



HAL
open science

Modélisation du chromosome bactérien en vue de la conception de réseaux biologiques de régulation dans l'espace cellulaire

Thibaut Lepage

► **To cite this version:**

Thibaut Lepage. Modélisation du chromosome bactérien en vue de la conception de réseaux biologiques de régulation dans l'espace cellulaire. Biophysique [physics.bio-ph]. Université d'Évry Val d'Essonne, 2015. Français. NNT: . tel-01111152

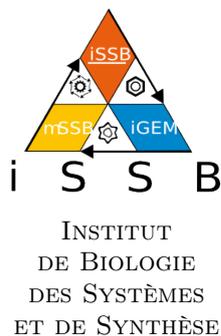
HAL Id: tel-01111152

<https://hal.science/tel-01111152>

Submitted on 29 Jan 2015

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

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



ÉCOLE DOCTORALE DES GÉNOMES AUX ORGANISMES

THÈSE

présentée pour obtenir le grade de
Docteur de l'Université d'Évry Val d'Essonne
spécialité Physique

par

Thibaut LEPAGE

**Modélisation du chromosome bactérien
en vue de la conception
de réseaux biologiques de régulation
dans l'espace cellulaire**

Soutenue publiquement le 12 janvier 2015
devant le jury composé de :

François KÉPÈS	Directeur de thèse
Ivan JUNIER	Co-encadrant
Vincent CROQUETTE	Rapporteur
Ralf EVERAERS	Rapporteur
Jean-Marc VICTOR	Examineur
Christophe LAVELLE	Examineur

Remerciements

J'É DOIS le succès de ma thèse à l'aide et au soutien de nombreuses personnes. En premier lieu, je veux remercier François et Ivan pour leurs précieux conseils, leurs explications salutaires, leur patience et leurs efforts pour m'initier sans douleur aux savoirs, aux questions et au jargon de la biologie. Il faut rendre grâce à la disponibilité d'Ivan, en dépit des montagnes qui nous séparent et des sautes d'humeur de Skype. Je ne serais pas allé très loin sans les longues discussions que nous avons pu avoir ensemble et les nombreuses idées qui en jaillissaient (généralement agrémentées de schémas d'autant plus incompréhensibles que l'idée en question était ingénieuse). Je suis également reconnaissant d'avoir pu, grâce à François, assister à plusieurs conférences en France et en Europe, y présenter parfois mes travaux, et y rencontrer des chercheurs de tous horizons, intéressés par les mêmes sujets que nous.

J'ai beaucoup apprécié de travailler au sein de l'iSSB, et particulièrement dans l'équipe MEGA : j'ai pu constater que, si une discussion entre des biologistes, des informaticiens, des physiciens, nécessite quelques efforts de la part de tous les interlocuteurs, les bénéfices que chacun tire de cet environnement pluridisciplinaire en valent la peine. Les conseils et les explications des membres de l'équipe, chacun abordant le vivant avec un regard différent, m'ont été très utiles. Merci en particulier à Costas et à Brian pour nos nombreuses discussions (pas toujours à propos de science).

Je tiens à remercier David Bensimon et Olivier Martin, qui ont accepté de faire partie de mon comité de thèse et qui m'ont grandement aidé à orienter mes recherches dans une direction fructueuse, ainsi qu'Henri Orland pour d'intéressantes discussions. Je tiens également à remercier les membres de mon jury : Vincent Croquette, Ralf Everaers, Jean-Marc Victor et Christophe Lavelle, pour leurs remarques pertinentes, leurs critiques avisées et leurs corrections.

J'ai grandement bénéficié du support de l'Institut des Systèmes Complexes, qui m'a permis d'obtenir une bourse d'allocataire « Domaine d'Intérêt Majeur » financée par la région Île-de-France, et m'a également fourni un accès à ses ressources de calcul. Celles de l'iSSB m'ont elles aussi été d'une grande aide, ainsi que l'assistance de Joan, à qui je présente mes excuses pour l'avoir trop souvent dérangé à cause d'« `ssh` qui marche plus ».

Merci à Sylvie et Bernadette pour leur gestion des questions administratives et logistiques (un service relativement peu coûteux en pâtisseries), ainsi que pour leur gentillesse et leur bonne humeur.

J'ai eu la chance d'être soutenu par une famille adorable sur qui j'ai régulièrement pu me décharger de ma frustration et de mon inquiétude, lorsque les bugs surgissant de tous les recoins de mon programme menaçaient d'avoir raison de ma santé mentale. Merci à mes parents, à mes sœurs et mes beaux-frères, et à la ribambelle de cousins, cousines, oncles, tantes, neveux et nièces, avec qui j'ai passé des moments toujours agréables, en Mayenne ou en région parisienne. J'ai une pensée souriante pour mon grand-père, qui n'a malheureusement pas pu voir s'achever ma thèse.

Je suis extrêmement reconnaissant envers Franck et toute la bande de la Fabryk, pour m'avoir permis durant trois ans d'oublier mes soucis, mes bugs et mon stress le temps d'un cours, d'un spectacle ou d'un anniversaire. Merci en particulier à la « team Bretagne » : Pauline, Caroline, Margot, Gaël, Laura, Akli et Isabel, pour les éclats de rire à répétition saupoudrés d'accent marseillais.

Merci à Florian et à tous les amis avec qui j'ai eu l'occasion de me détendre autour d'une bière.

Enfin, il me reste à rendre hommage à tous ceux qui ont contribué à la réussite de ma thèse, à des degrés divers : la boulangerie et le kebab du Bras de Fer, stackoverflow.com, Trivium, le service des urgences de l'Hôtel-Dieu, Robin Hobb, Ninkasi, Sid Meier et la SNCF. Ah non, pas la SNCF.

Table des matières

Remerciements	3
Table des matières	5
1 Motivations	7
2 Tour d’horizon de l’ADN	11
2.1 Structure de l’ADN	11
Structures primaire et secondaire	11
Linking number	12
Surenroulement	15
2.2 Expériences de molécule unique	16
Extension d’une molécule étirée	17
Détermination du couple	18
Dynamique des plectonèmes	20
2.3 Modèles et simulations	21
Le modèle Worm-Like Chain	21
Simulations de Monte-Carlo	26
2.4 Objectifs	28
3 Un algorithme adapté aux longues molécules	31
3.1 Calcul direct du linking number	31
Twist local	31
Continuité du brin et rotations	32
Conservation du linking number	33
Molécules linéaires	34
3.2 Mouvements invalides	36
Conditions aux bords	36
Auto-évitement de la chaîne	37
Détection des croisements	39
Détection des nœuds	43
3.3 Performances	45
Complexité algorithmique	45
Parallélisation	47
Implémentation	48

4	Retour sur les expériences de molécule unique	51
4.1	Étirement de l'ADN surenroulé	51
	Le modèle Worm-Like Chain non surenroulé	51
	Le modèle Worm-Like Chain surenroulé	52
	Couple	54
	Effet de la discrétisation	59
	Énergie de torsion globale ou locale?	62
	Convergence des simulations	62
	Comparaison avec les simulations existantes	64
4.2	Structures créées	66
	Visualisation des structures	66
	Nature des structures	75
	Nombre de plectonèmes	79
5	Interactions à longue portée	85
5.1	Fréquence de contact	85
	Contacts entre sites distants	86
	Contacts en fonction de la distance	88
5.2	Domaines indépendants	92
	Modélisation d'un isolant topologique	92
	Isolation de deux domaines	93
6	Conclusion	97
	Bibliographie	101

Chapitre 1

Motivations

Régulation de la transcription

L'ÉTUDE de la régulation de la transcription de l'ADN, amorcée par la découverte de l'opéron lactose par Jacob et Monod [1], a permis de dévoiler l'importance des interactions, parfois à longue portée, entre différentes régions du chromosome. Chez les bactéries, la régulation de la transcription peut nécessiter la formation de boucles le long de l'ADN, de manière à placer face-à-face les opérateurs sur lesquels viennent se fixer les facteurs de transcription. Ces opérateurs peuvent se trouver séparés (entre eux ou du promoteur) de quelques dizaines ou centaines de paires de bases (répresseurs *lac* ou *gal*, par exemple [2, 3, 4]), mais ils peuvent aussi être beaucoup plus éloignés (2,4 kb pour l'opérateur O_L lié au répresseur CI du bactériophage λ [5]).

Plus généralement, de nombreuses études suggèrent que le fonctionnement optimal des cellules bactériennes est basé sur la capacité des gènes à être proches dans l'espace cellulaire. Tout d'abord, les gènes régulés par un même facteur de transcription ont tendance à se trouver à proximité du gène codant pour ce facteur de transcription [6, 7], ou bien disposés périodiquement le long du chromosome [6, 8], la périodicité favorisant la co-localisation spatiale de gènes éloignés en interaction [9]. La prise en compte de la position génomique dans la recherche de sites de fixation de facteurs de transcription permet d'ailleurs d'améliorer la sensibilité et la spécificité des résultats, par rapport à une approche purement séquentielle [10]. Ensuite, des observations utilisant la fluorescence ont mis en évidence la co-localisation spatiale des ARN polymérase, en particulier dans un milieu riche [11], mais également la co-localisation de facteurs de transcription [12, 13], même lorsque ceux-ci ont un grand nombre de cibles et existent en un grand nombre de copies, comme H-NS [14]. Ces observations suggèrent la présence d'usines de transcription chez les bactéries [15], similairement aux eucaryotes, permettant de contrôler et d'optimiser la régulation de la transcription, notamment par un effet d'augmentation de la concentration locale de facteurs de transcription parfois peu nombreux [16]. Plus précisément, à l'échelle locale, l'aggrégation des gènes proches du gène régulateur permet de réduire le temps de recherche des cibles

par les facteurs de transcription [17, 18, 19, 20]. À l'échelle globale, les effets de concentration locale sont censés permettre une réponse optimale des réseaux de régulation [21, 22].

