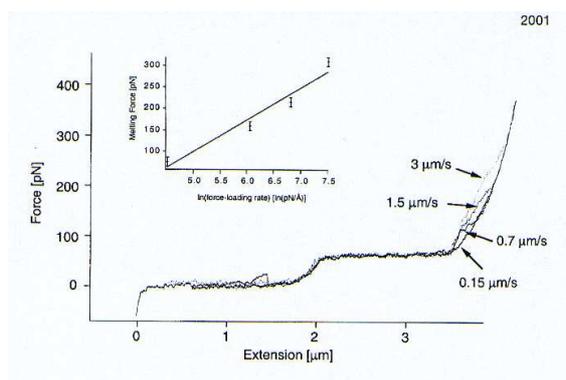


FIG. 2.24 – Courbe force-extension dans les expériences faites par le groupe de Gaub en AFM. On voit les deux transitions et la dépendance en la vitesse d'étirement dans la deuxième transition.

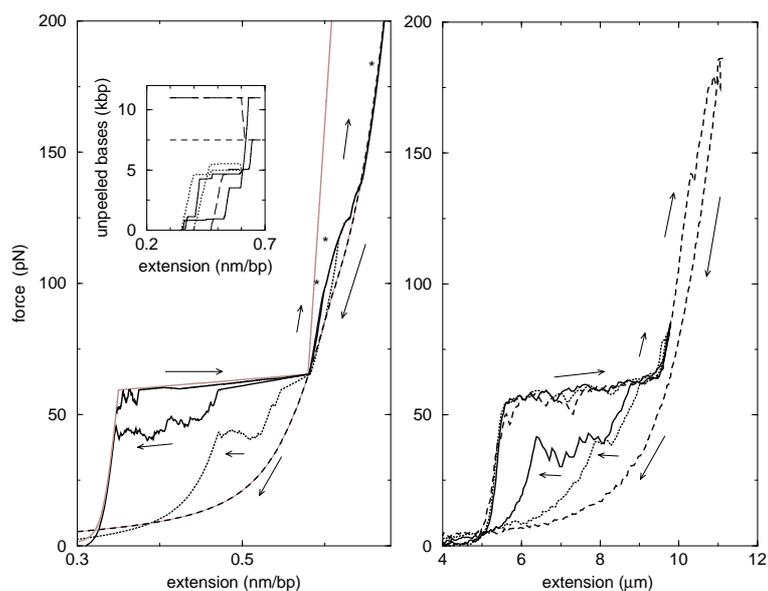


FIG. 2.25 – Courbes force-extension pour la molécule Gmac à 37 degrés. Gauche : théorie. Droite : expérience.

expérimentales de la Figure 2.24. L'autre caractéristique intéressante observée dans les expériences est que, pour des séquences riches en AT et à basse concentration ionique ou à des températures élevées, la transition de détachement (unpeeling) peut entrer en compétition avec la transition dsDNA/SDNA. Ainsi, avant la transition dsDNA/SDNA, une région riche en AT et près d'un nick peut commencer à s'ouvrir puis cette ouverture s'arrêtera sur une région riche en GC. Le reste de la molécule transite donc à la forme S, ce qui amène des hysteresis quand on relâche la force dans un cycle d'étirement-relaxation. Dans la Figure 2.25, nous comparons les données expérimentales et les résultats de la théorie pour la séquence Gmac avec 70% de base AT et à une température de 35 C. L'hysteresis est assez bien reproduit et augmente avec la force maximale atteinte pendant l'extension, c'est-à-dire le pourcentage de séquence dénaturée.

Chapitre 3

Etude de la dynamique persistante d'un neurone isolé

Ce travail a été motivé par les expériences faites au Laboratoire de Dynamique des Fluides Complexes (LDFC) à Strasbourg par L. Bourdieu, D. Chatenay et C. Wyart. Parmi les deux manuscrits soumis pour publication, A9 décrit les méthodes et résultats expérimentaux et A10 décrit la théorie. Dans les expériences, on fait croître des neurones, extraits de l'hypocampe de foetus de rats, en culture [36]. Les neurones se développent de manière satisfaisante jusqu'à 5 semaines [36]. On peut aussi guider la croissance des axones pour dessiner des réseaux à architecture contrôlée.

3.1 Système expérimental et persistance de l'activité.

Nous nous sommes particulièrement intéressés à un système très simple fait d'un seul neurone connecté avec lui-même (autapse). Pour obtenir ce système, il suffit de réaliser une culture de neurones très peu dense : le neurone doit développer des connexions pour survivre et, s'il n'y a pas d'autres neurones dans le voisinage, il se connecte avec lui-même. Comme schématisé en Figure 3.1 (gauche), le neurone développe de plus en plus de connexions au fur et à mesure qu'il vieillit. Avec des pipettes, on peut mesurer l'activité électrique et aussi déclencher des potentiels d'action. De manière surprenante, les autapses de plus de 2 semaines possèdent une activité spontanée et persistante. Ces neurones, en effet, présentent des bouffées de potentiels d'actions illustrées sur la Figure 3.1 pour des âges de 2, 3 et 4 semaines respectivement. Les bouffées peuvent commencer soit de manière spontanée soit à la suite d'un potentiel d'action déclencheur. La durée des intervalles entre potentiels d'action (ISI) dans une bouffée tout comme la durée des bouffées (BD) sont aléatoires. Leurs distributions mesurées sur des neurones de 2, 3 et 4 semaines sont montrées en Figures 3.3 et 3.4. En particulier, les ISI sont centrés autour de 5 ms pour 2 semaines, 200 ms pour 3 semaines et 1 seconde pour 4 semaines. Les deux questions que nous nous sommes posées sont les suivantes : comment peut-il y avoir une activité persistante dans un neurone isolé ? Pourquoi la fréquence d'émission de potentiel d'actions dans une bouffée diminue-t-elle avec

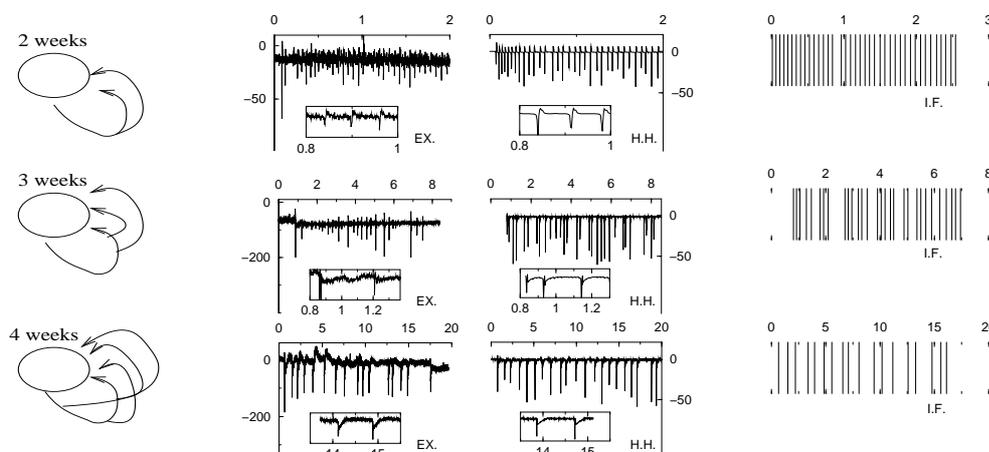


FIG. 3.1 – Schéma du système autaptique (gauche) montrant l'accroissement des connexions synaptiques avec l'âge (2, 3 et 4 semaines de haut en bas) et relevés expérimentaux (EX, milieu) et théoriques (modèle HH, droite) des courants. Les unités sont les pA (EX) & $\mu\text{A}/\text{cm}^2$ (HH) et les secondes. Les données expérimentales témoignent de la présence de bouffées d'activité pour 3 neurones représentatifs parmi 20 enregistrements. La simulation du modèle HH montre la dynamique des courants pour des bouffées de durées équivalentes. Les encarts agrandissent deux potentiels d'action consécutifs et choisis dans les bouffées, avec des ISI typiques de 0.05, 0.2 et 1 sec pour 2, 3 et 4 semaines respectivement. Le courant AMPA (négatif immédiatement après le potentiel) est très petit et invisible pour deux semaines, grand et observable pour trois semaines, très important pour 4 semaines.

l'âge du neurone, c'est-à-dire le développement de connexions? La présence de l'activité persistante est assez intrigante parce que l'on est plutôt enclin à penser que le maintien de l'activité est lié à la présence d'un réseau de neurones afférents s'activant suivant un certain ordre, par exemple, un réseau constitué de populations de neurones différents (excitateurs ou inhibiteurs), et donnant lieu à des oscillations de l'activité [37, 38, 39, 40, 41, 42]. La distribution d'ISI expérimentale montre que le système autaptique, lui-aussi, possède un rythme d'émission bien établi. Il existe donc un mécanisme de régulation interne de l'activité. Plus précisément, il est nécessaire qu'une rétroaction négative soit présente afin que, au cours du développement de connexions, le rythme d'émission n'augmente pas mais au contraire diminue. Mon activité théorique a consisté, d'une part, en une contribution à l'analyse des données (à travers le développement d'un programme pour extraire la présence de potentiels d'actions à partir de la mesure de l'intensité de fluorescence d'une sonde calcique et dans l'analyse des courants d'émission spontanée appelés miniatures); d'autre part, j'ai effectué des simulations du comportement d'un neurone autaptique à partir d'un modèle de type Hodgkin–Huxley (HH)[43] incorporant les courants neuronaux déduits des données expérimentales, et j'ai étudié un modèle du type Integrate–and–Fire (IF), moins détaillé mais soluble analytiquement. Les deux modèles, que je décris plus en

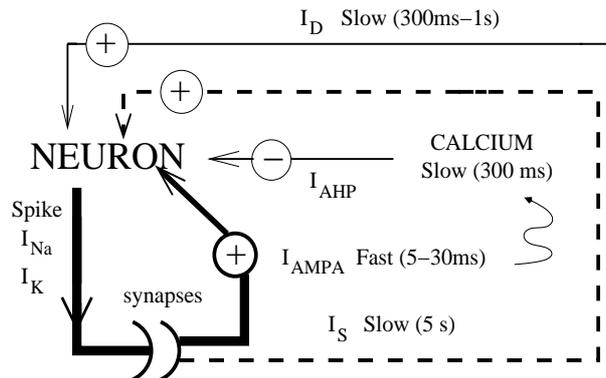


FIG. 3.2 – Courants dans le modèle de neurone autaptique. Les épaisseurs des traits sont proportionnelles aux amplitudes des courants. Les courants présynaptiques I_{Na} et I_K suivent un potentiel. Les courants postsynaptiques sont, voir texte : I_{AMPA} (trait épais), responsable de la dépolarisation de la membrane et de l'entrée Ca dans la cellule (ligne ondulée) ; I_D , un courant dépolarisant lent (trait fin) ; I_S , dû à l'excès de courants synaptiques spontanées et aléatoires (miniatures) pendant une bouffée. La rétroaction négative est représentée par un courant de Potassium hyper-polarisant, I_{AHP} , dont l'amplitude dépend du niveau en Calcium.

détail par la suite, permettent de reproduire les données expérimentales sur l'activité et de conjecturer que le mécanisme de rétroaction se fonde sur la présence d'un courant inhibiteur dépendant du calcium et que le bruit synaptique (miniature) joue un rôle important dans le maintien de l'activité et dans l'arrêt des bouffées de potentiels d'actions.

3.2 Modèle de Hodgkin-Huxley pour une autapse

Le modèle HH décrit le système autaptique sous la forme d'une dynamique de membrane :

$$C \frac{dV}{dt} = -(I_l + I_{NA} + I_K + I_{AMPA} + I_{AHP}(Ca) - I_D + I_S) \quad (3.1)$$

Les différents courants sont représentés en Figure 3.2 ; I_l est le courant de fuite qui fixe le potentiel de repos de la membrane à sa valeur d'équilibre de -60 mV, I_{NA} et I_K sont les courants qui génèrent les potentiels d'action et sont décrits dans le modèle HH par des variables de porte avec une dynamique d'activation dépendante du potentiel. Chaque potentiel d'action entraîne la production d'un courant d'auto-stimulation constitué, selon les mesures expérimentales (voir Article A9), d'une composante AMPA de grande amplitude et de courte durée suivie par une composante I_D de faible amplitude et de longue durée. Les composantes I_{AMPA} et I_D entrent dans le neurone environ 5 ms après l'émission du potentiel et sont fittées à partir des données expérimentales de la manière suivante : $I_{AMPA} = g_{AMPA} V$ où la conductance g_{AMPA} décroît exponentiellement (l'amplitude et la

constante de temps étant fixées à 0.1, 3, 20 mS/cm² et 10, 20, 30 ms pour 2, 3, 4 semaines respectivement), I_D a une amplitude fixée à $I_D^0 = 2$ mA/cm² et décroît exponentiellement avec une constante de temps de 0.3 s pour 2 semaines et 1 seconde pour 3 et 4 semaines. I_S est le courant synaptique spontané ou miniatures. Nous avons observé que, après un potentiel d'action et pendant quelques secondes ($\simeq 5$ sec), les intervalles de libération spontanée se rapprochent et passent d'environ 0.1 sec pour un neurone au repos à 10 (respectivement 20) ms pour un neurone actif et jeune c'est-à-dire âgé de 2 semaines (respectivement, vieux c'est-à-dire âgé de 3 ou 4 semaines), voir article A9. Nous avons fitté les miniatures, c'est-à-dire la distribution des temps d'émission et la distribution des amplitudes, à partir des données expérimentales. Dans les simulations du modèle de HH, j'ai considéré soit les deux distributions ou seulement celle des amplitudes, en fixant l'intervalle d'émission à sa valeur moyenne (cette approximation est très utile pour la résolution analytique du modèle IF). La distribution des amplitudes est fittée par une Poissonienne. On peut prendre en compte la distribution des temps d'émission de manière effective en modifiant la distribution des amplitudes; en particulier, les intervalles rapprochés ont tendance à favoriser la sommation de l'amplitude d'une miniature avec celle de la miniature précédente qui n'a pas eu le temps de relaxer à zero. Ainsi la distribution effective des amplitudes présente une queue positive dûe à cet effet de sommation.