Enfin, à plus grande échelle, le chromosome d'*Escherichia coli* est divisé en macrodomaines, régions longues de 600 kb à 1 Mb environ, et caractérisées par une fréquence de contacts entre gènes intra-domaine bien plus élevée que la fréquence de contacts inter-domaines [23]. Cette organisation se traduit par une compartimentation des macrodomaines dans l'espace cellulaire, qui évolue selon le stade du cycle cellulaire et influant sur le déroulement de la ségrégation des chromosomes [24, 25].

Rôle du surenroulement

Comprendre ces phénomènes de structuration spatiale du chromosome, et plus particulièrement de la co-localisation des gènes, nécessite de comprendre en détail l'impact du surenroulement, celui-ci ayant un effet à toutes les échelles de structuration. En effet, localement, la modification de l'enroulement de l'hélice peut affecter l'accessibilité des sites de fixation des facteurs de transcription [26] et perturber l'alignement des opérateurs de part et d'autre d'une boucle, réduisant la capacité de régulation du facteur de transcription [27, 4]. À plus grande distance, le repliement de l'ADN surenroulé en plectonèmes favorise la juxtaposition de sites distants de quelques kb [28] à plusieurs dizaines de kb [29].

Ainsi, le surenroulement est un régulateur global de la transcription [30], connu pour affecter plusieurs centaines de gènes d'*Escherichia coli* [31], par exemple, et est même soupçonné être le principal régulateur de la transcription pour *Mycoplasma genitalium*, chez qui peu de facteurs de transcription sont connus [32].

Ainsi, pour répondre aux questions liées à la régulation transcriptionnelle, la modélisation doit donc nécessairement passer par une approche multi-échelle du chromosome bactérien [33, 34, 35], prenant en compte l'effet du surenroulement.

Simulations multi-échelles

Bien que plusieurs approches pour modéliser et simuler l'ADN surenroulé existent depuis quelques décennies déjà [36], aucune ne permet à l'heure actuelle de simuler de longues molécules. La dynamique de Langevin et la dynamique brownienne sont limitées à des molécules de quelques kb [37, 28, 38, 39], tandis qu'une approche de type Monte-Carlo permet difficilement de dépasser 10 kb [40, 41] à cause de la nécessité de respecter, à l'échelle globale, la contrainte topologique de la conservation du linking number.

La majeure partie de ma thèse (Chapitre 3) a donc consisté à améliorer cette dernière approche afin de l'adapter aux longues molécules, sans altérer la qualité des modèles. Grâce à ce nouvel outil, j'ai alors pu m'intéresser aux propriétés de l'ADN au delà de 10 kb (Chapitre 4), mais également commencer à traiter le problème de la structuration en micro-domaines de super-hélicité (Chapitre 5.2), qui est inhérente à la structuration du chromosome bactérien [42]. Cette structuration peut être induite par le pontage de facteurs de transcription [43, 44], par l'ancrage à la membrane [45] ou encore par la machine de

transcription [46, 47, 48], et a un effet important sur la formation de longues boucles d'ADN et donc sur la régulation de la transcription [49].

Chapitre 2

Tour d'horizon de l'ADN

2.1 Structure de l'ADN

Structures primaire et secondaire

L A DÉCOUVERTE de la structure de la molécule d'acide désoxyribonucléique (ADN) en 1953 [50, 51, 52] est une étape fondamentale pour la biologie. À partir de ce moment, les processus se déroulant à l'intérieur de la cellule ont commencé à être décrits à l'échelle moléculaire et, progressivement, la compréhension des mécanismes de la réplication, de la transcription, de la régulation, a mis en lumière les rouages internes du vivant, ouvrant la voie à la biologie moléculaire, puis au génie génétique et à la biologie de synthèse.

L'acide désoxyribonucléique est une chaîne de paires de nucléotides, eux-mêmes composés d'une base azotée (adénine, thymine, guanine ou cytosine), d'un désoxyribose et d'un groupe phosphate (Fig. 2.1). L'appariement de deux nucléotides est spécifique (l'adénine ne s'apparie qu'à la thymine et la guanine ne s'apparie qu'à la cytosine) et se fait par des liaisons hydrogène entre leurs bases azotées : deux liaisons pour la paire adénine-thymine et trois pour la paire guanine-cytosine.

Ce polymère de nucléotides prend la forme d'une double hélice dont les brins sont constitués de l'enchaînement des désoxyriboses et des groupes phosphate des nucléotides consécutifs. Suivant la composition de l'environnement, cette double hélice peut présenter plusieurs aspects [53] (Fig. 2.2). La forme B, la plus commune (et la seule forme étudiée dans mes simulations), est une hélice droite comprenant 10,4 paires de bases par tour. Les paires de bases sont perpendiculaires à l'axe de la molécule et séparées de 0,34 nm le long de cet axe. Elles ne sont pas centrées sur celui-ci, ce qui forme un grand et un petit sillon. La forme A présente une forme beaucoup plus inclinée : les bases ne sont plus perpendiculaires à l'axe de l'hélice. La forme Z, quant à elle, est une hélice gauche.

Lorsque la molécule est étirée et surenroulée positivement, elle peut également adopter une forme très étirée, appelée forme P, où les bases sont à l'extérieur du squelette [54]. Inversement, lorsqu'elle est étirée et surenroulée

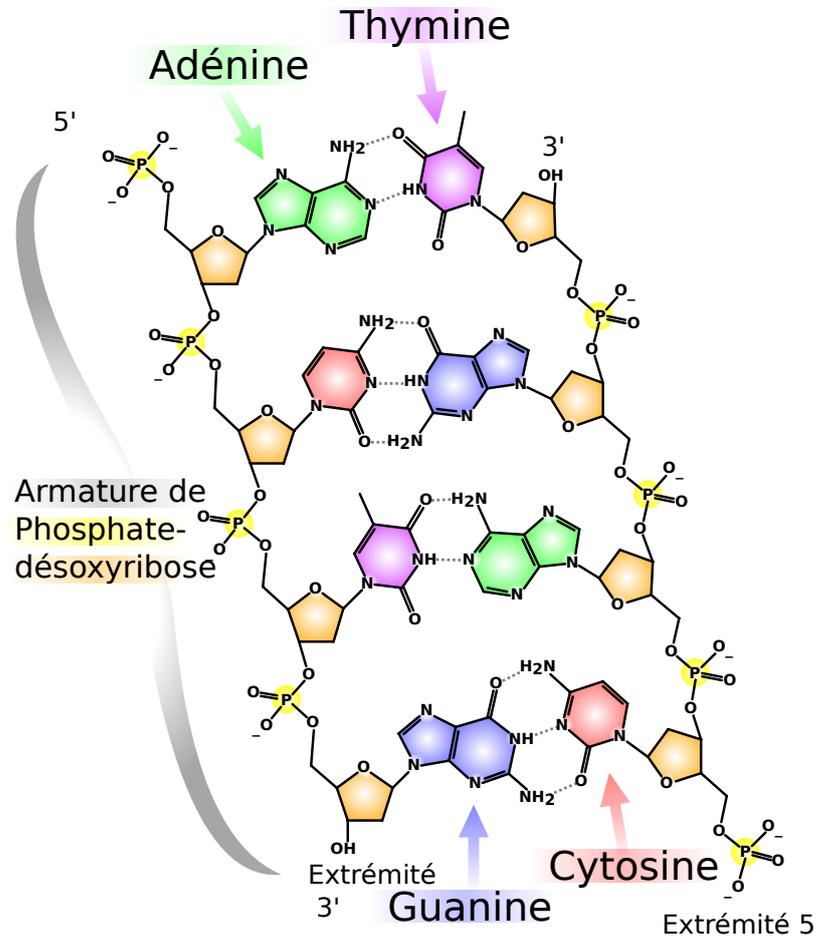


FIGURE 2.1: Schéma de la structure primaire de l'ADN (M. P. Ball, CC BY-SA).

négativement, l'appariement de certains nucléotides se brise et des bulles de dénaturation peuvent apparaître, régions constituées de deux morceaux d'ADN simple brin [54].

Linking number

Dès les années 1960 [55], la découverte du surenroulement a dévoilé la structure tertiaire adoptée par l'ADN, c'est-à-dire la manière dont la double-hélice se replie sur elle-même dans l'espace (Figs. 2.3 et 2.4). Ce phénomène apparaît pour des molécules circulaires, qui sont soumises à une contrainte topologique forte : la conservation du linking number.