Le mécanisme de maintien de l'activité est esquissé en Figure 3.2. Après un potentiel d'action, la composante excitatrice AMPA dépolarise le potentiel de membrane mais ne peut pas déclencher à elle seule un nouveau potentiel d'action parce que les variables de porte des canaux Sodium et Potassium n'ont pas eu le temps de revenir à leurs valeurs de repos apres le potentiel d'action, et le neurone n'est pas activable. En d'autres mots, le neurone est dans sa période réfractaire; la dépolarisation du potentiel de membrane, qui est visible dans la bosse sur le flanc droit du potentiel d'action, dans les enregistrements en 'Current Clamp'(voir par exemple la Fig. 9 de la Publication A9), fait entrer du calcium. Cette dépolarisation, et donc la quantité de calcium entrante, augmente avec l'âge du système et les développements d'auto-connexions. Un nouveau potentiel d'action peut être au contraire déclenché par la composante lente I_D à laquelle les miniatures s'ajoutent. La dernière composante introduite, $I_{AHP} = g_{AHP}Ca/k_d(V - V_K)$, est un courant potassique et donc inhibiteur, qui dépend du calcium [37, 38]. Dans notre cas, le calcium est un bon candidat pour être à la base du mécanisme de rétroaction qui produit des intervalles de plus en plus élevés avec l'âge du système, parce nous savons que la quantité de calcium qui rentre après chaque potentiel d'action augmente avec l'âge du système. Le courant I_{AHP} est à la base du mécanisme d'adaptation de la fréquence d'émission d'un neurone sujet à un courant exciteur constant : la fréquence des potentiels d'actions est d'abord élevée; le calcium, à cause de sa relaxation lente, s'accumule; l'inhibition augmente ensuite et la fréquence d'émission, par conséquent, diminue jusqu'à atteindre une valeur d'équilibre [37, 38]. Cette adaptation est également observable dans les neurones de notre préparation. On obtient à partir de l'équation dynamique 3.1, par intégration numérique, l'allure du potentiel de membrane et du courant neuronal en fonction du temps, pour des entrées synaptiques aléatoires. Trois bouffées d'activité (courant en fonction du temps) obtenues pour 2, 3 et 4 semaines sont montrées en Figure 3.1. J'ai aussi obtenu, à partir des simulations, les

histogrammes des Intervalles entre Spikes (ISI) et de la durée des bouffés (BD) pendant l'activité persistante d'un neurone (voir Fig. 3.3 et 3.4). Notons qu'on obtient plusieurs bouffées parce que le neurone est réexcité par un courant extérieur dix secondes après la fin de la bouffée précédente.

3.3 Modèle Integrate-and-Fire pour une autpase

En m'appuyant sur ce modèle de type HH, je suis passée à un modèle de type IF permettant un traitement analytique complet. Dans le modèle IF classique, la forme du potentiel d'action due à l'activation des courants Na et K n'est pas décrite en détail. On impose seulement qu'un potentiel d'action est déclenché dès que le potentiel de membrane dépasse un seuil θ et que le neurone est ensuite inactif pendant une période réfractaire de durée t_r . J'ai ajouté dans les courants effectifs aussi la composante AMPA, qui est prise en compte de manière effective dans le modèle IF par la valeur du calcium qui entre après chaque potentiel d'action, c_{sp} . Ce paramètre, extrait comme t_r et θ du modèle HH, dépend exclusivement de l'âge du neurone. Les courants lents, I_l , $I_{AHP}(Ca)$, I_D et I_S , font partie intégrante du modèle IF. On se concentre sur la dynamique couplée de l'émission de potentiels d'actions et de l'entrée et la relaxation du calcium. Le modèle IF est donc décrit par les équations

$$\begin{aligned} C \frac{dV}{dt} &= -g_l (V - V_l) - g_{AHP} \frac{Ca}{kd} (V - V_K) + I e^{-(t-t_n)/tl} - g_s s(t) V \\ \frac{dCa}{dt} &= c_{sp} \sum_{i=1}^n \delta(t - t_n - \delta t_{ca}) - \frac{Ca}{t_{Ca}} \\ \frac{ds}{dt} &= \sum_{j=1}^k n_s(j) \delta(t - t_j) - \frac{s}{t_s} \end{aligned} \quad (3.2)$$

Pour chaque âge du système (codé par la valeur de c_{sp}), il y a un temps optimal, après un potentiel d'action, pour l'émission d'un autre potentiel d'action. Ce temps correspond à l'intervalle auquel le niveau en calcium a suffisamment décréu mais la composante lente est encore présente. L'émission ou pas d'un potentiel d'action dépend aussi de la valeur des miniatures qui entrent dans le neurone aux environs de ce temps optimal. Ceci explique la distribution des ISI. Le modèle IF permet d'abord d'avoir une idée des valeurs des paramètres qui permettent la persistance de l'activité et de la valeur moyenne des ISI lorsque l'on fixe l'amplitude des miniatures à leur valeur moyenne. En effet, les neurones, dans la population étudiée, ne sont pas tous identiques; ainsi, la conductance de membrane peut varier entre 0.1 et 0.2 mS/cm². La distribution des miniatures varie également de neurone à neurone. En conséquence, certains neurones sont plus actifs que d'autres et la dynamique de l'activité, caractérisée par la distribution des ISI et des BD, s'en ressent même si elle est en général en bon accord avec la description donnée ci-dessus. Dans la publication A8, je montre (Fig. 4) le diagramme de phase de l'activité dans le plan de la constante de temps de la composante lente, t_D , et la conductance moyenne des miniatures pour 2, 3 et

4 semaines et pour $g_l = 0.1, 0.2$. L'intervalle typique entre deux potentiels d'action peut être calculé en insérant la condition de stationnarité pour le calcium, c'est-à-dire que la diminution du niveau en calcium entre deux potentiels d'action est exactement compensée par son augmentation après un potentiel d'action, dans la dynamique d'émission donnée par la première équation de 3.2. L'intégration numérique du modèle IF permet d'observer la dynamique de l'activité et d'obtenir les histogrammes des ISI et BD en présence de bruit synaptique. Le modèle IF permet aussi de calculer de manière analytique la distribution des ISI et BD quand on considère que les miniatures sont distribuées de manière poissonnienne. Les distributions des ISI et des BD dépendent de la distribution stationnaire du calcium après un potentiel, distribution centrée autour de la valeur d'équilibre stationnaire du calcium que l'on obtient lorsque l'on fixe les miniatures à leur valeur moyenne. Si cette distribution est très large, il en sera de même pour celles des ISI parce que la valeur du niveau en calcium après un potentiel influence fortement celle de l'ISI. En outre, la présence d'un niveau élevé en calcium peut suffire à arrêter une bouffée. Dans ce cas, la composante longue peut décroître avant que ce niveau soit suffisamment bas. Une fluctuation positive du calcium a lieu si par exemple une fluctuation positive des miniatures (c'est-à-dire une entrée plus grande que la normale) a provoqué un rapprochement temporel de deux potentiels d'action successifs (voir la dynamique du calcium dans la Figure 3 de la Publication A8).

En discrétisant les intervalles entre potentiels en multiples des temps d'émission des miniatures δ_s , ce qui est possible car la valeur de $\delta_s = 10, 20$ ms pour 2 et 3-4 semaines est toujours plus petite que l'ISI centré autour de 50 ms, 200ms et 1 s respectivement, on peut calculer la probabilité d'émission pendant l'intervalle i après le potentiel précédent étant donné le niveau de calcium Ca après le potentiel d'action précédent. Cette probabilité se calcule à partir de la valeur minimale s_m de l'amplitude des miniatures qui permet pour chaque intervalle de temps et de la valeur Ca l'émission d'un potentiel. La probabilité de ne pas avoir émis un potentiel est simplement donnée par la distribution cumulée des miniatures jusqu'à s_m ; la probabilité d'émission pendant l'intervalle i s'écrit donc comme la probabilité de ne pas avoir émis avant, multipliée par la probabilité d'émettre pendant l'intervalle i précisément. La probabilité d'émission pendant l'intervalle $i\delta_s$ étant donné le niveau en calcium Ca juste après le potentiel d'action précédent donne accès à la matrice de transition entre deux niveaux successifs du Calcium (voir Formule 4 de la Publication A8). Pour obtenir la distribution stationnaire du Calcium, il suffit de diagonaliser cette matrice et de chercher le vecteur propre de valeur propre nulle (et maximale). Dans ce but, j'ai diagonalisé numériquement cette matrice avec une procédure itérative. La distribution des ISI obtenue analytiquement est en très bon accord avec celle obtenue par les simulations numériques (HH et IF) et également avec les distributions expérimentales des 3 neurones de la Figure 3.1. Notons que les histogrammes expérimentaux et issus des simulations incluent aussi les premiers spikes dans une bouffée, avant que l'équilibre stationnaire ne soit rejoint. Les ISI correspondants sont petits et donnent lieu à la queue de gauche de la distribution qui n'est pas présente dans les résultats analytiques. La valeur médiane des ISI obtenus est de 0.06, 0.2, 1 sec pour 2, 3, 4 semaines respectivement (voir Figure 3.3). Pour calculer la

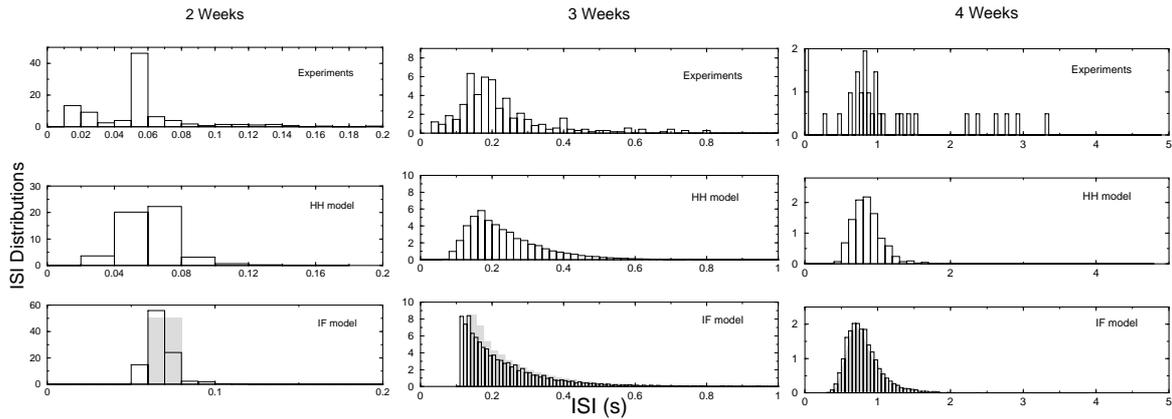


FIG. 3.3 – Distributions des intervalles entre spikes (ISI) de trois neurones de 2, 3, 4 semaines : Haut : résultats expérimentaux, milieu : résultats de simulations sur le modèle de Hodgkin-Huxley (HH), bas : résultats du modèle Integrate-and-Fire (IF). Les histogrammes en trait plein, respectivement remplis de gris, correspondent aux simulations, respectivement aux résultats analytiques.

distribution des BD, j'ai utilisé la technique des fonctions génératrices. Soit

$$G(x) = \sum_i Q(i)x^i$$

où $Q(i)$ est la probabilité que deux potentiels d'action successifs soient séparés par un intervalle $i \delta_s$. Le coefficient de x^k dans $G(x)^n$ correspond à la probabilité d'avoir un bouffée de n potentiels d'action de longueur $k \delta_s$. En sommant sur n , on obtient donc la probabilité qu'un burst ait une longueur $k \delta_s$ sous la forme du coefficient de x^k dans la série

$$\tilde{G}(x) = \frac{G(x)}{1 - G(x)} \quad .$$

On obtient une décroissance exponentielle de la distribution avec une constante de temps de 5-10 s pour les trois neurones, sans forte dépendance avec l'âge du système. La distribution de BD est montrée en Figure 3.4. Il faut remarquer que nous disposons d'un nombre limité de données (9, 13, 11 bouffées pour 2, 3, 4 semaines) pour calculer les distributions.

En conclusion, la comparaison entre données expérimentales et modélisation montre qu'on peut proposer un mécanisme de persistance de l'activité dans un neurone isolé. Le courant lent I_D met le neurone dans un état depolarisé (cet état est aussi observé dans différents réseaux oscillatoires et est dû d'habitude aux courants afférents d'autres neurones) sur lesquels les entrées spontanées synaptiques s'ajoutent pour déclencher un nouveau spike et faire persister l'activité. En outre, le mécanisme d'adaptation dépendant du calcium peut expliquer l'augmentation de l'ISI avec l'âge du neurone, et la présence couplée de ce mécanisme et du bruit synaptique peut suffire à expliquer l'arrêt des bouffées.

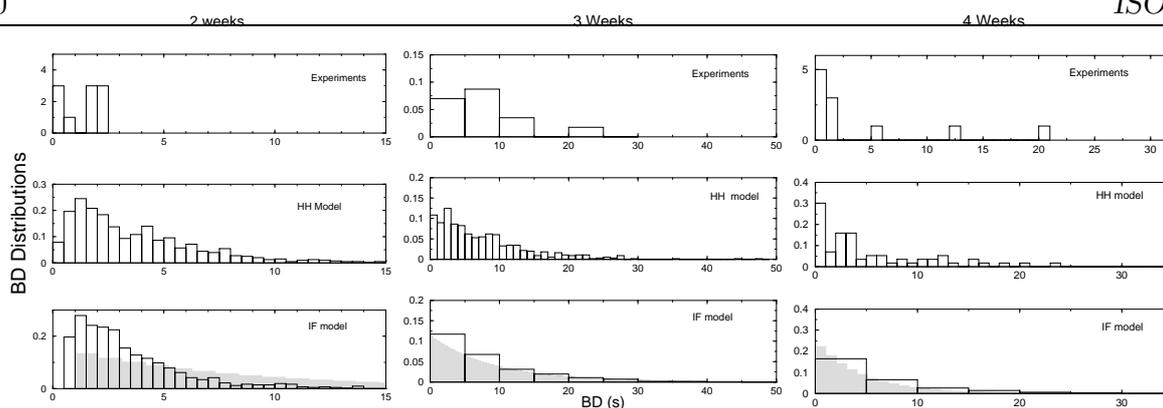


FIG. 3.4 – Distributions des durées des bouffées (BD) de trois neurones de 2, 3, 4 semaines : Haut : résultats expérimentaux, milieu : résultats de simulations sur le modèle de Hodgkin-Huxley (HH), bas : résultats du modèle Integrate-and-Fire (IF). Les histogrammes en trait plein, respectivement remplis de gris, correspondent aux simulations, respectivement aux résultats analytiques.

Nous avons négligé dans la modélisation d'autres effets qui pourraient aussi contribuer à l'arrêt des bouffées, comme la déplétion synaptique, ou une modulation de l'excitabilité cellulaire due à l'activité. Enfin, il serait intéressant de mettre en réseau plusieurs neurones de ce type pour en étudier l'activité et la comparer avec les données expérimentales qui sont disponibles.