Le linking number est un entier défini, pour deux courbes C_1 et C_2 fermées et orientées, comme le degré du lien qu'elles forment (Fig. 2.5). Un diagramme du lien (projection à deux dimensions) contient un nombre pair de croisements

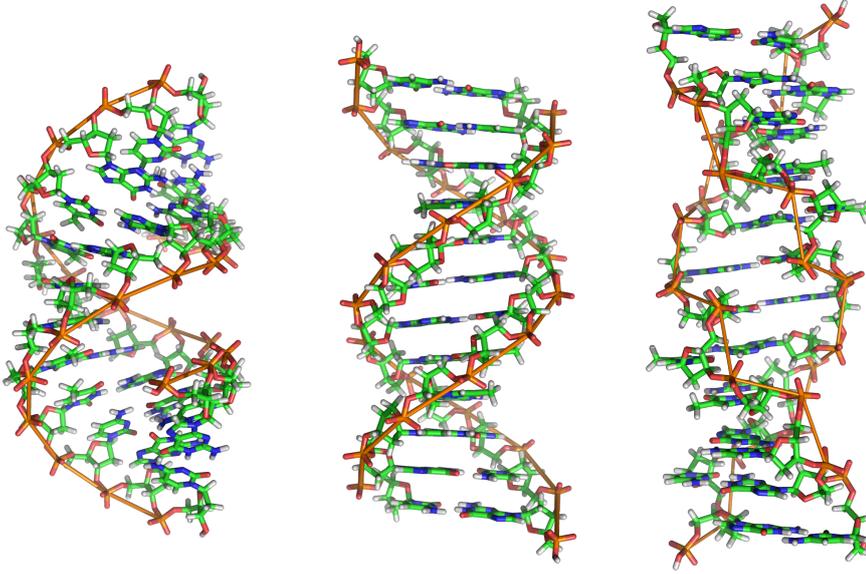


FIGURE 2.2: Schéma des structures secondaires possibles de l'ADN. De gauche à droite : formes A, B et Z (R. Wheeler, CC BY-SA).

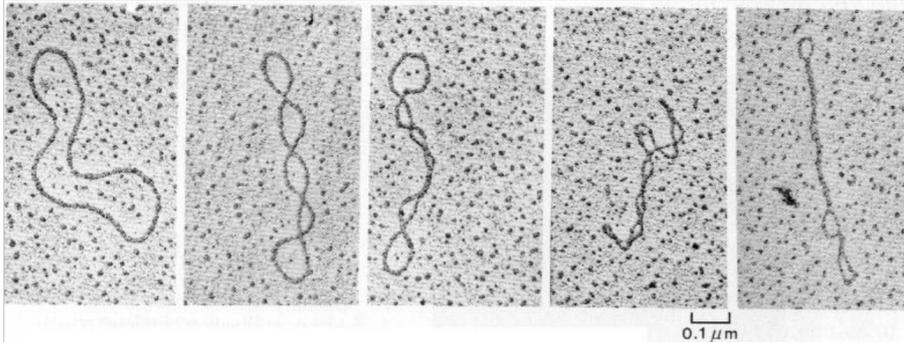


FIGURE 2.3: Structure tertiaire de l'ADN observée par microscopie électronique sur un plasmide de 5 kb [56]. Le degré de surenroulement augmente de gauche à droite. Ces structures composées d'une série de boucles sont appelées plectonèmes.

entre les deux courbes. Chacun de ces croisements contribue au linking number total de $\pm \frac{1}{2}$, le signe étant déterminé en regardant le croisement de manière à ce que la courbe du dessous soit orientée vers le haut. Le signe est alors + si la courbe du dessus va vers la droite, et - si elle va vers la gauche (Fig. 2.6). Le linking number peut également être défini par l'intégrale de Gauss [58] :

$$\text{Lk} = \frac{1}{4\pi} \oint_{C_1} \oint_{C_2} \frac{\vec{r}_1 - \vec{r}_2}{|\vec{r}_1 - \vec{r}_2|^3} \cdot (d\vec{r}_1 \wedge d\vec{r}_2) \quad (2.1)$$

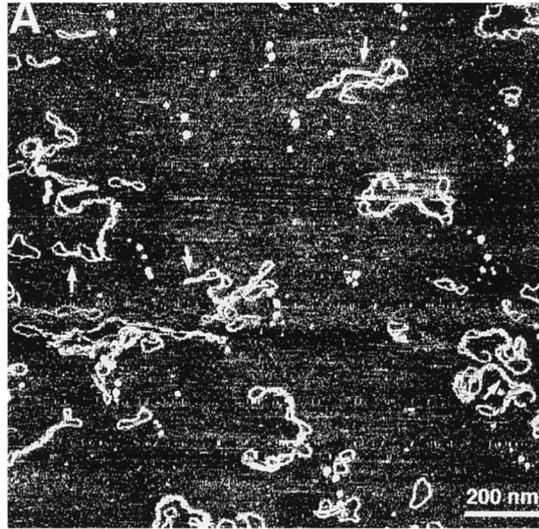


FIGURE 2.4: Plasmides de 3,8 kb surenroulés observés par microscopie à force atomique [57].

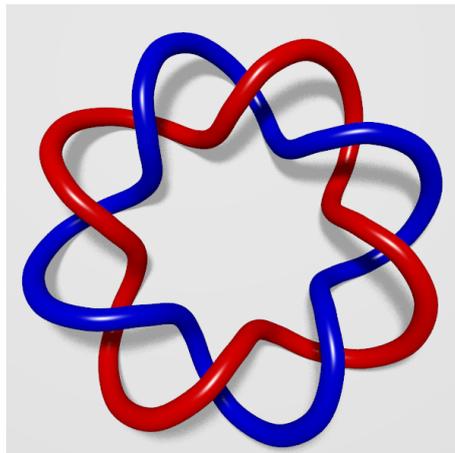


FIGURE 2.5: Lien entre deux courbes fermées (rouge et bleue). Le degré de ce lien, son linking number, vaut $Lk = \pm 4$ (le signe dépend de l'orientation relative des deux courbes).

Le linking number est un invariant topologique, c'est-à-dire qu'il reste inchangé par toute déformation des courbes, tant que l'une ne passe pas au travers de l'autre. Les deux courbes dont il est question peuvent représenter les deux brins de la double-hélice ou bien, de manière équivalente, un seul des deux brins et l'axe de la molécule, qui ne se croisent pas sans une intervention extérieure (d'une topoisomérase, par exemple). D'après le théorème de Călugăreanu [59], le linking number peut se décomposer en deux quantités : le twist Tw et le

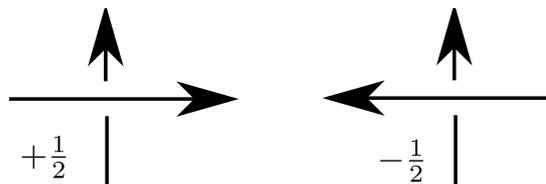


FIGURE 2.6: Contribution des deux types de croisements possibles au linking number total d'un lien.

writhe Wr :

$$Lk = Tw + Wr \quad (2.2)$$

Lorsque les deux courbes C_1 et C_2 correspondent à l'axe et à un brin de l'ADN, respectivement, le twist correspond au nombre d'hélices (nombre de tours d'un brin autour de l'axe) et le writhe correspond à l'entortillement de l'axe sur lui-même, dont l'expression analytique s'écrit :

$$Wr = \frac{1}{4\pi} \oint_{C_1} \oint_{C_2} \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} \cdot (d\vec{r} \wedge d\vec{r}') \quad (2.3)$$

La formation de plectonèmes dans une molécule surenroulée est la conséquence de cette conservation du linking number. En effet, le surenroulement introduit des défauts d'appariement des nucléotides, qui sont coûteux en énergie libre. Afin de rétablir l'appariement, la molécule tend à rétablir son twist. Mais du fait de la contrainte topologique à laquelle est soumise, la seule manière de diminuer son twist excédentaire consiste à augmenter son writhe en se repliant. La détermination de la nature exacte des structures formées (hélices, plectonèmes, série de boucles) est cependant un problème thermodynamique difficile à résoudre, impliquant l'énergie de torsion, l'énergie de courbure et l'entropie des structures formées.

Surenroulement

Chez la plupart des bactéries, l'ADN est circulaire (le chromosome comme les plasmides) et négativement surenroulé : la densité d'hélices est réduite (d'environ 6 % pour *E. coli* par exemple).

Une famille de protéines présente chez tous les êtres vivants, les topoisomérases, régule le surenroulement de l'ADN en réalisant des coupures simple-brin (type I) ou double-brin (type II) pour ajouter ou enlever des tours à la double-hélice [60, 61] (et également pour séparer les deux copies du chromosome emmêlées par la réplication). La transcription est également un processus générant du surenroulement positif en aval et négatif en amont, notamment lorsque certains sites de l'ADN sont ancrés à une large structure immobile, ou bien lorsque deux unités de transcription ont des orientations opposées [45].

Le surenroulement n'est pas distribué uniformément le long du chromosome. Le génome bactérien est divisé en domaines topologiquement indépendants. La machine de transcription ou la fourche de réplication ne sont pas les seules frontières possibles entre ces domaines : plusieurs protéines capables de se fixer sur

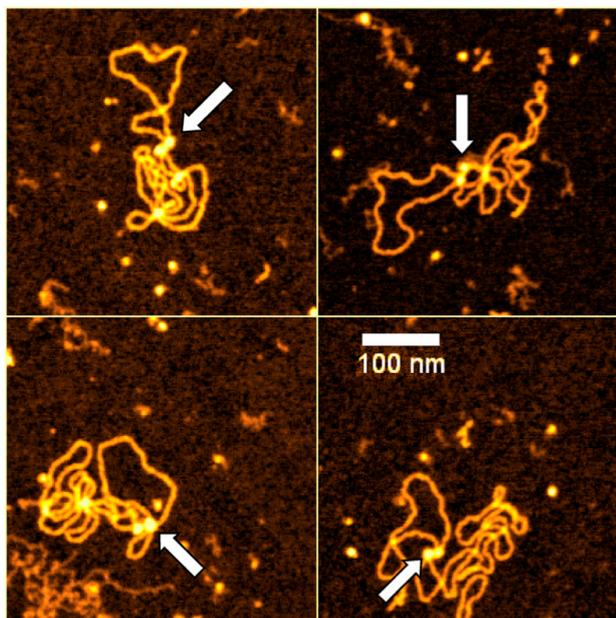


FIGURE 2.7: Plasmides divisés en deux domaines indépendants par la fixation de la protéine LacI sur deux sites, observés par microscopie à force atomique [43]. Seul l'un des deux domaines est surenroulé.

l'ADN (*nucleoid associated proteins* ou NAPs) peuvent également servir d'isolants topologiques [43] (Fig. 2.7). Toutefois, la manière dont ces protéines sont capables de constituer une barrière topologique n'est pas encore bien comprise à ce jour, et aucun modèle n'a encore essayé, à ma connaissance, de décrire un tel isolant (Section 5.2).

2.2 Expériences de molécule unique

Ces dernières décennies ont vu fleurir une variété de techniques expérimentales permettant de manipuler une unique molécule d'ADN linéaire [62, 63], notamment l'étirer, afin de mesurer ses propriétés physiques. Ainsi, il est possible d'attacher une extrémité de la molécule à une bille maintenue en place par la succion d'une pipette et l'autre extrémité à une fibre optique permettant de mesurer l'extension de cette molécule et la force de traction exercée par la pipette [64]. La seconde extrémité peut également être attachée à une autre bille maintenue par un piège optique, l'étirement de la molécule étant obtenu en éloignant la pipette du rayon laser formant le piège [65, 66]. La première extrémité peut aussi être attachée à une surface fixe, tandis que la seconde est attachée à une bille magnétique maintenue en place par des aimants, la traction étant d'origine magnétique [67, 68, 69] ou hydrodynamique [70]. La seconde extrémité peut, là encore, être maintenue par un piège optique [71].

Généralement, l'attachement entre l'extrémité de la molécule et la surface fixe ou la bille se fait via l'interaction spécifique entre un marqueur fixé sur la

molécule (biotine ou digoxigénine, typiquement) et une protéine recouvrant la surface ou la bille (streptavidine, antidigoxigénine) [72, 68].

Extension d'une molécule étirée

Parmi ces méthodes, les pinces magnétiques sont particulièrement intéressantes du fait de la possibilité qu'elles offrent de contrôler simultanément la force et le surenroulement auxquels la molécule est soumise [67, 73]. En effet, celle-ci est attachée à une surface par une extrémité et à une bille magnétique par l'autre (Fig. 2.8). En contrôlant la position et l'orientation d'aimants (les pinces magnétiques), il est possible d'étirer l'ADN avec une force f , tout en lui imposant un surenroulement ΔLk . La position moyenne de la bille est mesurée à partir de la figure de diffraction produite par la bille éclairée par une source monochromatique [73]. En comparant la taille de la figure de diffractions à une série de figures de calibration, il est possible de déterminer la position de la bille et d'en déduire l'extension moyenne $\langle z \rangle$ de la molécule. La mesure des fluctuations transverses δx^2 , dues au mouvement brownien, permet de connaître la force f appliquée :

$$f = \frac{k_B T \langle z \rangle}{\delta x^2} \quad (2.4)$$

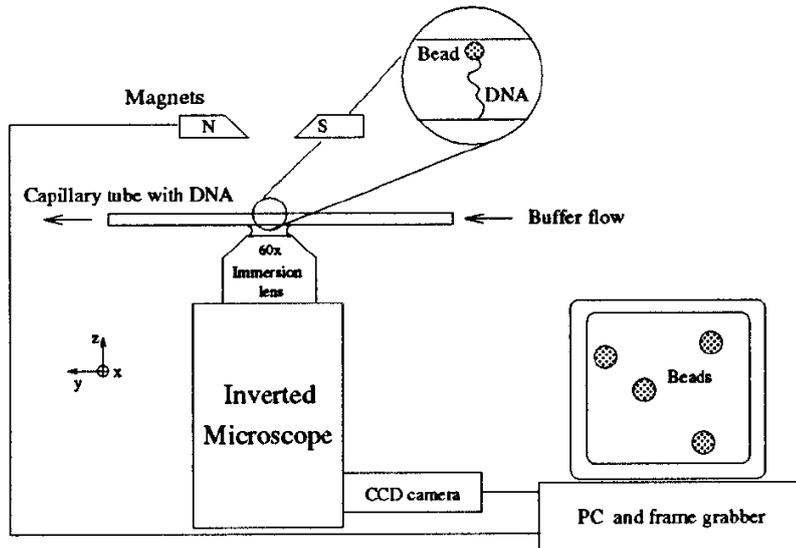


FIGURE 2.8: Schéma d'une expérience de molécule unique par piégeage magnétique [67]. La molécule d'ADN est fixée à la paroi du tube par une extrémité et à une bille magnétique par l'autre. Les mouvements de la bille peuvent être contrôlés par des aimants et observés à l'aide d'une caméra.

La figure 2.9 montre les résultats d'une telle expérience, c'est-à-dire l'extension d'une molécule étirée en fonction de la force appliquée et du surenroulement. Ce système présente une transition de flambage, d'autant plus singulière

que la force est grande. Avant cette transition, l'extension est approximativement constante et correspond à une phase étirée où la contrainte topologique est stockée sous forme d'énergie de torsion. À la transition, une part de cette énergie de torsion est soudainement convertie en énergie de courbure pour former des plectonèmes.

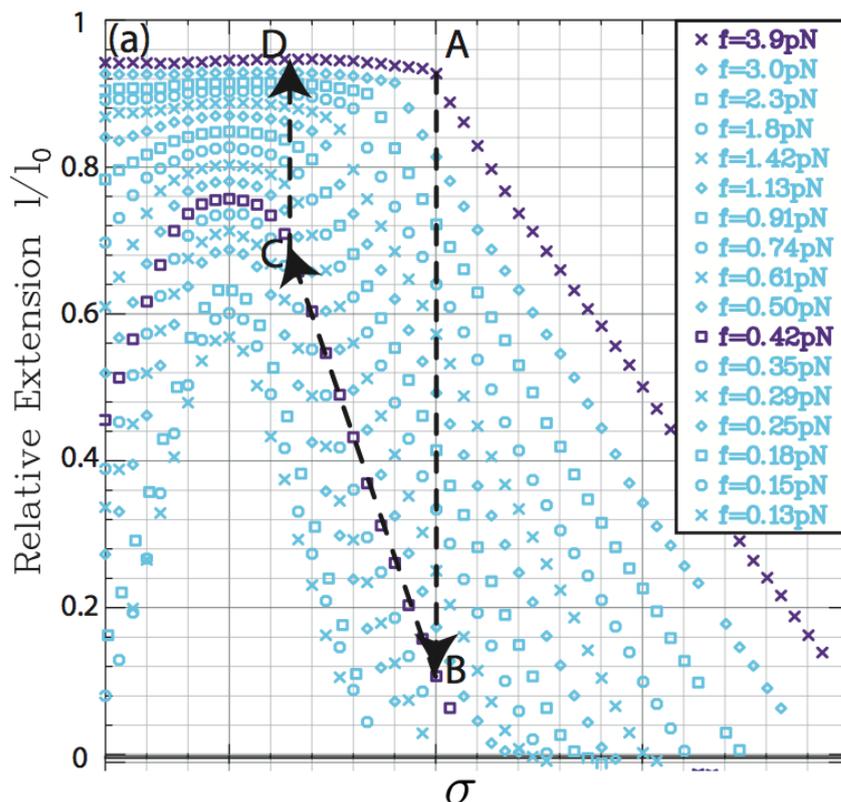


FIGURE 2.9: Extension relative d'une molécule de 16 kb en fonction du surenroulement, pour différentes forces [74].

Détermination du couple

Méthode directe

En attachant une seconde bille témoin, plus petite, à la première, Lipfert *et. al.* ont pu mesurer la position angulaire θ de cette bille témoin, et ses fluctuations $\delta\theta^2$, afin d'en déduire le couple Γ exercé par la molécule sur la bille magnétique [75]. En effet, en notant k_r la raideur du potentiel appliqué à la bille magnétique et à l'ADN, ce système répond par le couple :

$$\Gamma = -k_r \langle \theta \rangle \quad (2.5)$$

Or la raideur k_r est reliée aux fluctuations par le théorème d'équipartition de l'énergie :

$$k_r = \frac{k_B T}{\langle \delta\theta^2 \rangle} \quad (2.6)$$

Forth *et. al.* [71] ont également utilisé une technique de pinces optiques pour mesurer ce couple (Fig. 2.10). Dans ce type d'expériences, la bille magnétique est remplacée par un cylindre de quartz biréfringent piégé par un laser dont la polarisation peut faire tourner le cylindre pour engendrer du surenroulement [76, 71]. Le couple croît linéairement avec le surenroulement dans la phase étirée (l'énergie de torsion étant de plus en plus grande), avant de se stabiliser à la transition. Il est constant dans la phase repliée, le surenroulement étant changé en writhe (courbure) plutôt qu'en twist (torsion).