Chapitre 4

Mécanique statistique et problèmes d'optimisation

4.1 Dynamique d'algorithmes de décision.

En collaboration avec R. Monasson (voir les articles A11, A12 pour une introduction et A13, A14 pour les aspects plus techniques)

Le problème de la satisfaisabilité (SAT) consiste à se demander si un ensemble de contraintes logiques portant sur des variables booléennes (ayant seulement deux valeurs possibles, vrai V ou faux F) peuvent être satisfaites simultanément ou non. Considérons la restriction dite de la 3-SAT où chaque contrainte, appelée aussi clause, porte sur trois variables booléennes exactement. Une instance du problème 3-SAT est définie comme une liste de P clauses, chacune étant le OU logique de trois variables x_i ou de leurs négations logiques \bar{x}_i . L'indice i des variables est un nombre entier compris entre 1 et N , le nombre total des variables. Un exemple de clause est $(x_3 \vee \bar{x}_6 \vee x_7)$. Pour satisfaire cette clause, il suffit de donner à une de ces variables la valeur de vérité requise par la clause, par exemple $x_6 = F$. Une instance, c'est-à-dire une liste de clauses, est dite satisfaisable (sat) s'il existe, parmi les 2^N configurations possibles des N variables, au moins une configuration satisfaisant toutes les clauses. Dans le cas contraire, les clauses sont incompatibles entre elles et l'instance n'est pas satisfaisable (unsat). Résoudre une instance de 3-SAT consiste à trouver explicitement une solution si l'instance est sat ou à montrer qu'il n'en existe pas si l'instance est unsat. Même avec les meilleurs algorithmes disponibles, les temps de résolution, qui définissent la complexité computationnelle, peuvent augmenter dramatiquement (exponentiellement) avec la taille des problèmes, c'est-à-dire les nombres de variables N et de clauses P . 3-SAT est donc un paradigme de problème computationnel difficile (classe NP-Complet) et, à ce titre, joue un rôle central en informatique théorique.

Les informaticiens ont entrepris il y a quelques années l'étude d'un modèle statistique des problèmes SAT survenant dans la pratique : la 3-SAT aléatoire. Les P clauses sont tirées au hasard, indépendamment les unes des autres ; à l'intérieur d'une clause, chaque variable ou sa négation apparaît de manière équiprobable. Un paramètre important est

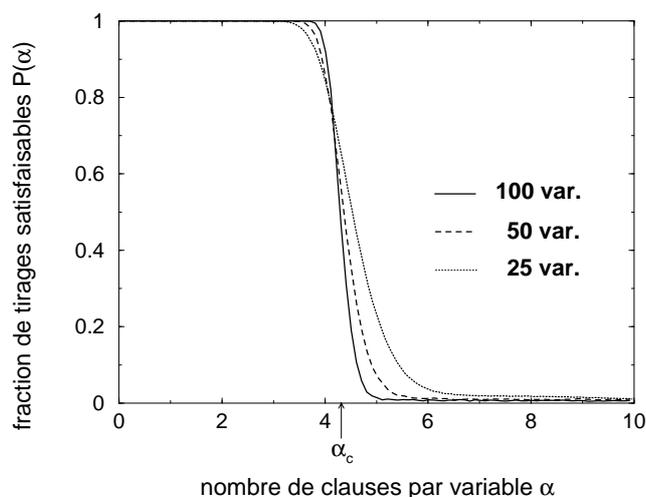


FIG. 4.1 – Probabilité $P(\alpha)$ qu'un ensemble de αN clauses formant une instance 3-SAT soit satisfaisable pour différents nombres N de variables. Cette probabilité est obtenue numériquement en générant, pour chaque valeur de α et N , 10000 tirages d'instances 3-SAT et en mesurant la fraction de celles-ci qui admettent au moins une solution. Dans la limite de taille infinie, $\alpha_C \simeq 4.3$ sépare une région satisfaisable ($\alpha < \alpha_C$, $P = 1$) d'une région non satisfaisable ($\alpha > \alpha_C$, $P = 0$). Une telle transition existe aussi pour le problème 2-SAT aléatoire, où toutes les clauses contiennent deux variables ; le seuil vaut alors $\alpha_C = 1$ exactement.

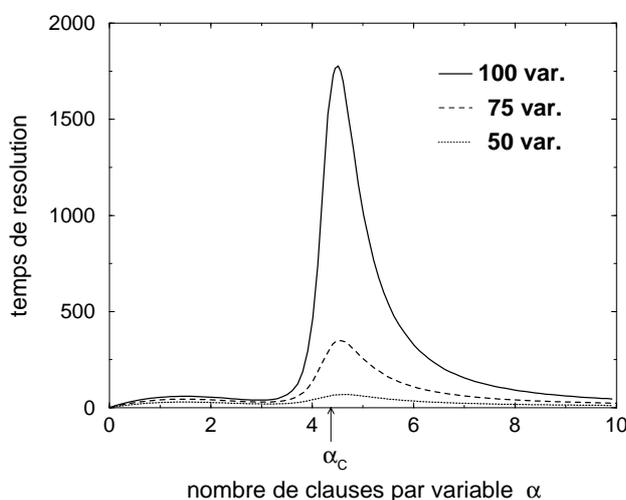


FIG. 4.2 – Temps de résolution (complexité calculatoire) de 3-SAT aléatoire en fonction du nombre de clauses par variable, α , et pour trois tailles N différentes. Le temps est estimé à partir de la grosseur de l'arbre de recherche engendré par l'algorithme de résolution, voir le nombre de noeuds en Fig. 4.4. La croissance avec la taille est maximale au seuil α_C . Elle est exponentielle autour du seuil et dans toute la région non satisfaisable, mais de moins en moins fortement lorsque α augmente.

le rapport $\alpha = P/N$ du nombre de contraintes par variable. Intuitivement, α représente un nombre de contraintes par degré de liberté. Il est légitime de penser que, à N fixé, un problème 3-SAT aléatoire admettra des solutions si α est petit (problème sous-contraint) et sera de plus en plus difficile à satisfaire au fur et à mesure que α augmente (problème sur-contraint). Les expérimentations numériques réalisées pour de grandes tailles N montrent qu'une séparation nette apparaît à une valeur critique $\alpha_c \simeq 4.3$: les instances ayant moins de $\alpha_c N$ clauses sont presque toujours sat, alors que celles comportant plus de $\alpha_c N$ clauses n'ont presque jamais de solution et sont unsat (voir Figure 4.1).

La transition survenant dans le problème 3-SAT peut être étudiée à l'aide des concepts et méthodes de la physique statistique de verres de spins. Monasson et Zecchina ont introduit en 1996 une énergie représentant le nombre de clauses non satisfaites pour chaque configuration des variables-spins [46]. A partir de la définition de cette énergie, ils ont calculé avec la méthode des répliques (pour moyenner sur le choix des clauses) la capacité critique de la transition sat/unsat dans l'approximation de symétrie des répliques, donnant un seuil $\alpha_C = 4.60$. En fait, cette symétrie est brisée car les solutions se regroupent en agrégats (voir section 4.3); Biroli, Monasson et Weigt ont réalisé un calcul avec un pas de brisure de symétrie des répliques qui faisait appel à une approximation variationnelle donnant $\alpha_c = 4.48$; récemment, Mézard, Parisi et Zecchina ont obtenu, toujours avec un pas de brisure mais sans approximation variationnelle, un seuil théorique valant $\alpha_c = 4.27$ en très bon accord avec les simulations numériques. Monasson et Zecchina ont ensuite introduit un modèle SAT mixte, mélangeant des clauses à 2 variables et à 3 variables. Ce modèle, appelé 2+p-SAT, est caractérisé par, outre α , un deuxième paramètre 'thermodynamique', à savoir la fraction p de clauses de longueur 3 par rapport à l'ensemble des clauses. Le seuil de transition $\alpha_c(p)$ peut être estimé théoriquement en fonction de p , et définit une ligne critique dans le plan p, α qui sépare une phase sat (au-dessous de la ligne) d'une phase unsat (au-dessus de la ligne), voir Figure 4.3. La ligne critique abrite un point tricritique situé en $p_0 \simeq 0.4$, qui sépare une transition du premier ordre ($p > p_0$), d'une transition du deuxième ordre ($p < p_0$). Ces résultats théoriques ont été accompagnés par les simulations numériques faites par Kirkpatrick et Selman, qui ont observé que la complexité computationnelle des instances 2+p-SAT critiques (avec pour paramètres $p, \alpha_c(p)$) passe de polynômiale à exponentielle en N lorsque p traverse la valeur tricritique p_0 .

Le but de notre travail était de comprendre cette observation, et plus généralement, comme dépend la complexité de résolution de 3-SAT en fonction du rapport α du nombre de clauses par variable. (voir Figure 4.2) L'algorithme que nous avons étudié est la procédure dite de Davis et Putnam (DP) [45]. Selon une stratégie très répandue en optimisation, la procédure de DP consiste en une exploration systématique de l'espace des phases grâce à une succession d'essais quant aux choix des variables et de corrections d'erreurs, qui peuvent être représentés sous forme d'un arbre de recherche (Figure 4.4). J'explique brièvement le principe de cette recherche et la routine que nous avons implémentée dans la simulation numérique :

(1) on choisit une variable et sa valeur V ou F . A chaque fois qu'une variable est choisie, un noeud est ajouté à l'arbre. Selon que la valeur est vraie ou fausse, on emprunte une des deux branches possibles.

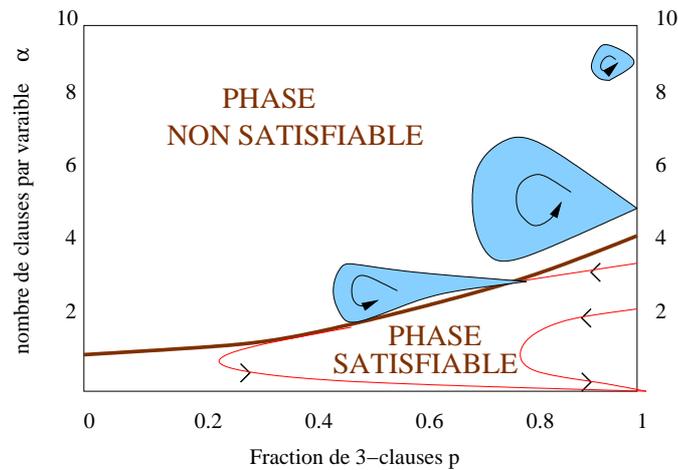


FIG. 4.3 – Diagramme de phase de 2+p-SAT. La ligne critique en trait épais $\alpha_c(p)$ sépare les phases satisfaisables et non satisfaisables. Les trajectoires des instances 3-SAT à résoudre sont représentées pour des rapports initiaux $\alpha = 2, 3.5, 5$ et 9 . Lorsque $\alpha = 3.5$, la trajectoire traverse la frontière entre les différentes phases (phase satisfiable, complexité polynômiale et phase non satisfiable, complexité exponentielle).

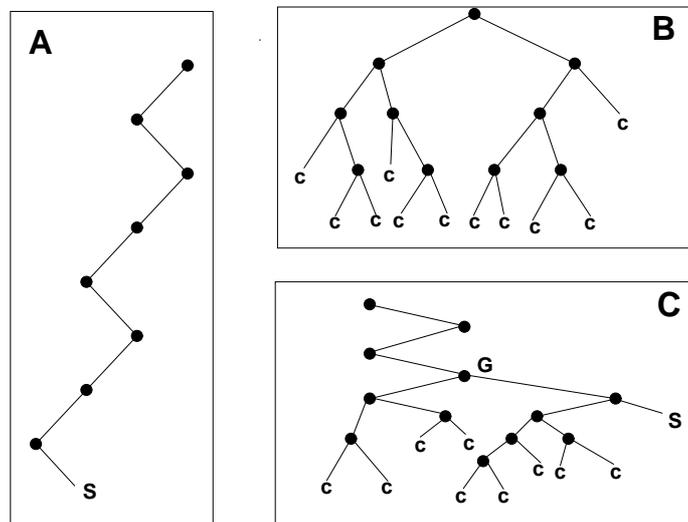


FIG. 4.4 – Arbres de recherche créés par DP. Les extrémités des branches indiquent l'apparition de contradictions (C) ou d'une solution (S). **A.** branche unique (région $0 < \alpha < \alpha_L \simeq 3$). **B.** arbre touffu (région $\alpha > \alpha_c$). **C.** cas mixte, branche puis arbre (région $\alpha_L < \alpha < \alpha_c$). La complexité de résolution par DP est définie comme le nombre de noeuds dans l'arbre.

(2) Le long d'une branche, on analyse toutes les implications logiques du dernier choix effectué. Par exemple, la variable 6 vient d'être fixée à faux F . On scrute une à une les clauses où figure x_6 ou \bar{x}_6 . Celles qui contiennent \bar{x}_6 sont satisfaites et éliminées. Celles qui comportent x_6 sont simplifiées ; par exemple $(x_3 \vee x_6 \vee \bar{x}_7)$ devient $(x_3 \vee \bar{x}_7)$. Plusieurs cas peuvent alors se présenter. **(a)** Il n'y a pas de clauses unitaires (comportant une seule variable) : on regarde si toutes les clauses sont satisfaites, dans ce cas on a trouvé une solution, la branche est marquée d'un S et on arrête la recherche ; sinon, on retourne à l'étape 1. **(b)** Des clauses unitaires sont apparues au cours de la simplification. On est obligé de fixer les variables de façon à les satisfaire. C'est l'étape de la propagation unitaire. Un tel choix forcé n'est pas représenté par un noeud dans l'arbre. Si, à un certain moment de la propagation, deux clauses unitaires contradictoires x et \bar{x} surgissent, la branche meurt et est affublée d'un C. On modifie alors le dernier choix (remontée dans l'arbre jusqu'au noeud le plus proche portant une seule branche) et on poursuit sur une nouvelle branche en allant à l'étape 2.

La recherche se conclut quand on trouve une solution ou bien quand on montre que les clauses sont incompatibles, c'est-à-dire que l'arbre est complet (à travers chaque noeud passent deux branches et toutes les branches se terminent par C). La complexité computationnelle est directement donnée par la taille de l'arbre de recherche, plus précisément par le nombre de ses noeuds. Les performances peuvent être améliorées plus ou moins empiriquement en jouant sur la règle heuristique de choix des variables (étape 1).

L'algorithme de DP définit un processus dynamique complexe. Pour l'étudier nous sommes partis de l'observation suivante. De noeud en noeud, la procédure récursive de DP transforme l'instance de 3-SAT initiale en un problème mixte 2+p-SAT contenant des clauses de longueurs deux et trois, les clauses comportant une seule variable étant éliminées à travers la propagation unitaire (étape 2c de l'algorithme ci-dessus). Ce processus de transformation de clauses est schématisé Figure 4.5. Donc, si l'on symbolise l'instance de 3-SAT initiale comme un point de coordonnées $p = 1, \alpha$ dans le diagramme de phase p, α (voir la Figure 4.3), ce point va évoluer sous l'action dynamique de l'algorithme et décrire une trajectoire dans le plan. C'est cette dynamique que nous avons étudiée pour accéder à la complexité de résolution, tout d'abord avec des expériences numériques puis par une théorie analytique.

L'emplacement des trajectoires dans le diagramme de phase du modèle 2+p-SAT donne des renseignements importants sur la complexité computationnelle. L'analyse montre que, si la trajectoire reste confinée à la phase sat, (voir trajectoire qui part de $\alpha = 3, p = 0$ de Figure 4.3) presque aucune contradiction ne peut survenir. Les clauses sont progressivement satisfaites et éliminées, et la trajectoire s'arrête sur l'axe $\alpha = 0$ où une solution est trouvée. Ce cas correspond à de petites valeurs du nombre initial de clauses par variable, $\alpha < 3.003$. L'arbre de recherche typique est constitué d'une seule branche (Figure 4.4A) ; la complexité est linéaire en N , parce qu'une branche ne peut pas comporter plus que N noeud.