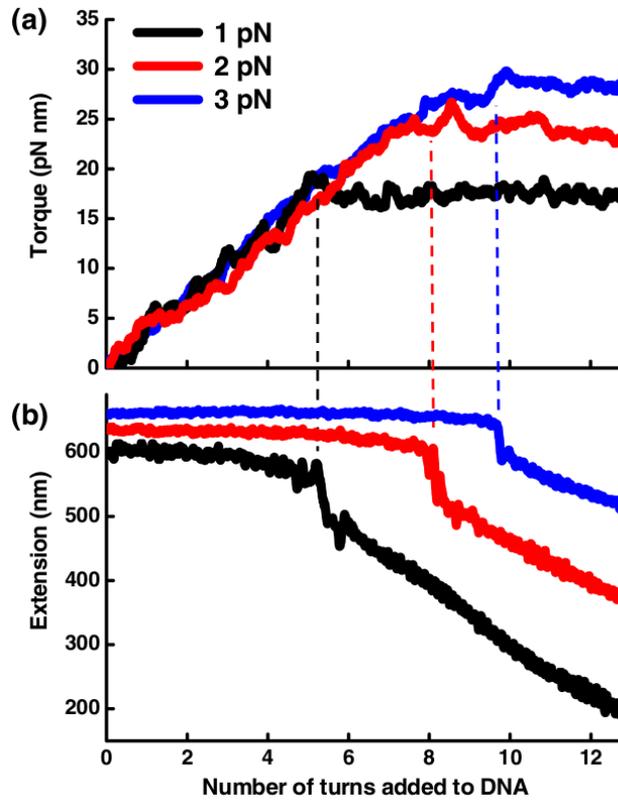


FIGURE 2.10: Extension et couple d'une molécule de 2,2 kb en fonction du surenroulement, pour différentes forces [71].

À la différence des expériences précédentes (Section 2.2), celles-ci montrent un saut prononcé de l'extension à la transition de flambage. Ce phénomène est dû à la taille bien plus faible de la molécule étudiée (2,2 kb au lieu de 16 kb), ce qui permet de voir l'effet de la formation du premier plectonème sur l'extension, à la transition.

Méthode indirecte

Il est également possible de déterminer le couple à partir des mesures de force et d'extension [74], qui ne requièrent pas les dispositifs expérimentaux plus complexes utilisés pour les mesures directes (bille témoin, quartz biréfringent). En effet, si le couple $\Gamma(f_0, \Theta)$ est connu pour une force f_0 et un surenroulement $\Theta = 2\pi\Delta Lk$, le couple pour n'importe quelle autre force f est donné, de façon relative, par :

$$\Gamma(f, \Theta) = \Gamma(f_0, \Theta) + \int_{f_0}^f \left(\frac{\partial l}{\partial \Theta} \right)_{f'} df' \quad (2.7)$$

De plus, en faisant l'hypothèse que Γ est constant dans la phase repliée, il est possible de connaître $\Gamma(f, \Theta)$ sur une large gamme de valeurs de Θ . Enfin, en considérant que le couple est nul en l'absence de surenroulement : $\Gamma(f, 0) = 0$ et qu'il augmente linéairement dans la phase étirée, l'ensemble des valeurs de Γ peut être calculé de manière absolue [74].

Dynamique des plectonèmes

Si les mesures d'extension et de couple permettent de distinguer les deux phases, étendue et repliée, de la molécule d'ADN, elles n'apportent aucun détail sur la structure exacte dans la phase plectonémique. Van Loenhout *et. al.* ont récemment réalisé des expériences de pinces magnétiques dans lesquelles l'ADN est rendu fluorescent par l'ajout de cyanine (Cy3) le long de la molécule [69]. La mesure de l'intensité de la fluorescence le long de l'axe de la force permet alors de visualiser la densité d'ADN le long de cet axe, avec une résolution de l'ordre de 500 nm, correspondant à la taille des taches fluorescentes observées (Fig. 2.11). Les zones de hautes densité correspondent à des plectonèmes (ou des groupes de plectonèmes).

En enregistrant l'évolution de cette densité au cours du temps, il est même possible de suivre le mouvement de ces structures et d'étudier la dynamique des plectonèmes. En l'occurrence, Van Loenhout *et. al.* ont observé un mouvement de diffusion des plectonèmes le long de l'axe, entrecoupé de sauts brusques vers des positions distantes, dévoilant ainsi une dynamique complexe soulevant plusieurs questions quant à la capacité de réorganisation de l'ADN à l'échelle de la dizaine de kb.

Cette expérience permet notamment de mesurer le nombre de plectonèmes formés le long d'une molécule de 21 kb, en fonction de la force et des conditions salines. La comparaison avec les résultats de mes simulations, qui prédisent systématiquement un nombre de plectonèmes plus élevé que celui observé, soulève la question de l'effet local de la séquence sur la formation de ces structures (Section 4.2).

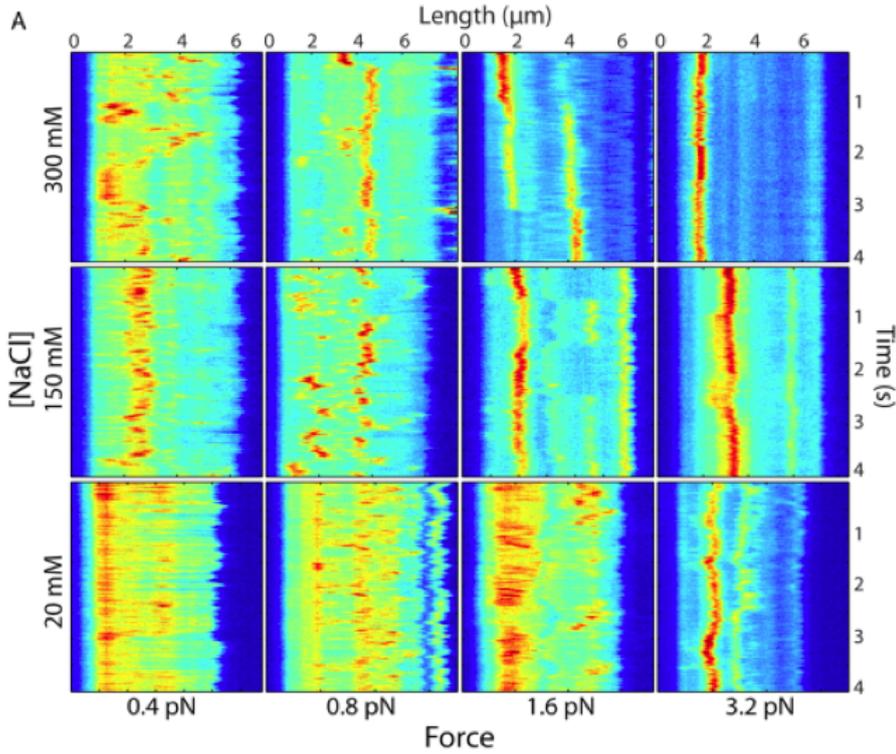


FIGURE 2.11: Évolution temporelle de la densité d'ADN le long de l'axe de la force, pour différentes forces et concentrations en sel (molécule de 21 kb) [69]. Le surenroulement est ajusté de manière à ce que l'extension soit inférieure de 25% à l'extension sans surenroulement. Les zones de hautes densité (jaunes et rouges) correspondent à un ou plusieurs plectonèmes.

2.3 Modèles et simulations

Simuler le repliement d'une molécule d'ADN à grande échelle nécessite un modèle physique qui permette de décrire quantitativement les propriétés macroscopiques (extension, couple, etc.) d'une molécule d'ADN de plusieurs milliers de paires de bases. Ces propriétés dépendent très peu du détail de la séquence, mais plutôt des caractéristiques statistiques des propriétés de courbure et de torsion dont les échelles caractéristiques sont de l'ordre de 50 à 100 nm [77].

Le modèle Worm-Like Chain

Modèle de Kratky-Porod

Le modèle de Kratky-Porod est le modèle hors-réseau le plus simple permettant d'étudier les effets de la longueur de persistance, et sert de base au modèle Worm-Like Chain et à des extensions prenant en compte le surenroulement. Dans ce modèle, l'ADN est représenté sous la forme d'une chaîne de N cylindres de longueur l et de directions \vec{t}_i ($i = 1 \dots N$) (Fig. 2.12). La longueur

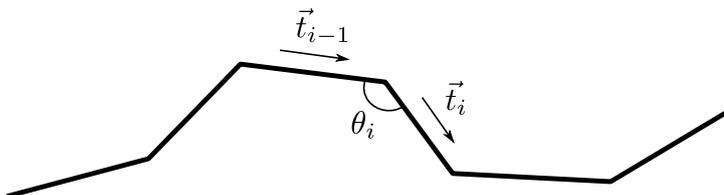


FIGURE 2.12: Schéma du modèle de Kratky-Porod. Deux segments adjacents $i-1$ et i forment un angle θ_i tel que $\cos \theta_i = \vec{t}_{i-1} \cdot \vec{t}_i$, ce qui contribue à l'énergie de courbure totale de la chaîne.

totale de la chaîne est notée L . La rigidité du polymère se traduit par l'énergie de courbure :

$$E_{\text{KP}} = -k_B T \frac{l_p}{l} \sum_{i=1}^N \vec{t}_i \cdot \vec{t}_{i-1} \quad (2.8)$$

avec $\vec{t}_0 = \vec{t}_N$ (pour une molécule circulaire ; dans le cas d'une molécule linéaire, la somme débute à $i = 2$). La longueur de persistance l_p correspond à la distance typique de l'autocorrélation de l'orientation de l'axe (Fig. 2.13) :

$$\langle \vec{t}_i \cdot \vec{t}_j \rangle = e^{-l \frac{|j-i|}{l_p}} \quad (2.9)$$

Ainsi, le polymère apparaît très rigide sur des longueurs inférieures à l_p , tandis que les fluctuations thermiques dominent sur des longueurs supérieures. Ce modèle est analogue au modèle de Heisenberg pour un système de N spins à une dimension [78], résoluble analytiquement et pouvant donc être facilement testé.

Le rayon de giration R_g représente l'extension spatiale typique du polymère. Il est défini par :

$$R_g^2 = \left\langle \sum_{i=1}^N (\vec{r}_i - \vec{\rho})^2 \right\rangle \quad (2.10)$$

où \vec{r}_i est la position du monomère i et $\vec{\rho} = \sum_{i=1}^N \vec{r}_i / N$ est le barycentre de la molécule. Dans la limite des grandes chaînes, le rayon de giration vérifie $R_g^2 = \frac{l_p L}{3}$ (Fig. 2.14). Le faible écart entre la théorie et les simulations à faible longueur peut s'expliquer par un effet de taille finie (en plus de l'effet l'échelle logarithmique).