Si la trajectoire reste au contraire confinée à la phase unsat, (voir trajectoire qui partent de $\alpha = 5, p = 0$ et $\alpha = 9, p = 0$ de Figure 4.3) des arbres complets dits de réfutation, avec toutes leurs branches qui se terminent par C (Figure 4.4B), sont construits par la procédure de DP pour s'assurer qu'il n'y a pas de solution possible. Le nombre de noeuds et donc

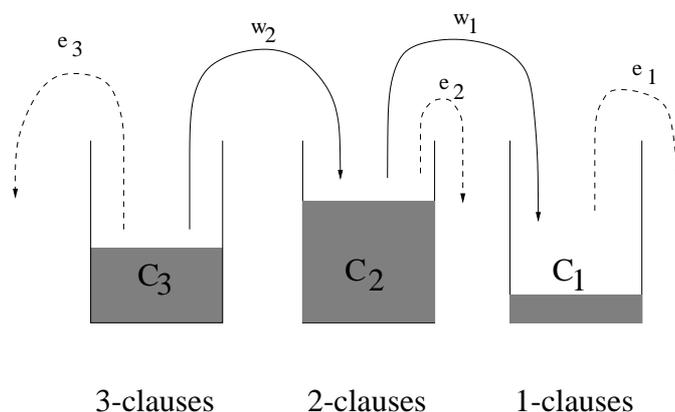


FIG. 4.5 – Représentation schématique de la dynamique des clauses. Les clauses sont séparées en trois récipients selon leur longueur, c'est-à-dire le nombre de variables qu'elles comportent. Chaque fois qu'une variable est assignée par DP, les clauses sont modifiées et les populations des récipients changent (lignes avec flèches). Les lignes en tirets représentent l'élimination des clauses satisfaites de longueur 1, 2 ou 3. Les lignes épaisses représentent la réduction de 3-clauses en 2-clauses ou de 2-clauses en 1-clause. Ces flux des clauses correspondent respectivement à e_1, e_2, e_3 et w_2, w_1 . Une solution est trouvée quand tous les récipients sont vides. Le niveau du récipient de droite donne le nombre de clause unitaires. Si ce niveau est bas (*i.e.* $O(1)$), la probabilité que deux clauses contradictoires x et \bar{x} soient présentes dans le récipient tend à zéro lorsque N est grand. Des contradictions auront lieu presque sûrement lorsque le niveau est haut (*i.e.* $O(\sqrt{N})$).

la complexité croît exponentiellement avec N et s'exprime comme $2^{N\omega}$, où ω dépend de la valeur initiale de α (> 4.3).

La troisième possibilité est obtenue lorsque le rapport initial est compris dans l'intervalle $3.003 < \alpha < 4.3$. Les trajectoires partent de la phase sat, mais croisent la ligne critique $\alpha_c(p)$ à un certain moment. DP a, sans le savoir, donné naissance à un problème 2+p-SAT sans solution. Le noeud de la branche qui se trouve sur la ligne critique sera donc suivi dans la majorité des cas par un sous-arbre complet de réfutation (Figure 4.4C). La trajectoire de la branche qui conduit à la solution est très atypique et ne contribue pas à la complexité, qui reste dominée par la taille du sous-arbre de réfutation.

Sur la Figure 4.3, on peut voir les trajectoires qui correspondent à ces trois possibilités. Nous avons distingué les trajectoires de branches uniques (phase sat) de celles d'arbres complets (phase unsat). Dans la région mixte $3.003 < \alpha < 4.3$, la trajectoire, initialement de branche unique, est suivie par une trajectoire de sous-arbre complet au-delà de la ligne critique.

Les trajectoires de branches uniques s'obtiennent simplement en suivant l'étude de l'évolution stochastique (la distribution des clauses et certains choix de l'algorithme sont aléatoires) des clauses réalisée par deux informaticiens, M.T. Chao et J. Franco [47]. A partir des flux représentés Figure 4.5, on peut écrire des équations différentielles couplées

pour les dérivées par rapport au ‘temps’ (c’est-à-dire le nombre de variables fixées) des valeurs moyennes des nombres de clauses de longueurs 2 et 3. Le temps auquel la trajectoire touche l’axe $\alpha = 0$ est relié directement à la complexité, linéaire en N dans ce cas.

Le logarithme ω de la complexité pour des arbres complets est bien plus difficile à calculer analytiquement et avait résisté aux efforts des mathématiciens. Il semble impossible d’étudier la vraie dynamique séquentielle, faite de descentes et remontées, que DP utilise dans sa construction de l’arbre complet. Nous avons imaginé un processus de croissance parallèle de toutes les branches de l’arbre, qui conduit au même arbre de recherche complet que la vraie dynamique de DP. Ce processus peut lui être analysé mathématiquement et suffit à calculer la taille finale de l’arbre. Nous avons pu écrire une équation d’évolution pour la distribution des branches, c’est-à-dire le logarithme du nombre de branches symbolisant une instance de 2+p-SAT avec un rapport clauses/variable égal à α et une fraction p de clauses de longueur trois. Cette distribution définit à chaque ‘temps’ t (profondeur dans l’arbre) une surface au-dessus du plan (p, α) et les coordonnées $(p(t), \alpha(t), \omega(t))$ du maximum de cette surface caractérisent les branches les plus répandues. Les trajectoires correspondantes sont reportées sur le diagramme de phase de la Figure 4.3. Une fois le nombre branches calculé, nous avons utilisé la relation, que l’on peut vérifier facilement sur un arbre complet, que le nombre de noeuds est égal au nombre de branches moins un. La complexité est donc donnée, à l’ordre exponentiel, par la valeur de $\omega(t)$ à l’endroit où les trajectoires des branches les plus répandues s’arrêtent parce qu’une contradiction survient, voir légende de la Figure 4.5. La ligne de contradiction est dessinée également sur le diagramme de phase de la Figure 4.3. L’obtention de l’équation aux dérivées partielles régissant la croissance des branches de l’arbre à demandé l’utilisation de techniques de physique statistique comme la diagonalisation d’une matrice à 6 indices, qu’on a pu réaliser à travers l’introduction d’une fonction génératrice des composantes des vecteurs propres ; la ligne de contradiction coïncide avec une transition de délocalisation des fonctions propres de cette matrice (non hermitienne) tout à fait similaire à celle qu’on a pour la transition entre les phases fermée et ouverte de l’ADN!. En outre les concepts de contribution dominante à la limite thermodynamique, point de col et transformée de Legendre sont à la base de notre calcul.

Nous avons ainsi obtenu une explication générale, à la fois qualitative et quantitative, de la courbe de la complexité de DP sur la 3 SAT en fonction de α observée dans les expériences numériques (Figure 4.2). Le logarithme ω de la complexité, pour différentes valeurs de $\alpha > 3.003$, est en très bon accord avec les résultats numériques ainsi qu’en témoigne le Tableau 1 de la Publication A13 et la Figure 4.7.

4.2 Procédure de redémarrage aléatoire.

Travail en collaboration avec R. Monasson (voir l’article A15).

Dans ce travail, nous avons étudié la distribution de temps de résolution d’instances du problème 3-SAT aléatoire avec l’algorithme DP. Le caractère stochastique de ce temps vient de l’instance (ensemble de clauses à satisfaire choisi au hasard) et des choix aléatoires

de l'algorithme quant aux variables à affecter ('run' de l'algorithme). Dans la limite des grands N , les simulations numériques suggèrent que la distribution d'instance à instance est équivalente à la distribution de run à run.

Des simulations numériques (voir l'article A15) dans la région upper-sat $3 < \alpha < 4.3$ montrent que, même si la complexité typique de résolution vaut $T = 2^{N\omega}$ avec $\omega > 0$ et est donc exponentielle grande en N , l'histogramme des temps de résolution admet aussi une région de résolution linéaire en N . Cette partie de la courbe est représentée par un pic dont la hauteur décroît très rapidement avec N . En effet, lorsque l'on mesure la probabilité cumulative d'avoir une complexité plus petite ou égale à N , on obtient $P_{lin} = 2^{-N\zeta}$. Cet exposant ζ s'avère être plus petit (en valeur absolue) que l'exposant ω de la complexité typique (à α fixé). Cela suggère une méthode alternative pour accélérer la recherche de solutions : si l'algorithme n'a pas trouvé une solution au bout du temps $T = N$, on l'arrête et on le relance (sur la même instance). Le nombre de redémarrages moyen avant de trouver une solution de cette manière est simplement $N_{rest} = 1/P_{lin}$. Donc, comme $\zeta < \omega$, cette méthode est plus rapide que l'algorithme DP habituel. Il faut remarquer que cette méthode n'est pas complète car on ne peut pas l'utiliser pour démontrer qu'il n'y a pas de solution.

Pour calculer analytiquement l'exposant ζ , nous avons utilisé la représentation en termes de trajectoires dans le diagramme de phase déjà introduite dans le cadre de l'étude de DP. Même si avec grande probabilité les trajectoires dans la région analysée trouvent une contradiction quand elle croisent la ligne critique et donc l'algorithme doit remonter dans l'arbre de recherche, certaines trajectoires atypiques réussissent à traverser la zone critique sans rencontrer de contradictions comme il est montré en Figure 4.6. Dans la théorie, on caractérise l'instance par les nombres de clauses, $\mathbf{C} = (C1, C2, C3)$, et on écrit une équation d'évolution pour la probabilité $\tilde{P}(\mathbf{C}, T)$ qu'une instance soit dans l'état \mathbf{C} après l'affectation de T variables (sans qu'il y ait eu de contradictions). L'introduction de la fonction génératrice $P(\mathbf{y}, t)$ permet d'écrire de manière compacte l'équation d'évolution de la probabilité d'un état \mathbf{C} . En outre, sous l'hypothèse de scaling $P(\mathbf{C}; T) = e^{N\tilde{\phi}(\mathbf{c}, t=T/N)}$ ou encore $P(\mathbf{y}; T) = e^{N\phi(\mathbf{y}, t=T/N)}$, on obtient que $\tilde{\phi}(\mathbf{c}, t)$ et $\phi(\mathbf{y}, t)$ sont liées par une transformée de Legendre et que $\phi(\mathbf{0}, t)$ est le logarithme de la probabilité que la branche n'a pas rencontré de contradiction après qu'une fraction t de variables ont été affectées. La valeur la plus probable des densités $c_j(t)$ des clauses portant sur j variables est égale à la dérivée partielle de ϕ en $\mathbf{y} = \mathbf{0}$.

Nous avons écrit et résolu l'équation aux dérivées partielles satisfaite par $\phi(\mathbf{y}, t)$. Cette équation s'écrit de manière différente dans la région initiale où la densité de clauses de longueur 1, c_1 , reste égale à zéro et dans la région au-delà de la ligne critique où $c_1 > 0$. Les résultats pour $\zeta = -\phi(\mathbf{0}, t_d)/\ln 2$ (où t_d est le temps auquel la trajectoire sort de la région dangereuse) sont montrés en Figure 4.7. Ils sont en très bon accord avec les résultats des simulations.

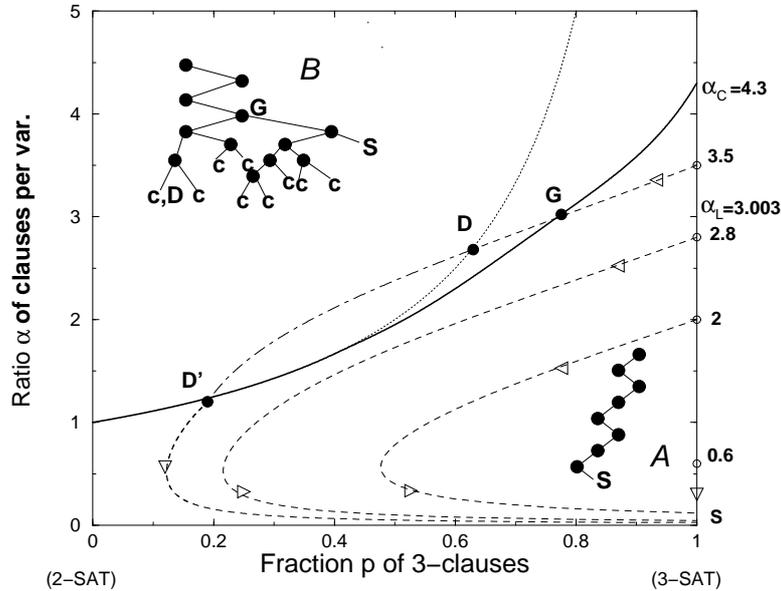


FIG. 4.6 – Diagramme de phase du modèle 2+p-SAT et trajectoires de la première branche pour des instances satisfaisables. La ligne de seuil $\alpha_C(p)$ (ligne en trait plein foncé) sépare les phases SAT (partie inférieure du plan) et non SAT (partie supérieure). Les points de départ pour les trajectoires de DP sont situés sur l'axe vertical de 3-SAT (cercles vides). Les flèches indiquent la direction du mouvement le long des trajectoires (courbes tiretées) paramétrisées par la fraction t des variables fixées par DP. Pour de petits rapports $\alpha < \alpha_L \simeq 3.003$ pour l'heuristique GUC), les trajectoires de branches restent confinées à la phase SAT et finissent en S de coordonnées (1,0) où une solution est trouvée après un processus de recherche symbolisé par l'arbre A. Pour des rapports $\alpha_L < \alpha < \alpha_C$, la trajectoire de branche intersecte le ligne de seuil en un point G. Une contradiction survient presque sûrement avant que la trajectoire ne croise la ligne $\alpha = 1/(1-p)$ (point D) et des backtrackings massifs jusqu'en G permettent de trouver une solution (arbre de recherche B). Avec une probabilité exponentiellement faible en N , la trajectoire (trait mixte, flèche pleine) est capable de traverser la région dangereuse où les contradictions surviennent avec grande probabilité (l'arbre de recherche est alors similaire à A); elle sort de cette région (point D') et se termine une fois qu'une solution est trouvée (trajectoire tiretée la plus proche de l'axe des abscisses).

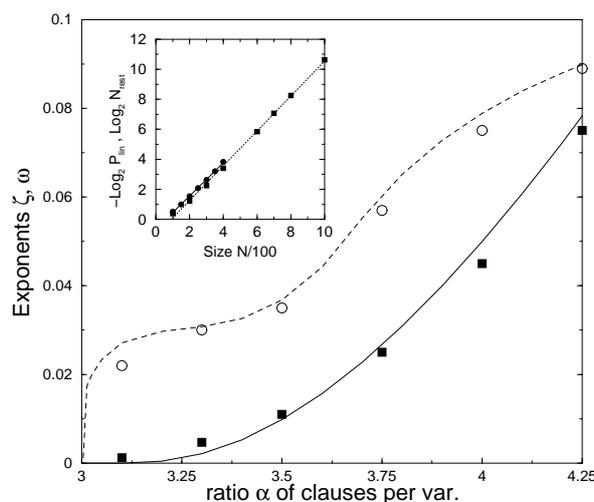


FIG. 4.7 – Logarithme de la complexité de résolution par DP (ω – simulations : cercles, théorie venant de la publication A13 : ligne pointillée) et par la méthode de redémarrages S&R (ζ – simulations : carrés, théorie : ligne pleine) en fonction du rapport α . Encart : moins le logarithme de la probabilité cumulée P_{lin} des complexités $\leq N$ en fonction de la taille pour $100 \leq N \leq 400$ (ligne pleine) ; logarithme du nombre de redémarrages N_{rest} nécessaires à la découverte d’une solution pour $100 \leq N \leq 1000$ (ligne pointillée) pour $\alpha = 3.5$. Les pentes sont $\zeta = 0.0011$ et $\bar{\zeta} = 0.00115$ respectivement.

4.3 Etats fondamentaux des modèles de verre de spins sur réseaux aléatoires.

En collaboration avec O. Dubois, J. Mandler, R. Monasson (Publication A16)

Une variante du problème SAT, appelée XOR-SAT où l’on remplace les OU entre variables par des OU exclusifs, est formellement équivalente à un modèle connu en physique statistique des systèmes désordonnés, le modèle dit à p-spins. Récemment, en collaboration avec R. Monasson et O. Dubois et J. Mandler, tous deux informaticiens théoriciens au Laboratoire d’Informatique de Paris VI, j’ai étudié ce modèle de verre de spins sur graphe aléatoire. Plus exactement, les spins sont placés aux sommets de plaquettes triangulaires, attachées les unes aux autres de manière aléatoire. Ce modèle peut être étudié par la méthode des répliques et admet une solution à un pas de brisure à la Parisi. La phase de basse température est caractérisée par une brisure d’ergodicité et par la division de l’espace des phases en états purs. Ces états purs, intimement reliés à la notion de brisure de symétrie des répliques, ne sont encore pas totalement compris.

Nous avons montré comment ces résultats pouvaient être retrouvés exactement par des techniques rigoureuses. Nous avons proposé une méthode explicite qui permet de construire les états purs du modèle pour un échantillon donné du désordre. Il est alors facile de compter ces états, de calculer leur énergie libre, leurs distances mutuelles... Notre approche est détaillée dans la publication 22. Je me contente donc d’en évoquer le principe sous-jacent.

L'idée essentielle est de regarder l'effet d'une procédure de décimation partielle, judicieusement choisie. On décime les spins un par un et on suit le flot du graphe des interactions entre spins au cours de cette décimation. Ceci peut être fait de manière rigoureuse par des techniques d'analyse d'algorithmes similaires à l'analyse de DP en l'absence de backtracking. Puis on utilise des outils de théorie de la probabilité (méthode des premier et deuxième moments) pour analyser le graphe simplifié restant à la fin de la décimation partielle.

Il est clair que cette approche a été rendue possible par la simplicité analytique relative du modèle considéré. Cependant, on peut espérer qu'elle est assez générale et pourrait s'appliquer à d'autres modèles de spins sur graphes aléatoires.

Chapitre 5

Travaux en cours et projets

5.1 Reconstruction de séquences d'ADN à partir de données de micromanipulation

Notre travail sur l'ouverture de l'ARN et ADN nous a permis de vérifier que le modèle dynamique d'ouverture défini par les taux d'ouverture et fermeture de l'équation 2.5 reproduisait bien les données expérimentales, comme le signal d'extension en fonction du temps, pour une séquence de molécule donnée (Fig. 2.12). Il est naturel de se demander si l'on peut, inversement, extraire des informations sur la séquence à partir des données expérimentales. Les cas à envisager sont celui de l'ouverture à force constante et celui de l'ouverture à vitesse d'éloignement des extrémités constante. Nous avons commencé par l'analyse du cas à force constante qui s'avère plus simple.

Considérons d'abord le cas idéal où le dispositif expérimental n'est affecté par aucun bruit si ce n'est les fluctuations thermiques et l'on dispose d'une résolution parfaite en temps et espace. Un résultat typique d'une telle expérience idéale d'étirement à force constante est schématisé en Figure 5.1. La molécule s'ouvre ou se referme en passant un certain temps sur chaque base avant que la frontière entre régions ouverte et fermée ne se déplace d'une base vers l'avant ou l'arrière. Toutes les transitions sont détectées. Une représentation commode des données consiste à compter le nombre d'intervalles élémentaires $t_i = n_i dt$ passés sur chaque base i , les nombres u_i et d_i de transitions $i \rightarrow i + 1$ et $i \rightarrow i - 1$ vers les bases suivante et précédente respectivement. Même en absence de bruit dû au dispositif expérimental, le signal est stochastique à cause du bruit thermique. Deux répétitions de l'expérience ne donnent pas le même signal. Si l'on définit l'ensemble des données d'un signal expérimental comme $\mathbf{x} = \{t_i, u_i, d\}$ ou $d = \sum_i d_i$, on peut se demander quelle est la probabilité d'avoir une séquence S produisant ce signal. Cette probabilité peut être écrite, selon les idées de l'inférence Bayésienne comme

$$P(S/\mathbf{x}) = \frac{P(\mathbf{x}/S) P(S)}{P(\mathbf{x})} \quad (5.1)$$

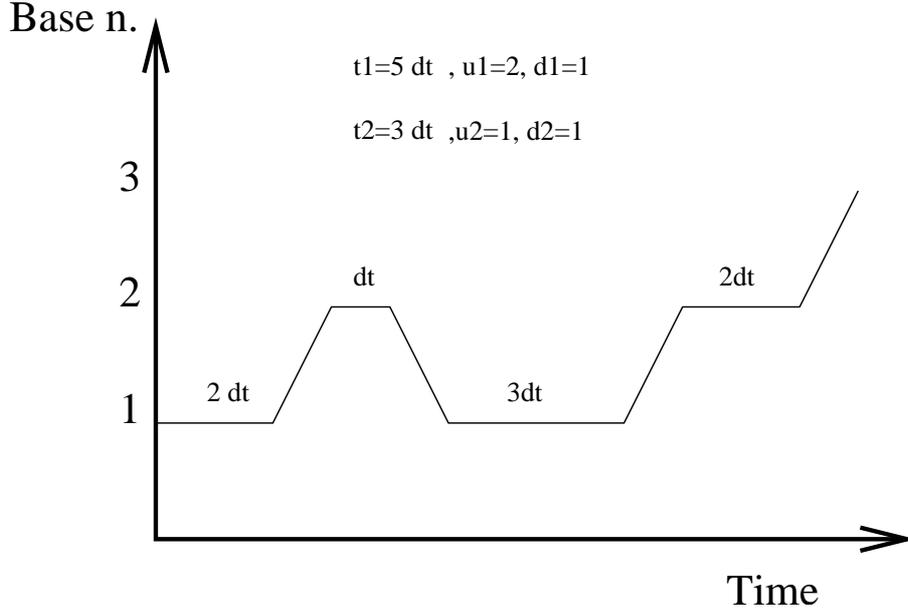


FIG. 5.1 – Transitions dynamiques dans l'ouverture à force constante. Au cours du temps, le nombre de bases ouvertes, i , reste constant ou varie de plus ou moins une unité lors de sauts.

Donc on peut faire une prédiction sur la séquence en cherchant à maximiser cette probabilité parmi toutes les S possibles. Si l'on considère une distribution a priori des séquences $P(S)$ uniforme, la maximisation de $P(S/\mathbf{x})$ se réduit à celle de $P(\mathbf{x}/S)$. $P(\mathbf{x}/S)$ est définie à partir des taux élémentaires de (2.5) comme

$$\begin{aligned}
 P(\mathbf{x}/S) &= \prod_i \left\{ 1 - r dt [\exp(-g_0(s_i, s_{i+1})) + \exp(2\mathcal{F}_{ss}(f))] \right\}^{t_i/dt} \\
 &\times \prod_i \left\{ r dt \exp(-g_0(s_i, s_{i+1})) \right\}^{u_i} \times \left\{ r dt \exp(2\mathcal{F}_{ss}(f)) \right\}^d
 \end{aligned} \quad (5.2)$$

et peut s'écrire comme un produit de matrices de transfert

$$P(\mathbf{x}/S) = \prod_i M(s_i, s_{i+1}; u_i, t_i) \quad (5.3)$$

Pour trouver la séquence qui maximise (5.2), on fait appel à l'algorithme de Viterbi, qui permet de trouver le maximum en partant de la première base et en choisissant la valeur optimale de cette base pour chaque valeur de la deuxième base; ainsi on affecte une probabilité à chaque valeur de la deuxième base comme

$$P_2(s_2) = \max_{s_1} M(s_1, s_2) \quad (5.4)$$

Puis on optimise sur la deuxième base en obtenant

$$P_3(s_3) = \max_{s_2} (P_2(s_2) M(s_2, s_3)) \quad (5.5)$$

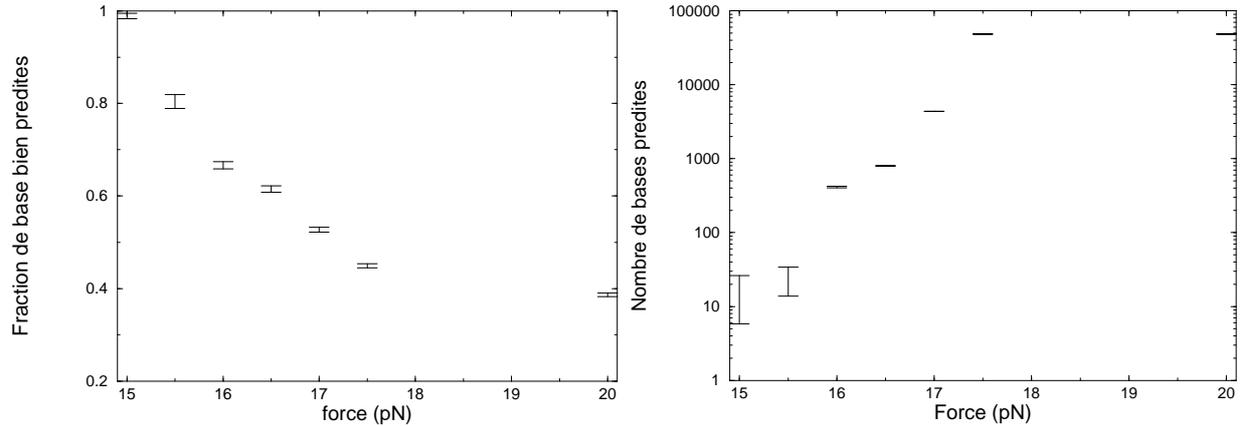


FIG. 5.2 – Gauche : fraction de bases prédites correctement, ω , en fonction de la force. Droite : nombre de bases prédites en fonction de la force.

et ainsi de suite,

$$P_{j+1}(s_{j+1}) = \max_{s_j} (P_j(s_j) M(s_j, s_{j+1})) \quad (5.6)$$

jusqu'à arriver à la dernière base N de la séquence. En choisissant la valeur s_N qui optimise $P_N(s_N)$, on sélectionne toute la séquence optimale des bases précédentes. Pour mettre au point cette procédure d'inférence Bayésienne et avoir une idée des performances dans le cas idéal, nous avons généré les données à partir de la séquence $\{s_i^L\}$ du λ -phage, à l'aide d'une procédure de type Monte Carlo se fondant sur le modèle dynamique (2.5). Une fois les données "expérimentales" produites, un deuxième programme qui ignore la séquence du phage et basé sur l'algorithme de Viterbi permet d'effectuer une prédiction sur la séquence, $\{s_i^V\}$. Une mesure a posteriori (c'est-à-dire en connaissant la séquence S^L) de la validité de la prédiction est le pourcentage de bases prédites correctement,

$$\omega = \frac{1}{N} \sum_{i=1}^N \delta(s_i^V - s_i^L) \quad (5.7)$$

La valeur moyenne de ω sur 100 prédictions est montrée en Figure 5.2 (gauche) avec sa variance; ω décroît avec la force parce que le nombre de base ouvertes (voir Figure 5.2 droite), et donc prédites, augmente et le temps que la fourche d'ouverture passe sur chaque base diminue. Pour une force de 16 pN, 67% des bases prédites le sont correctement.

On peut aussi obtenir une estimation a posteriori de la précision de la prédiction de la base i à l'aide de l'indicateur

$$\omega_i = \delta(s_i^V - s_i^L) \quad (5.8)$$

Cette précision, en raison du paysage d'énergie libre constitué de minima et de barrières, est fortement dépendante de la séquence. Dans la Figure 5.4 (courbe en tirets), je montre ω_i pour les 300 premières bases du λ phage à une force de 16 pN. La comparaison avec l'énergie libre du dégraissage $G(f = 16pN, n)$ des n premières bases (eqn.2.1) montre que

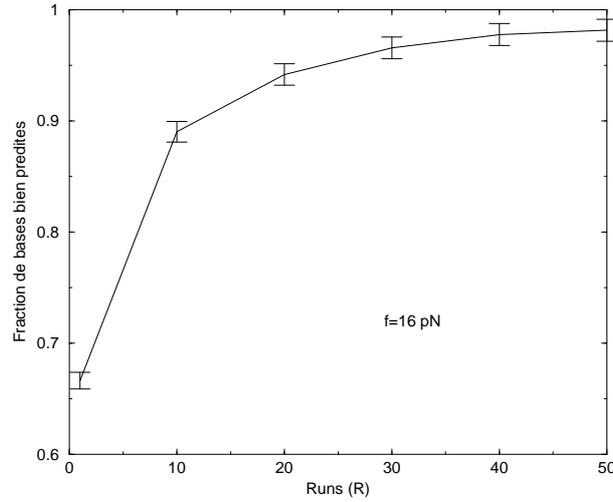


FIG. 5.3 – Fraction de bases prédites correctement ω en fonction du nombre de répétition de l’expérience, (runs), pour une force $f = 16 \text{ pN}$. ω augmente de la valeur de 0.66 à 0.9 pour R de 1 à 10 et atteint la valeur de 0.98 pour R=50.

les bases prédites le mieux sont celles qui correspondent à des minima locaux et des plats du paysage de l’énergie libre (Figure 5.4 milieu). Par contre, il semble difficile d’établir une corrélation directe entre une bonne prédiction et la composition de base GC par rapport aux bases AT dans la séquence (Figure 5.4 bas).

Pour avoir une estimation a priori (sans connaître la vraie séquence) de la justesse de la prédiction sur la base i , on peut calculer avec l’algorithme de Viterbi la probabilité $P^V(s_i)$ de chacun des quatre choix possibles ($s_i = \text{A, T, G, ou C}$) de la base i et évaluer la précision sur cette base comme

$$1 - \sigma_i = 1 - \sum_{s_i} P^V(s_i) \log_4 P^V(s_i) \quad (5.9)$$

L’allure de σ_i suit en effet celle de ω_i le long de la séquence.