Limite continue

Le modèle Worm-Like Chain correspond à la limite continue du modèle de Kratky-Porod : $N \rightarrow \infty$ et $l \rightarrow 0$ (à longueur totale $L = Nl$ constante). Dans ce cas, l'énergie de courbure devient :

$$E_b = \frac{1}{2} k_B T l_p \int_0^L \left(\frac{d\vec{t}}{ds} \right)^2 ds \quad (2.11)$$

Il est possible de modéliser les expériences de molécule unique (Chapitre 4) en ajoutant à l'énergie un terme correspondant au travail de la force f :

$$E_f = -f (z(L) - z(0)) \quad (2.12)$$

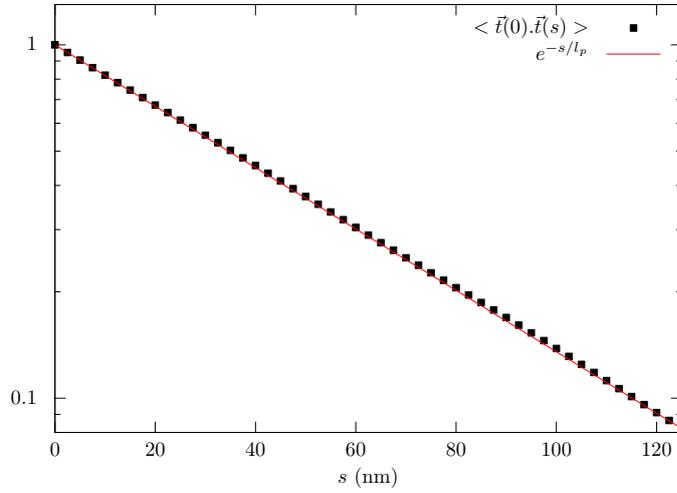


FIGURE 2.13: Autocorrélation de la direction $\vec{t}(s)$ où $s = |j - i|$ est la distance entre deux points le long de la chaîne ($l_p = 50$ nm), pour une chaîne linéaire fantôme. Les points noirs sont obtenus par simulation de Monte-Carlo.

pour un polymère étiré selon z . Il est souvent pratique d'utiliser l'extension relative $\chi = (z(L) - z(0))/L$. Si ces deux termes d'énergie sont minimisés lorsque la molécule est totalement étirée (extension maximale et courbure nulle), cette configuration minimise également l'entropie, et n'est donc pas favorable thermodynamiquement à température non nulle. Une configuration typique présentera plutôt des fluctuations de la courbure, réduisant l'extension, et d'autant plus faibles que la force f est élevée.

Comme le montre la figure 2.15, les mesures d'extension en fonction de la force, à surenroulement nul, sont très bien décrites par le modèle Worm-Like Chain, et permettent de déterminer la longueur de persistance de l'ADN double-brin, autour de 50 nm. Le modèle Worm-Like Chain extensible [79, 80], en tenant compte de l'élasticité de l'ADN à travers son module élastique (de l'ordre de 1 000 pN), étend le domaine de validité aux forces très importantes (plusieurs dizaines de pN). À faible force cependant, l'élasticité du squelette de l'ADN, mettant en jeu des énergies de liaisons covalentes bien supérieures à $k_B T$, est négligeable.

Modèle Worm-Like Chain surenroulé

Afin de prendre en compte le surenroulement de la molécule, plusieurs extensions au modèle Worm-Like Chain ont été développées pour intégrer la torsion [82, 83, 84, 85]. Une manière de procéder particulièrement adaptée aux simulations consiste à définir l'énergie de torsion E_t sous la forme [86, 41] :

$$E_t = \frac{1}{2} k_B T C \int_0^L (\phi'(s) - \phi'_0)^2 ds \quad (2.13)$$

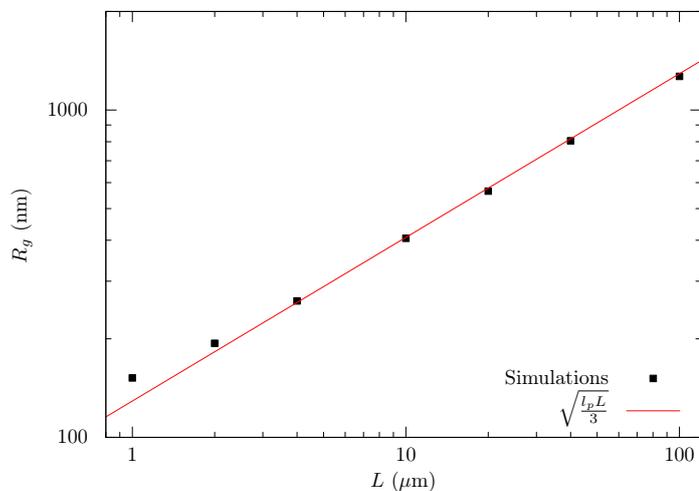


FIGURE 2.14: Rayon de giration R_g d'une chaîne linéaire fantôme en fonction de sa longueur L . Les points noirs sont obtenus par simulation de Monte-Carlo.

Où C est le module de torsion de l'ADN, $\phi'(s) = \frac{d\phi(s)}{ds}$ est le taux d'enroulement à l'abscisse curviligne s (Fig. 2.16), $\phi(s)$ étant l'angle d'enroulement du brin autour de l'axe, et ϕ'_0 représente le taux d'enroulement au repos (correspondant à un pas d'hélice de 10,4 paires de bases par tour pour la forme B).

En notant tw_i le twist du segment i et tw_0 le twist au repos d'un segment, la forme discrète de cette énergie s'écrit alors :

$$E_t = 2\pi^2 k_B T \frac{C}{L} \sum_{i=1}^N (tw_i - tw_0)^2 \quad (2.14)$$

Cette énergie peut cependant se décomposer en la somme de deux termes [86] :

$$E_t = 2\pi^2 k_B T \frac{C}{L} (Tw - Tw_0)^2 + 2\pi^2 k_B T \frac{C}{L} \sum_{i=1}^N \left(tw_i - \frac{Tw}{N} \right)^2 \quad (2.15)$$

Le premier est un terme global faisant intervenir le twist total Tw et le twist au repos $Tw_0 = Ntw_0$ de la molécule, tandis que le second représente les fluctuations locales du twist. Ce deuxième terme n'affecte pas les propriétés de repliement de l'ADN [41], ce que j'ai pu vérifier dans mes simulations (Section 4.1). Il peut donc être ignoré au profit d'une énergie de torsion globale [87], ce qui raccourcit fortement le temps nécessaire pour atteindre l'équilibre. En revanche, si la distribution exacte du twist est importante pour les simulations, par exemple dans le cas de la présence de domaines indépendants ou d'interactions avec des protéines nécessitant un alignement donné des brins (Chapitre 5.2), la forme globale ne peut plus être utilisée.

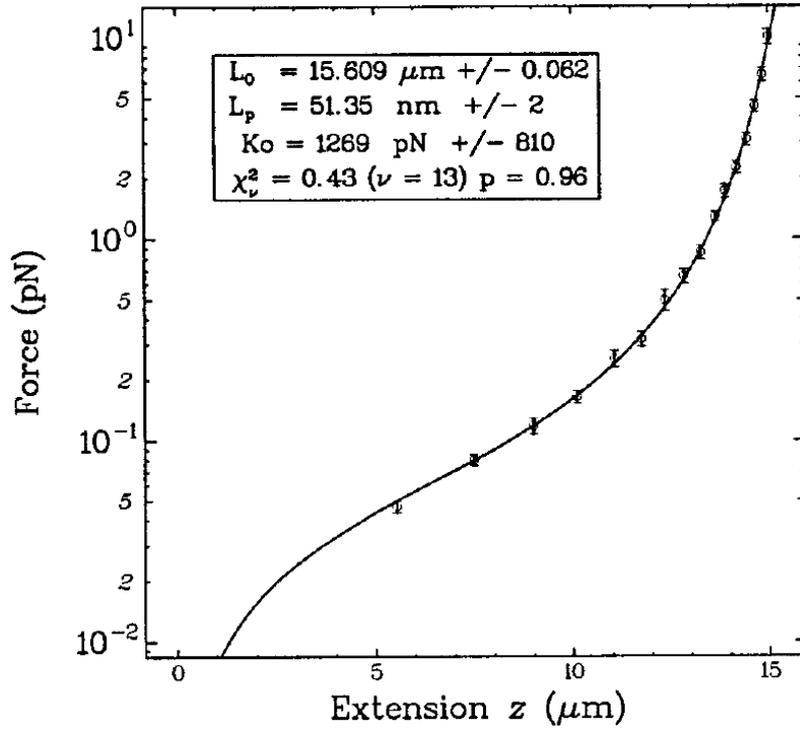


FIGURE 2.15: Courbe force-extension d'une molécule de 48,5 kb obtenue par expérience de molécule unique (à surenroulement nul). Le modèle Worm-Like Chain reproduit les données avec une remarquable précision [81].