On peut par la suite se demander comment améliorer la prédiction, en jouant sur la force, les répétitions de l’expériences (runs) sur la même molécule et sur la durée d’une expérience. Dans notre expérience idéale, nous avons ainsi pu comprendre qu’un paramètre important est le nombre de répétitions de l’expérience. En effet, en ouvrant et refermant plusieurs fois la molécule, on est certain de repasser plusieurs fois sur la même partie de la séquence ce qui permet d’accumuler des informations sur les temps de passage et de transitions sur chaque base, informations qui peuvent être ensuite traitées ensemble par l’algorithme de Viterbi. On voit sur la Figure 5.3 que le taux de prédictions correctes moyen, $\langle \omega \rangle$, s’améliore grandement avec le nombre de répétitions de l’expérience. La Figure 5.4 (haut) montre ω_i à 16 pN pour 1, 20 et 40 répétitions de l’expérience.

A coté de ces expériences numériques nous avons aussi commencé un traitement analytique. Dans ce but, nous considérons d’abord un cas simple, celui où la force appliquée est très grande. Une estimation rapide montre que pour des force égales ou supérieures

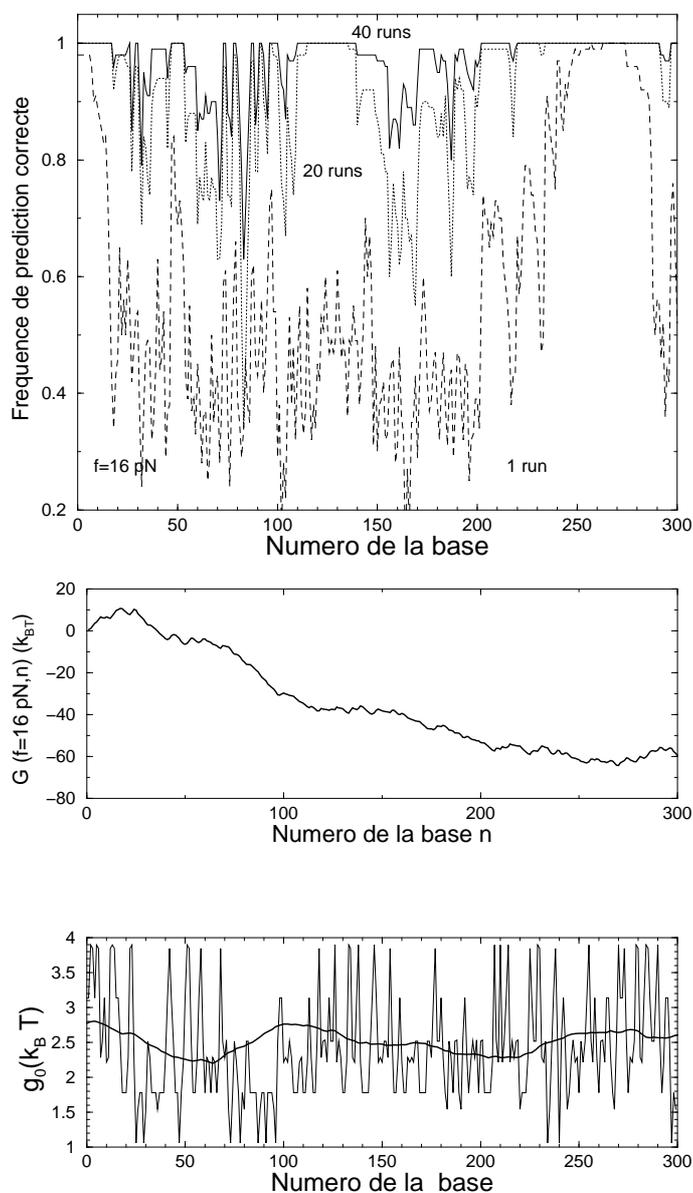


FIG. 5.4 – Haut : fréquence de prédiction correcte en fonction du numéro de la base pour une force de 16 pN et 1 run (ligne en tirets), 20 runs (ligne en pointillés), 40 runs (ligne pleine). Milieu : paysage d'énergie libre de dégraissage à une force $f = 16 \text{ pN}$; les minima locaux dans ce paysage correspondent à la partie de la séquence que l'on prédit le mieux. Bas : énergie libre d'appariement en fonction du numéro de la paire de bases, sans moyenne et moyennée avec un poids gaussien sur 20 bp. A l'oeil, on ne voit pas de corrélation directe entre la richesse en bases GC ou AT et l'exactitude de la prédiction.

à 40 pN, la frontière entre régions ouverte et fermée ne revient jamais en arrière lors de l'ouverture de la molécule. La matrice $M(s_i, s_{i+1}; u_i, t_i)$ se simplifie et ne dépend plus de u_i . En outre, si l'on élimine la possibilité de revenir en arrière, les moyennes sur les t_i (dues au bruit thermique) se decorrèlent. On peut aussi considérer deux types de bases seulement au lieu de quatre afin de simplifier quelque peu les calculs.

Pour commencer, nous avons calculé $\langle \omega \rangle$ en fonction du nombre de runs dans l'approximation où il n'y a pas d'interaction d'empilement. Il y a alors seulement deux valeurs de l'énergie libre $g(s_i)$ et leur différence est appelée $\Delta = g_- - g_+$. Un premier résultat théorique est que, à grande force, l'algorithme de Viterbi a tendance à prédire des bases de type faible (g_-). Nous avons étudié dans ce cas comment la valeur de $\langle \omega \rangle$ varie avec la différence Δ entre les deux énergies. Les résultats montrés en Figure 5.5 (haut) sont en parfait accord avec les résultats numériques.

Si l'on veut, par exemple, seulement distinguer dans la séquence les bases de type A ou T des bases de type G ou C, la valeur de Δ est environ 2.8. Si au contraire on veut parvenir à une résolution plus fine permettant de distinguer les bases A des bases T (ou les bases G des bases C), $\Delta \simeq 0.5$ seulement. Nous avons calculé analytiquement, en fonction du Δ , le nombre de runs typique permettant de parvenir à une précision tendant à 1, voir Figure 5.6.

Par la suite, nous avons considéré deux types de bases avec empilement, ce qui nécessite de prendre en compte trois valeurs de l'énergie d'appariement données en Figure 5.5 (bas). Nous avons effectué le calcul analytique des probabilités de prédire correctement une base faible ou une base forte, $\langle \omega^- \rangle$, $\langle \omega^+ \rangle$, dans le cas d'une séquence homogène de bases faibles ou fortes. Nous avons étudié comment cette probabilité varie en fonction des répétitions de l'algorithme (runs). Les résultats montrés en Figure 5.5(b) sont en parfait accord avec les résultats numériques.

Pour aller plus loin dans la théorie, nous voulons considérer d'abord le cas de valeurs quelconques des énergies libres de pairing et non plus le cas particulier (mais important car proche des valeurs réelles issues des données biochimiques) de la Figure 5.5b. Par la suite, nous voudrions calculer la valeur de ω pour une séquence quelconque et non pas uniforme. Pour ce, nous pensons d'abord calculer la longueur de corrélation de ω_i (à partir de la fonction à deux points) le long de la séquence, pour se ramener à une moyenne sur un nombre fini de bases autour de la base que l'on considère. Il serait très intéressant, mais semble pour l'instant difficile, de parvenir à une compréhension analytique du cas d'une force finie.

Du point de vue des simulations numériques, nous sommes en train d'étudier comment étendre l'algorithme de Viterbi au cas d'une résolution temporelle finie. Il faut prendre en compte le fait que l'on ne voit pas toutes les transitions. Donc, pendant l'intervalle de temps entre deux acquisitions de données successives, la fourche d'ouverture peut effectuer des sauts de plusieurs bases.

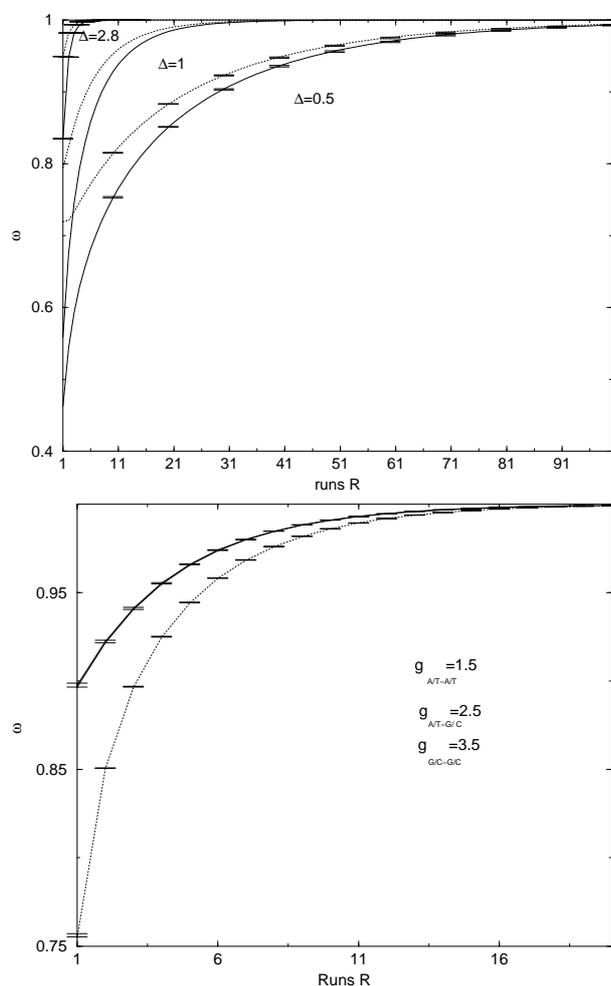


FIG. 5.5 – Haut : valeur moyenne du pourcentage de bases faibles ou fortes $\langle \omega^+ \rangle$ (ligne pleine) et $\langle \omega^- \rangle$ (ligne pointillée) prédites correctement pour une expérience à force élevée et dans le cas simplifié de deux bases sans empilement en fonction du nombre de répétitions de l'expérience et pour différentes valeurs de Δ . Les résultats des simulations sont montrés pour $\Delta = 0.5, 2.8$ avec les barres d'erreurs. Bas : $\langle \omega^+ \rangle$, $\langle \omega^- \rangle$ (pourcentage de bases prédites correctement pour une séquence homogène de bases fortes, faibles respectivement) pour une expérience à force élevée, deux types de bases avec empilement. Ces résultats sont à comparer avec le cas sans empilement et $\Delta = 1$. Les résultats des simulations sont montrés avec les barres d'erreurs.

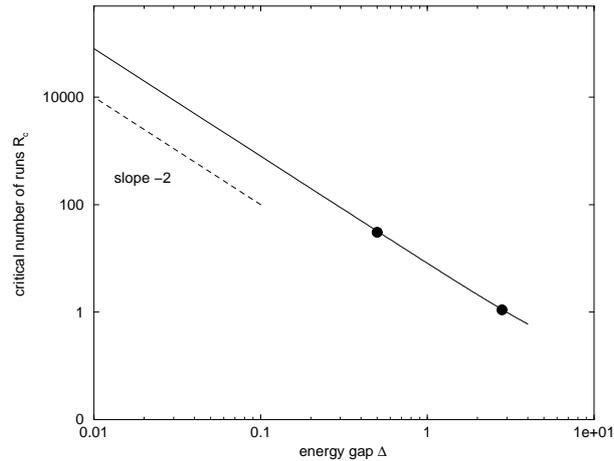


FIG. 5.6 – Nombre de répétitions de l’expérience permettant d’avoir un nombre de bases correctement prédites tendant à un, en fonction de la différence Δ d’énergie libre (cas de force élevée et sans empilement, avec deux types de bases seulement).

5.2 Formation de boucles dans l’ADN

En collaboration avec N. Douarche (étudiant en thèse)

La formation de boucles est un mécanisme très important dans l’organisation structurale et fonctionnelle de l’ADN. En effet, la formation de boucles entre en jeu dans l’empaquetage, la réparation et l’initiation ou la répression de la transcription [48]. Dès les années 70, la probabilité que l’ADN forme une boucle fut mesurée à l’aide d’expériences dites de cyclisation. Il s’agit de faire interagir en solution des molécules d’ADN avec des protéines qui joignent les deux extrémités de la molécule de façon permanente quand elles sont au contact. Ensuite, on lave les protéines et on mesure par électrophorèse le nombre d’ADN circulaires par rapport au nombre d’ADN qui sont restés lineaires, ce qui définit la probabilité de cyclisation. Une expérience de cyclisation sur des petites molécules d’ADN a été faite assez récemment par Widom et collaborateurs [49].

A la suite du développement des expériences sur molécule unique, la formation de boucle a été étudiée sur un ADN isolé, tout d’abord par Finzi et Bustamante [15] dans le cadre de l’action de la protéine Lac répresseur sur un ADN de λ phage. Les molécules d’ADN λ sont d’abord manipulées génétiquement pour insérer les deux séquences qui interagissent avec la protéine (opérons) à une distance de 150 paires de bases l’une de l’autre. Ensuite, elles sont attachées à une lamelle de verre d’un côté et aux billes de l’autre; la hauteur d’une bille à laquelle est attachée un ADN est mesurée à l’aide de pinces optiques. Quand on introduit les protéines dans la solution, on observe que la molécule, alternativement, se raccourcit (formation d’une boucle) et s’allonge (détachement de la protéine). Ceci permet de visualiser directement la dynamique de formation et rupture de boucles. Une expérience similaire a été faite par Lia en collaboration avec Allemand, Bensimon, Croquette et Finzi

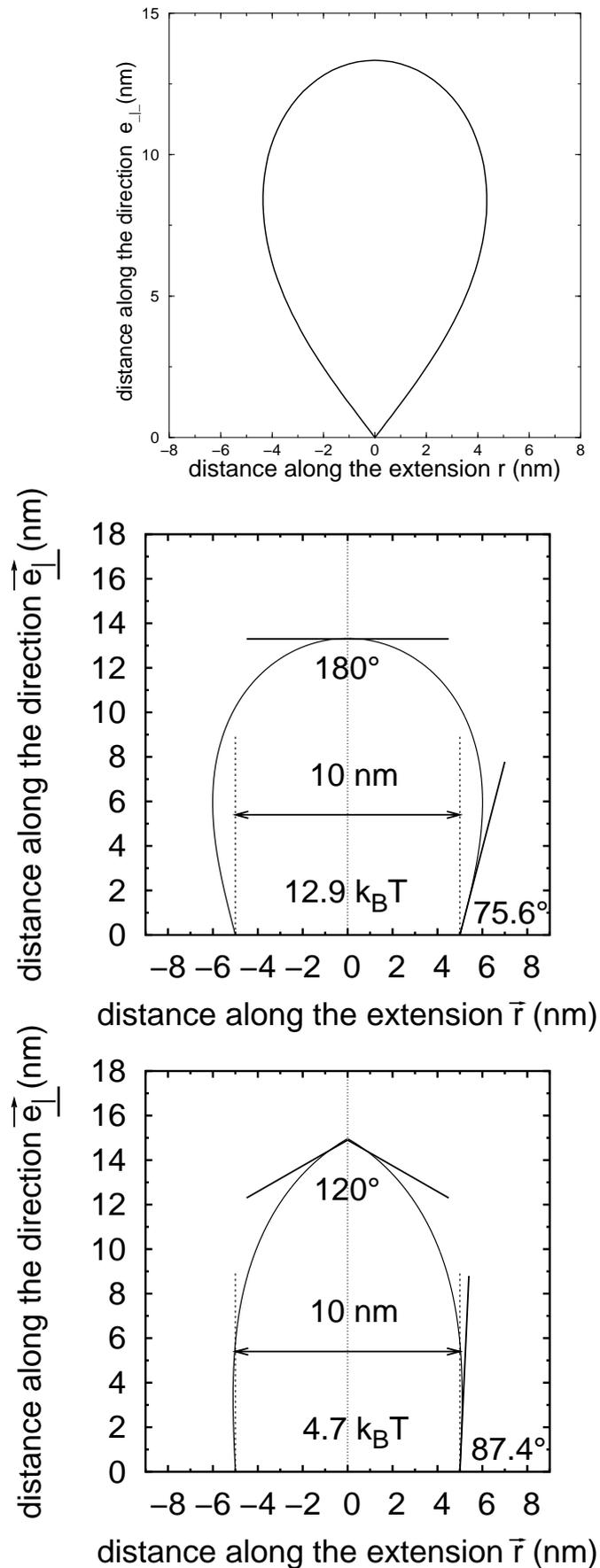


FIG. 5.7 – Configuration optimale pour une 'boucle' d'une séquence de 100 paires de bases fermée (haut), avec une distance entre les extrémités de 10 nm sans pliement (milieu) et avec un angle de 120 degrés entre les deux bases centrales (bas).

[50] pour étudier la formation de boucle par une protéine similaire, le Gal répresseur. Les résultats conduisent dans ce cas à supposer que dans ce processus de formation d'une boucle, une autre protéine (Cro) intervient afin de créer une bulle de dénaturation au milieu de la partie de la séquence qui doit se replier. Ceci favorise la création de la boucle parce que l'ADN simple brin est plus flexible que l'ADN double brin. En outre, la dynamique de formation de la boucle a été également étudiée en présence d'une force d'étirement.

Dans notre travail, nous avons calculé la probabilité de formation d'une boucle à partir du modèle classique du ver (WLC) (voir article A2) pour l'ADN. Cette quantité est très difficile à calculer parce qu'elle suppose que l'on sache évaluer une intégrale de chemin dont l'action est l'énergie de courbure du polymère dont, ce qui introduit une contrainte globale, les extrémités se confondent. Aucune résolution analytique exacte n'est disponible mais ce problème a été très étudié jusqu'ici et plusieurs approches d'approximation ont été proposées [51, 52].

Les origines de la difficulté qu'a l'ADN à faire une boucle sont multiples et varient avec la longueur de la molécule. Pour de longues molécules (qui font plus que dix longueurs de persistance environ), la mise en contact de deux bouts de la molécule n'est pas difficile du point de vue de l'énergie de courbure, mais à cause de la perte d'entropie consécutive à la baisse du nombre de configurations accessibles une fois que la boucle est formée. La probabilité de faire une boucle peut être calculée simplement avec le modèle de polymère gaussien,

$$P(r = 0) = \frac{1}{(2\pi L b)^{3/2}} \quad (5.10)$$

Pour de petites molécules (jusqu'à quelques longueurs de persistance), la molécule est rigide et le coût énergétique nécessaire à la plier pour mettre en contact ses deux extrémités est élevé. Si l'on considère une configuration bouclée optimale, en forme de cercle, l'énergie de courbure de cette configuration est simplement,

$$E/k_B T = A L / (R^2) \quad (5.11)$$

où $R = L/(2\pi)$ est le rayon du cercle et donc

$$P(r = 0) \propto e^{-A/(4\pi^2 L^2)} \quad (5.12)$$

La vraie configuration optimale n'est en fait pas un cercle si, comme c'est le cas ici, les vecteurs tangents de la première et de la dernière bases ne doivent pas être alignés dans la même direction. Cette configuration optimale, et son énergie, peuvent être calculées dans la limite du continuum puisqu'elles sont reliées à la solution de l'équation de Mathieu pour le mouvement du pendule. La configuration optimale pour une molécule de 100 paires de bases est montrée en Figure 5.7 (haut). Les fluctuations entropiques autour de la configuration optimale doivent être aussi englobées dans les calculs pour avoir une estimation plus précise. La probabilité de cyclisation tend donc à zéro pour de petites longueurs et de grandes longueurs, et elle présente un maximum entre les deux. Ce maximum se situe pour une longueur de la molécule égale à 3.5 longueurs de persistance, voir Figure 5.8.

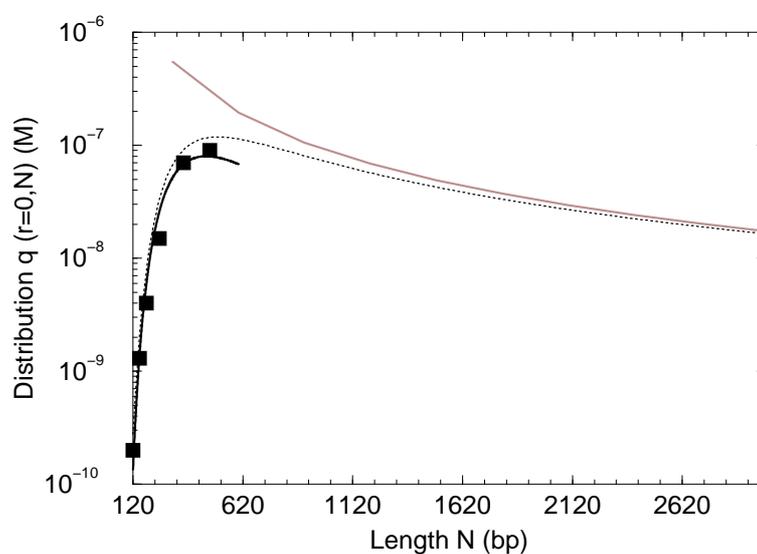


FIG. 5.8 – Facteur de cyclisation pour $r = 0$ en fonction de la longueur de la molécule, en nombre de paires de bases, et un angle de pliage de 180 degrés. Ligne en trait plein : approximation de point col, carrés : simulations de dynamique moléculaire par Langowski, ligne en pointillés : résolution numérique, en gris : approximation gaussienne.

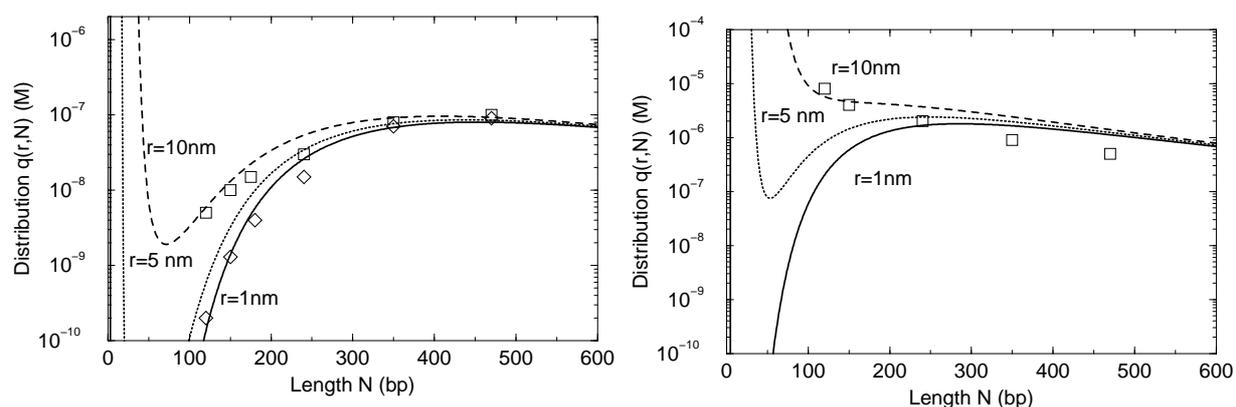


FIG. 5.9 – Facteur de cyclisation pour $r = 1, 5, 10$ nm et un angle de pliage de 180 (gauche), 120 (droite) degrés. Lignes : approximation de point col, symboles : simulations effectuées par Langowski.

Pour calculer le facteur de cyclisation, nous avons utilisé deux approches. D'une part, nous avons calculé numériquement l'intégrale de transfert qui donne cette probabilité et, d'autre part, nous avons calculé pour de petites molécules la configuration optimale, l'énergie de courbure correspondantes et les fluctuations quadratiques autour de cette configuration.

Nous n'avons pas considéré seulement le cas de deux extrémités se refermant parfaitement mais avons aussi calculé la probabilité de cyclisation en tenant compte des dimensions de la protéine. La probabilité de cyclisation augmente sensiblement si l'on considère que l'ADN se referme quand les extrémités se trouvent à une distance inférieure aux dimensions linéaires de la protéine. Pour s'en rendre compte, il suffit de considérer que 100 bases font 34 nm en longueur et la dimension linéaire d'une protéine est d'environ 10 nm (voir Figure 5.7). Comme il est montré dans la Figure 5.9 le facteur de cyclisation augmente, par exemple, de 10^{-9} à 10^{-8} pour une molécule de 150 paire de bases si on considère une dimension de fermeture de $r = 10$ nm par rapport à $r = 1$ nm.

Ensuite, nous avons pris en compte l'autre facteur qui réduit nettement la probabilité de cyclisation : le pliement de la séquence en son milieu. Nous avons donc introduit un angle de courbure, sans coût énergétique, de 120 degrés entre les deux bases à mi-séquence (voir Figure 5.7), et calculé le facteur de cyclisation dans ces conditions. Il augmente, dans les deux cas $r = 1$ nm, $r = 10$ nm à 10^{-7} et 10^{-5} respectivement si on introduit un angle de courbure de 120 degrés entre les deux bases à mi-séquence. Pour finir, nous sommes en train de calculer le facteur de cyclisation lorsque l'on tire les extrémités de la molécule avec une force. Introduire une force est intéressant non seulement dans le contexte général des expériences de micromanipulation mais aussi parce que, habituellement, l'ADN dans la cellule n'est pas libre mais attaché à des protéines d'empaquettement qui exercent des forces variées.

5.3 Distribution de fluorescence dans une population de bactéries monoclonale.

Etude de l'évolution statistique d'une population bactérienne monoclonale, en collaboration avec M. Castelnovo R. Monasson, K. Sekimoto (LDFC) (Projet ACI DRAB commencé en 2004).

Sous la direction de Didier Chatenay, Michael Poirier et Christian Rick ont mis au point un système expérimental pour étudier la distribution de l'expression d'une protéine fluorescente (GFP) dont le gène est situé sur des plasmides (ADN circulaire ne faisant pas partie du chromosome) dans une population monoclonale de la bactérie *Escherichia Coli* (*E. Coli*). La distribution de fluorescence est liée aux fluctuations de plusieurs processus biochimiques. D'abord, le nombre de plasmides peut varier de bactérie à bactérie. Les fluctuations de ce nombre sont liées au bruit dans la réplication, dans la durée du cycle cellulaire qui n'est pas forcément synchronisée avec la réplication des plasmides, et à la répartition des plasmides dans les deux cellules filles pendant la division cellulaire. En

outre, le gène codant la protéine fluorescente sur chaque plasmide doit être exprimé, ce qui fait intervenir d'autres sources de bruit liées aux phases de transcription et traduction.

Notre effort théorique portera d'abord sur la conception et l'analyse de modèles pour la dynamique stochastique de réplication des plasmides et de la division cellulaire de la bactérie. On tiendra compte aussi des paramètres de contrôle du système comme par exemple la concentration d'antibiotique. Nous voulons commencer par comprendre comment les taux de réplication des plasmides doivent dépendre du nombre C de plasmides (via les mécanismes de contrôle de la réplication) pour permettre d'atteindre l'équilibre dans la distribution du nombre de plasmides de génération en génération, en présence de bruit dans la réplication, la répartition et la durée du cycle cellulaire. Notre objectif sera ensuite, d'une part, de chercher à coupler l'équation d'évolution pour C avec l'analyse du réseau biochimique afin d'obtenir une théorie complète des fluctuations d'expression de la GFP pour des plasmides ayant, en moyenne, une ou plusieurs copies et, d'autre part, de confronter les prédictions obtenues à partir de ce modèle aux résultats expérimentaux. La résolution du modèle fera appel à des techniques analytiques (formalisme de Fokker-Planck) permettant de décrire la distribution stationnaire des espèces moléculaires et numériques (résolution approchée d'équations analytiques, simulations numériques). Parmi les quantités que l'on cherchera à calculer, mentionnons :

- la distribution $P(C)$ des nombres C de copies du plasmide ;
- la distribution $Q(I/C)$ de l'intensité de fluorescence à nombre C de copies fixées ;
- la distribution $R(I/I_{\text{chromosome}})$ de l'intensité de fluorescence de la protéine dont le gène est localisé sur les plasmides par rapport à l'intensité de celle dont le gène est placé sur le chromosome. Cette quantité permettra d'éliminer certains bruits extrinsèques dus à l'environnement cellulaire et influençant l'expression des protéines.

Ces quantités seront confrontées directement avec les expériences. La convolution de P et Q donne la distribution globale de l'intensité de fluorescence. La comparaison avec les données permettra de fixer les valeurs de paramètres (taux de réactions) figurant dans la description du réseau biochimique.

L'étape suivante consistera à étudier les effets non stationnaires affectant les distributions précédentes. Il s'agira tout d'abord de calculer le nombre de générations nécessaires à la relaxation de ces distributions vers leurs valeurs stationnaires lors d'une perturbation faible (variation d'un paramètre extérieur) ou importante (par exemple la sélection d'un individu avec un phénotype anormal). Les techniques utilisées seront d'une part l'analyse de stabilité linéaire de notre modèle de réseau dans le cas de faible perturbation, et d'autre part des méthodes essentiellement numériques dans le cas de fortes perturbations.

Un dernier point que nous souhaitons étudier est celui de la pression de sélection exercée dans notre système expérimental par l'antibiotique. Alors que trop d'antibiotique empêche la croissance, une faible concentration ne semble pas trop affecter l'expression de GFP. Une hypothèse spéculative est que la présence massive d'antibiotique, juste en dessous du seuil où la population cesse de croître, induit une pression d'évolution forte sur la population et sélectionne les bactéries avec un nombre de plasmides élevés ; dans ce cas il y aurait réduction des fluctuations, un phénomène susceptible d'être modélisé théoriquement.

Chapitre 6

Autre activités

6.1 Encadrement d'étudiants

2002

Encadrement, avec C. Marques, d'Emmanuel Collé en Stage de DEA .
Sujet : Empilement de cyclodextrines sur des tubes de polymères.