Répulsion électrostatique

L'ADN étant négativement chargé, il est nécessaire de tenir compte de la répulsion électrostatique des brins, surtout pour étudier le surenroulement et le repliement de la molécule en plectonèmes, puisque de nombreuses régions du polymère vont entrer en contact. Afin de simplifier le modèle et d'accélérer les simulations, cette répulsion est généralement modélisée par un potentiel de cœur dur [88, 86, 85]. En effet, du fait de l'écrantage dû aux cations présents dans la solution, la répulsion électrostatique est une interaction à courte portée, disparaissant au-delà d'une distance de l'ordre de la longueur de Debye λ_D de la solution. Pour une solution de NaCl à la concentration c (en mM), cette longueur (en nm) est donnée par [39] :

$$\lambda_D = \frac{10}{\sqrt{c}} \quad (2.16)$$

Cette distance vient alors s'ajouter au rayon de la double-hélice d'ADN (1 nm) pour donner le rayon effectif r_e de la molécule.

Il est à noter que la longueur de persistance dépend elle-aussi en principe de la concentration en sel, puisqu'une plus grande répulsion électrostatique augmente la raideur de la molécule. Toutefois, l'effet de la salinité sur l_p ne de-

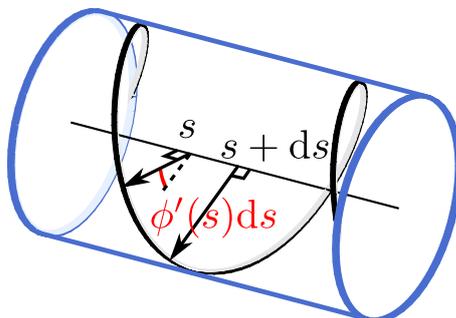


FIGURE 2.16: Schéma illustrant la définition du taux d'enroulement du brin autour de l'axe, $\phi'(s)$, utilisé pour calculer l'énergie de torsion.

vient détectable que pour des concentrations inférieures à $[\text{NaCl}] = 10 \text{ mM}$ [89], faibles en comparaison des concentrations généralement étudiées (quelques dizaines à quelques centaines de mM, la concentration physiologique de chlorure de sodium étant environ égale à 150 mM).

Simulations de Monte-Carlo

Pour pouvoir réaliser des simulations, le modèle Worm-Like Chain est discrétisé en une chaîne de N cylindres. Le niveau de discrétisation est un compromis entre performances et précision : les simulations ne représentent le modèle Worm-Like Chain qu'à condition de faire tendre N vers l'infini, mais le temps de calcul augmente rapidement avec le nombre de cylindres.

En pratique, les résultats obtenus (extension, writhe, etc.) dépendent de la discrétisation $n = Nl_p/L$ (le nombre de cylindres par longueur de persistance) avant de converger vers des valeurs indépendantes de n , pour $n > n_0$. Cette valeur n_0 est donc un bon compromis. Sa valeur précise sera étudiée dans les prochains chapitres (Sections 4.1 et 4.2).

L'objectif étant de pouvoir simuler de longues molécules, une approche de type dynamique de Langevin n'est pas très adaptée. En effet, la limite actuelle pour une molécule surenroulée semble être de quelques milliers de paires de bases [39], le problème de cette approche étant la nécessité d'utiliser un pas de temps très faible à l'échelle de la dynamique du système (de l'ordre de 10^{-13} s) afin que les billes composant la chaîne se déplacent peu à chaque étape, une molécule de 3 kb atteignant alors l'équilibre après 10^7 pas [39].

Les simulations de Monte-Carlo, en revanche, permettent d'effectuer des mouvements de grande amplitude et donc d'atteindre l'équilibre thermodynamique bien plus rapidement (quelques millions de pas pour une molécule de 16 kb, voir Section 4.1). L'inconvénient étant toutefois l'impossibilité d'étudier la dynamique du système, puisque celle-ci est artificielle.

Le principe des simulations de Monte-Carlo est d'échantillonner l'espace des conformations de la molécule, pondérées par leur poids statistique correspondant à l'équilibre thermodynamique. Ces conformations doivent donc être distribuées selon la loi de Boltzmann, c'est-à-dire que la probabilité d'obtenir

une conformation d'énergie E doit être :

$$p(E) = e^{-\frac{E}{k_B T}} \quad (2.17)$$

Pour obtenir cette distribution, les configurations sont générées l'une après l'autre, chacune étant obtenue à partir de la précédente *via* une rotation « test » (susceptible d'être rejetée) [88, 87]. Plus précisément, un bloc de taille aléatoire est tiré au hasard le long de la chaîne. Il va alors subir une rotation autour de son axe (Fig. 2.17), d'un angle aléatoire uniformément distribué entre $-\alpha$ et α (α étant choisi de manière à obtenir un taux de rejet de l'ordre de 50 %).

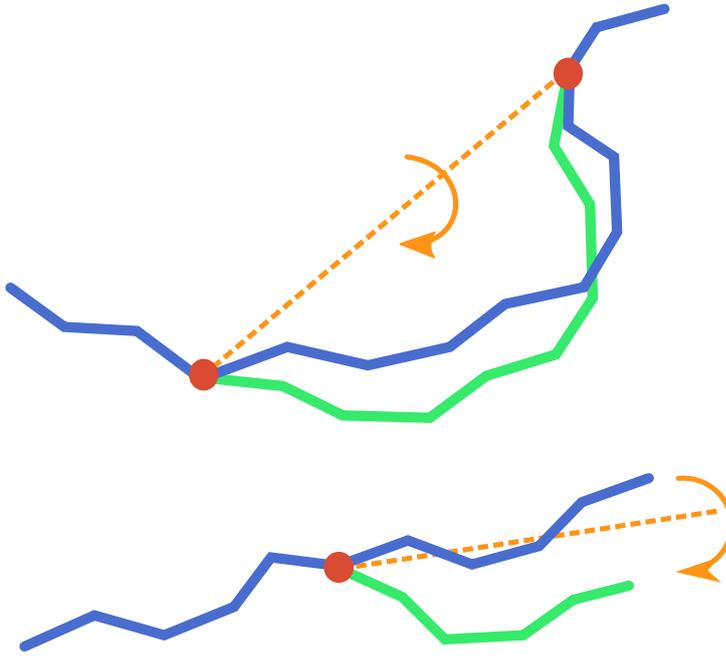


FIGURE 2.17: Schéma des rotations élémentaires. En haut : un bloc de taille aléatoire, délimité par deux cylindres « pivots » (points rouges), subit une rotation autour de son axe (en jaune) d'un angle aléatoire. En bas : pour les molécules linéaires, lorsque le bloc dépasse une extrémité de la chaîne, l'axe de rotation est tiré au hasard.

Dans le cas de molécules linéaires, il est nécessaire de permettre aux extrémités de la chaîne de se déplacer. Pour ce faire, lorsque la taille et la position du bloc sont telles que ce dernier dépasse une extrémité, l'axe de rotation est choisi aléatoirement.

Si le mouvement est invalide (notamment si deux cylindres entrent en collision, voir Section 3.2), il est immédiatement rejeté. Sinon, la différence d'énergie ΔE par rapport à la précédente conformation est calculée, et la rotation est acceptée selon la méthode de Metropolis afin que les configurations sélectionnées soient effectivement distribuées selon la loi de Boltzmann [90] :

$$p(\Delta E) = \begin{cases} 1 & \text{si } \Delta E < 0, \\ e^{-\frac{\Delta E}{k_B T}} & \text{si } \Delta E > 0. \end{cases} \quad (2.18)$$

Pour calculer l'énergie de la molécule, et notamment son énergie de torsion, il est nécessaire de connaître son twist. Jusqu'à présent, en l'absence de détails locaux du twist, les simulations de Monte-Carlo ont utilisé la formule globale de l'énergie de torsion [87] en calculant le twist total de la molécule à partir de son writhe et du théorème de Călugăreanu (Section 2.1). Toutefois, le writhe étant défini par une double intégrale, son calcul numérique nécessite une somme sur toutes les paires de cylindres de la chaîne [91]. Par conséquent, le temps nécessaire, à chaque pas, pour calculer l'énergie d'une molécule composée de N cylindres augmente comme N^2 (Section 3.3). Si le calcul de la différence de writhe entre une courbe de référence et la configuration d'intérêt semble permettre un calcul linéaire en N [92], cette méthode nécessite en fait l'hypothèse que les orientations de ces deux courbes ne sont anti-parallèles en aucun point et à aucun moment, hypothèse qui n'est pas remplie dans le cas de l'ADN surenroulé formant des plectonèmes [93].

La simulation de longues molécules nécessite donc une nouvelle approche du calcul de l'énergie de torsion, plus performante, qui est présentée au prochain chapitre.

2.4 Objectifs

Au regard de cet état de l'art, je me suis donné pour objectif d'implémenter la méthode de Monte-Carlo pour simuler le repliement d'une molécule d'ADN surenroulée (Chapitre 3). L'algorithme que j'ai mis au point introduit plusieurs avancées :

- Tout en conservant le bénéfice d'une implémentation efficace par la méthode de Monte-Carlo, il introduit la superhélicité, caractéristique essentielle de la structure des chromosomes bactériens.
- Il est rapide grâce à une approche locale de la conservation du linking number et de la détection des croisements du polymère ; une variante en calcul parallèle existe également ; tout ceci permet d'aborder le repliement de molécules d'ADN longues de plusieurs centaines de milliers de paires de bases.
- Il permet une mesure directe du couple exercé par la molécule.
- Il est validé par comparaison avec des expériences d'étirement et de torsion sur molécules uniques (Chapitre 4).
- Il offre un cadre approprié à l'étude des interactions à longues portées et notamment, la formation de domaines topologiquement indépendants 5.

Les principaux paramètres utilisés dans ce chapitre et les suivants sont résumés dans la table 2.1.