Encadrement de Jules Grucker et Alan Ott en stage de licence.
Sujet : les mécanismes d'ouverture de l'ADN.

2003

Encadrement, avec D. Chatenay, de Nicolas Douarche en thèse de doctorat.
Sujet : Modèles d'élasticité de l'ADN et de la formation de boucles.

Encadrement, avec R. Monasson, de Philippe Beckrich en stage de DEA.
Sujet : Modèles dynamiques de l'activité spontanée de réseaux de neurones.

2004

Encadrement, avec E. Marinari (Université de Rome, La Sapienza) et R. Monasson, de V. Baldazzi en thèse de doctorat.
Sujet : Reconstitution de séquence partir de données de micromanipulation sur l'ADN.

6.2 Organisation de congrès et d'écoles.

Organisation avec P. Bassereau et J.J Benattar du Minicolloque Physique et systèmes biologiques : de la molécule à la cellule ; 8ièmes Journées de la Matière Condensée , Société Française de Physique, Marseille du 27 au 30 août 2002.

Organisation avec D. Chatenay, O. Krichevsky, R. Monasson, D. Thieffry de l'école d'été des Houches, sur le thème "Aspects multiples de l'ADN et l'ARN :de la bioinformatique à

la biophysique ” du 2 au 27 août 2004.

6.3 Diffusion de l’information scientifique.

Publication d’articles dans Physics World (Pub. A1), Pour la Science (Pub. A11) et Images de la Physique (Pub. A12).

Traduction, en collaboration avec R. Monasson, du livre de Piergiorgio Odifreddi : *Les mathématiques à l’aube du XXIe siècle*, éditions Belin (2004).

6.4 Allocations de recherche.

2002 ACI Jeunes chercheurs sur le thème : Décoder et coder les chemins de repliement de l’ARN. Equipe : H. Isambert (chef du projet), S. Cocco, F. Thalmann, A. Xayaphoummine.

2004 ACI Nouvelles Interfaces des mathématiques sur le thème : le paradigme XORSAT, de la complexité aux transitions de phases. Equipe constituée d’informaticiens, mathématiciens et physiciens statisticiens dirigée par H. Daudé (Marseille).

2004 ACI Dynamique et réactivité des objets biologiques sur le thème : Etude des variations phénotypiques dans une population bactérienne monoclonale. Equipe : D. Chatenay (chef du projet), S. Cocco, V. Hakim, R. Monasson, J. Robert.

Chapitre 7

Liste des publications

- 1] S. Cocco, R. Monasson and R. Zecchina.
Analytical and Numerical Study of Internal Representations in Multilayer Neural Networks with Binary Weights.
Physical Review E, **54**, 717-725 (1996).
- 2] S. Cocco, R. Monasson, R. Zecchina.
The weight space structure of the parity machine with binary weights : analytical and numerical results.
Proceedings of the Varenna School of physics Enrico Fermi course CXXXIV, edited by F. Mallamace and H. E. Stanley (IOS Press, Amsterdam) 1997, 738-739.
- 3] S. Cocco, M. Barbi, M. Peyrard.
Vector Nonlinear Klein-Gordon Lattices : General Derivation of Small Amplitude Envelope Soliton Solutions.
Physics Letters A, **253** 161-167 (1999).
- 4] M. Barbi, S. Cocco, M. Peyrard.
Helicoidal model for DNA opening.
Physics Letters A, **253**, 358-369 (1999).
- 5] M. Barbi, S. Cocco, M. Peyrard, S. Ruffo.
A twist opening model for DNA.
Journal of Biological Physics, **24** 97-114 (1999).
- 6] S. Cocco, R. Monasson.
Statistical Mechanics of Torque Induced Denaturation of DNA.
Physical Review Letters, **83** 5178-5181 (1999).
- 7] S. Cocco, R. Monasson.
A Theoretical study of collective modes in DNA at ambient temperature.

Journal of Chemical Physics **112**, 10017 (2000).

8] S. Cocco, R. Monasson.

Trajectories in phase diagrams, growth processes and computational complexity : how search algorithms solve the 3-Satisfiability problem.

Physical Review Letters **86**, 1654 (2001) .

9] G. Biroli, S. Cocco, R. Monasson.

Le temps d'un choix : transitions de phase et complexité en informatique.

Images de la Physique 2001, CNRS Editions.

10] S. Cocco, R. Monasson.

Analysis of the computational complexity of solving random satisfiability problems using branch and bound search algorithms.

Eur. Phys. J. B **22**, 505 (2001)

11] S. Cocco, J.F. Marko, R. Monasson

Force and kinetic barriers to unzipping of the DNA double helix.

Proceedings of the National Academy of Science U S A. **98(15)**, 8608 (2001)

12] S. Cocco, J.F. Marko, R. Monasson

Force and kinetic barriers to initiation of DNA unzipping

Physical Review E, **65** 041907 (2002).

13] S. Cocco, O. Dubois, J. Mandler, R. Monasson

Au seuil de la complexité calculatoire

Pour la Science **295**, 52 (2002).

14] G. Biroli, S. Cocco, R. Monasson

Phase transitions and Complexity in computer science :

An overview of the statistical physics approach to the random satisfiability problem.

Proceedings of the StatPhys 2001 meeting. Physica A **306**, 381-394 (2002).

15] S. Cocco, R. Monasson

Restart method and exponential acceleration of random 3-sat instances resolutions : a large deviation analysis of the Davis-Putnam-Loveland-Logemann algorithm.

to appear in "Annals of Mathematics and Artificial Intelligence" (2004).

16] S. Cocco, R. Monasson

Heuristic average-case analysis of the backtrack resolution of random 3-Satisfiability instances,

Theoretical Computer Science **A320**, 345 (2004).

-
- 17] S. Cocco, R. Monasson
Exponentially hard problems are sometimes polynomial, a large deviation analysis of search algorithms for the random satisfiability problem, and its application to stop-and-restart resolution
Physical Review E, **66** 037101.(2002) .
- 18] S. Cocco, R. Monasson, J. Marko
Slow nucleic acid unzipping kinetics from sequence-defined barriers,
Eur. Phys. J. **10**, 153 (2003).
- 19] S. Cocco, R. Monasson, J. Marko
Theoretical models for single molecule experiments : from elasticity to unzipping
Comptes rendus de l' Académie des Sciences Physiques **3**, 569-584 (2002).
- 20] S. Cocco, R. Monasson, J. Marko
Unzipping Dynamics of Long DNAs
Physical Review E, **66** 051914 (2002) .
- 21] Simona Cocco, Rémi Monasson, Jie Yan, Abhijit Sarkar and John F. Marko
Force-extension behavior of folding polymers.
Eur. Phys. E **10**, 249 (2003).
- 22] S. Cocco, O. Dubois, J. Mandler, R. Monasson
Rigorous decimation-based construction of ground pure states for spin glass models on random lattices.
Phys. Rev. Lett. **90**, 047205 (2003).
- 23] S. Cocco, J.F. Marko
The micromechanics of DNA
Physics World **16**, 37 (2003).
- 24] S. Cocco, R. Monasson, A. Montanari, G. Semerjian.
Approximate analysis of search algorithms with “physical” methods.
Chapter for Phase Transition and Algorithmic Complexity edited by G. Istrate, C. Moore, A. Percus (2004).
- 25] S. Cocco, L. Ein-Dor, R. Monasson
Analysis of backtracking procedure for random decisions problems
Chapter for New optimisation algorithm in Physics edited by A. Hartmann, H. Rieger, Wiley, Springer-Verlag (2004).
- 26] S. Cocco, Y. Yan, J-F. Léger, D. Chatenay, J.F. Marko,
Overstretching and force-driven strand separation of double-helix DNA

Phys. Rev. **E70**, 011910 (2004).

27] C. Wyart, S. Cocco, J. F. Léger, C. Herr, L. Bourdieu, D. Chatenay
Dynamics of excitatory synaptic components in persistent bursting activity at low firing rates
Submitted to J. Neurophysiology (2004).

28] S. Cocco
Role of calcium and noise in the persistent activity of an isolated neuron
Submitted to Phys. Rev. Lett. (2004)

Bibliographie

- [1] S.B. Smith, L. Finzi, C. Bustamante, *Science* **258**, 1122 (1992).
- [2] P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J.L. Viovy, D. Chatenay, F. Caron, *Science* **271**, 792 (1996).
- [3] S. Smith, Y. Cui, C. Bustamante, *Science* **271**, 795 (1996).
- [4] C. Bustamante, J.F. Marko, E.D. Siggia, S. Smith, *Science* **265**, 1599 (1994).
- [5] C. Bustamante, S.B. Smith, J. Liphardt, D. Smith, *Curr. Op. Struct. Biol* **10**, 279 (2000).
- [6] M.N. Dessinges, B. Maier, Zhang, M. Peliti, D. Bensimon, V. Croquette, *Phys. Rev. Lett.* **89**, 248102 (2002).
- [7] J. Liphardt, B. Onoa, S.B. Smith, I. Tinoco JR, C. Bustamante, *Science* **292**, 733-737 (2001).
- [8] P. Thomen, U. Bockelmann, F.Heslot, *Phys. Rev. Lett.* **88**, 248102 (2002).
- [9] S. Harlepp, T. Marchal, J. Robert, JF Léger, A. Xayaphoummine A, H. Isambert, D. Chatenay, *Europhys. Journal E* **12**, 605 (2003).
- [10] T.R. Strick, J-F. Allemand, D. Bensimon, A. Bensimon, V. Croquette, *Science* **271**, 1835 (1996).
- [11] J-F. Léger, G. Romano, A. Sarkar, J. Robert, L. Bourdieu, D. Chatenay, J.F. Marko, *Phys. Rev. Lett.* **83**, 1066 (1999).
- [12] B. Maier, D. Bensimon, V. Croquette, *Proc. Natl. Acad. Sci. (USA)* **97**, 12002 (2000).
- [13] J.F. Marko, E.D. Siggia, *Macromolecules* **28**, 209 (1995).
- [14] Y. Cui, C. Bustamante, *Proc. Natl. Acad. Sci. (USA)* **97**, 127-132 (2000).
- [15] L.Finzi and J. Gelles *Science* **267**, 378 (1995).
- [16] J.F. Marko, E.D. Siggia, *Biophys. J.* **73**, 2173 (1997).
- [17] M. Rief, H. Clausen-Schaumann, H.E. Gaub, *Nat. Struct. Biol.* **6**, 346 (1999).
- [18] K. Breslauer, R. Frank, H. Blocker, L.A. Marky, *Proc. Natl. Acad. Sci. (USA)* **83**, 3746 (1986).
- [19] D.K. Lubensky, D.R. Nelson, *Phys. Rev. Lett.* **85**, 1572 (2000).

- [20] S. Cocco, R. Monasson, *Phys. Rev. Lett.* **83**, 5178 (1999).
- [21] A. Sarkar, J-F. Léger, D. Chatenay, J.F. Marko, *Phys. Rev. E* **63**, 051903 (2001).
- [22] M. Zuker, *Curr. Opin. Struct. Biol.* **10**, 303-310 (2000).
- [23] M. Peyrard, A.R. Bishop *Phys. Rev. Lett.* **62**, 2755 (1989).
- [24] D. Pörschke *J. Mol. Biol.* **62**,361-368 (1971).
- [25] G. Bonnet, O. Krichevsky, A. Libchaber *Proc.Natl. Acad. Sci. Usa***95**, 8602-8606.
- [26] T. Strunz, K. Oroszlan, R. Schäfer, H-J. Güntherhodt. *Proc. Nat. Acad. Sci. (USA)* **96**, 11277 (1999).
- [27] A. Montanari, M. Mzard *Phys. Rev. Lett.* **86**, 2178 (2001).
- [28] H. Orland, A. Zee *Nuclear Phys. B* **620** 456-476 (2002)
- [29] D. Moroz, P. Nelson *Proc. Nat. Acad. Sci. (USA)* **94**, 12295 (1998).
- [30] J. L. Barrat, J. F. Joanny, *Europhys. Lett.***24**,333 (1993).
- [31] D. Bartolo, I. Derenyi, A. Ajdari, *Phys. Rev. E* **65** 051910 (2002)
- [32] B. Essevaz-Roulet, U. Bockelmann, F. Heslot, *Proc. Natl. Acad. Sci. (USA)* **94**, 11935 (1997).
- [33] U. Bockelmann, B. Essevaz-Roulet, F. Heslot, *Phys. Rev. E* **58**, 2386 (1998).
- [34] U. Bockelmann, P. Thomen, B. Essevaz-Roulet, V. Viasnoff, F. Heslot, *Biophys. J.***82**, 1537 (2002).
- [35] M. Mañosas, F. Ritort Thermodynamic and kinetic aspects of RNA pulling experiments cond-mat/0405035 (2004)
- [36] C. Wyart, C. Ybert, L. Bourdieu, C. Herr, C. Printz, D. Chatenay, *J. Neurosci. Methods* **117**, 123 (2002).
- [37] X.J. Wang *J. Neurophysiol.***79**, 1549 (1998).
- [38] Y.H. Liu, X.J. Wang *J. Comp. Neuroscience***10**, 25 (2001).
- [39] A. Renart, N. Brunel X.J. Wang in *Computational Neuroscience : A Comprehensive Approach.*, J. Feng Ed., CRC Press, Boca Raton (2004).
- [40] D. Hebb, *The Organization of behavior* Wiley, New York (1949).
- [41] X.-J. Wang Neural Oscillations. in *Encyclopedia of Cognitive Science* MacMillan Reference Ltd (2003)
- [42] J. Rinzel, G.B. Ermentrout, In *:Methods in Neural Modelling*, edited by C.Koch and I. Segev. Cambridge, M.A. : MIT Press, (1989) p. 135-170.
- [43] A.L. Hodgkin, A.F. Huxley *J. Physiol. (Lond)* **117**, 500 (1952).
- [44] Hayes, B. *American Scientist* **85**(2), 108–112 (1996).
- [45] Davis, M. & Putnam, H. *J. Assoc. Comput. Mach.* **7**, 201–215 (1960).
- [46] Monasson, R., Zecchina, R., Kirkpatrick, S., Selman, B. & Troyansky, L. *Nature* **400**, 133–137 (1999).

-
- [47] Chao, M.T. & Franco, J. *Information Science* **51**, 289-314 (1990).
- [48] K. Rippe, P.H. von Hippel, J.Langowski *TIBS*, 500 (1995).
- [49] T. Cloutier, J. Widom *Molecular Cell* **14**, 355-362 (2004)
- [50] G. Lia, D. Bensimon, V. Croquette, J.-F. Allemand, D. Dunlap, D.E.A. Lewis, S. Adhya and L. Finzi, *PNAS USA* **100**, 11373 (2003).
- [51] H. E. Daniels *Proc. Roy. Soc. Edinburgh A* **63**, 290 (1952).
- [52] J.Shimada and H. Yamakawa *Macromolecules* **17**, 689 (1984).