Symbole	Valeur typique	Description
Paramètres globaux		
L	0,7 à 13 μm (2,2 à 40 kb)	longueur de la molécule
N	70 à 1400	nombre de cylindres
Lk		linking number
Lk_0		Lk au repos (sans surenroulement)
σ	0 à 0,1	surenroulement relatif $(Lk - Lk_0)/Lk_0$
Tw		twist (nombre d'hélices)
Tw_0		Tw au repos (sans surenroulement)
Wr		writhe
R_g		rayon de giration
l_p	50 nm	longueur de persistance de l'ADN
C	86 nm	module de torsion
f	0,25 à 3 pN	force extérieure étirant l'ADN
χ	0,1 à 1	extension relative de l'ADN étiré
Paramètres locaux		
n	5 à 10	nombre de cylindres par l_p
r_e	1,5 à 3 nm	rayon effectif (répulsion électrostatique)
\vec{r}_i		position du cylindre i
\vec{t}_i		orientation du cylindre i
l	5 à 10 nm	longueur d'un cylindre (L/N)
tw_i		twist du cylindre i
tw_0	1,5 à 3	twist au repos par cylindre (Tw/N)

TABLE 2.1: Principaux paramètres utilisés et propriétés calculées dans les simulations.

Chapitre 3

Un algorithme adapté aux longues molécules

DANS CE CHAPITRE, je présente ce nouvel algorithme en mettant l'accent sur ses spécificités : une approche locale de la conservation du linking number et la détection des croisements de cylindres susceptibles de créer des nœuds ou de modifier le linking number. Je teste également la validité et les performances de cette implémentation, ainsi qu'une variante parallélisée capable de profiter d'une architecture multi-cœurs.

3.1 Calcul direct du linking number

Twist local

Afin de définir localement le twist de la molécule, un brin est enroulé en hélice autour de celle-ci. Ce brin représente l'un des deux brins de l'ADN, le second étant présumé constamment apparié au premier sur toute sa longueur. La clé de la conservation du linking number réside alors dans la continuité de ce brin : tant qu'il n'est pas coupé, la contrainte topologique sera respectée.

Localement, le twist tw_i du cylindre i correspond simplement au nombre de tours formés par ce brin autour de l'axe du cylindre (Fig. 3.1). Le twist total Tw de la chaîne est alors :

$$Tw = \sum_{i=1}^N tw_i \quad (3.1)$$

La nature discrète de la chaîne présente cependant une difficulté : à chaque intersection entre deux cylindres i et $i + 1$, le brin sera en général discontinu, c'est-à-dire que la position $\vec{b}_i(l)$ du brin à la « sortie » du cylindre i ne correspond pas à sa position $\vec{b}_{i+1}(0)$ à l'« entrée » du cylindre $i + 1$ (Fig. 3.2). La continuité peut être rétablie en liant ces deux extrémités par un morceau de brin supplémentaire, mais celui-ci ne doit pas contribuer au twist de la molécule, sans quoi le linking number ne serait plus conservé. Cette contrainte

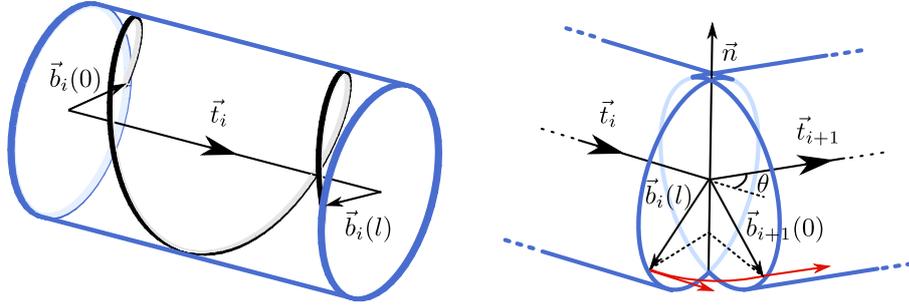


FIGURE 3.1: À gauche : schéma d'un cylindre de longueur l et d'axe \vec{t}_i , autour duquel est enroulé un brin définissant le twist tw_i (ici $\text{tw}_i = 1, 5$). À droite : schéma du joint entre deux cylindres formant un angle θ . Pour ne pas créer de twist à l'intérieur du joint, le brin (en rouge) passe de la position $\vec{b}_i(l)$ à $\vec{b}_{i+1}(0)$ en tournant autour du vecteur \vec{n} normal aux axes des cylindres.

est respectée en imposant au brin de tourner « à plat » autour de l'intersection, c'est-à-dire dans le plan formé par les axes \vec{t}_i et \vec{t}_{i+1} des deux cylindres (Fig. 3.1). Plus précisément, en notant $\vec{n} = \frac{\vec{t}_i \times \vec{t}_{i+1}}{|\vec{t}_i \times \vec{t}_{i+1}|}$ la normale à ce plan et $\theta = (\vec{t}_i, \vec{t}_{i+1})$ l'angle entre les deux axes, la position $\vec{b}_{i+1}(0)$ du brin doit correspondre à la rotation de $\vec{b}_i(l)$ autour de \vec{n} , d'un angle θ .

En effet, de cette manière, la direction du brin dans le joint est à la fois parallèle à \vec{t}_i à l'extrémité de i et parallèle à \vec{t}_{i+1} à l'extrémité de $i+1$ (tangentes indiquées par des flèches rouges sur la figure 3.1). Or, c'est précisément lorsque le brin est parallèle à l'axe de la molécule que sa contribution au twist est nulle.

Continuité du brin et rotations

Cette contrainte imposée au brin aux intersections des cylindres prend tout son sens lorsque l'on considère les rotations élémentaires de la simulation de Monte Carlo. Au cours de la rotation d'un bloc quelconque dont les extrémités sont les cylindres i et j , le brin contenu dans le bloc subit évidemment le même mouvement que les cylindres. Un problème se pose alors aux extrémités du bloc : l'axe \vec{t}_i ayant bougé par rapport à \vec{t}_{i-1} , la normale à leur plan ainsi que leur angle relatif ont également changé, ce qui signifie que la contrainte décrite précédemment n'est *a priori* plus respectée pour cette intersection (Fig. 3.2). Il en va bien sûr de même entre les cylindres j et $j+1$.

Il faut alors modifier la position du brin au niveau de ces intersections pour ne pas introduire de twist dans le joint. Plusieurs manières sont possibles : il est par exemple possible de garder fixe la position du brin à l'extrémité d'un cylindre (choisi au hasard) et calculer la nouvelle position du brin à l'extrémité de l'autre cylindre, par rotation autour de la normale \vec{n} . Dans ce cas, le twist du second cylindre varie. Cette variation peut facilement être calculée : à un facteur 2π près, elle est égale à l'angle orienté entre l'ancienne et la nouvelle position du brin, compris entre $-\pi$ et π (l'orientation étant donnée par l'axe du cylindre, \vec{t}_{i-1} ou \vec{t}_i). Cette méthode suppose donc que la variation du twist est

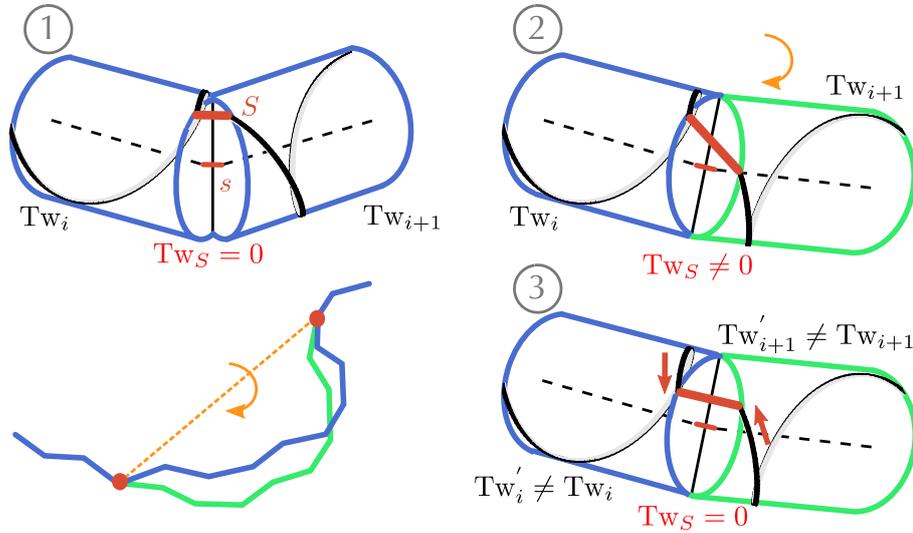


FIGURE 3.2: Conservation de la continuité du brin au cours de la rotation d'un bloc (en bas à gauche). 1 : L'enroulement du brin (ruban noir) autour de l'axe des cylindres (poillés noirs) définit leur twist. La continuité du brin impose que le segment S joignant leurs extrémités doit être parallèle à s (en rouge), afin que le twist Tw_s dans l'intersection soit nul. 2 : Au cours d'une rotation, s et S sont modifiés et ne sont plus alignés en général. Du twist a été créé à l'intersection. 3 : La continuité du brin est restaurée en étirant (ou contractant) le brin pour aligner à nouveau s et S . Le twist des deux cylindres est alors modifié mais peut être facilement calculé par l'angle entre les positions initiale et finale du brin.

inférieure à un demi-tour (en valeur absolue) mais cette hypothèse est toujours vérifiée puisque l'amplitude des rotations est elle-même limitée.

Il est également possible de faire varier la position du brin aux deux extrémités de l'intersection pour, par exemple, répartir équitablement la différence de twist entre les deux cylindres. Lorsque l'énergie de torsion est globale, la méthode choisie importe peu puisque la variation totale de twist sera toujours la même. En revanche, lorsqu'il est nécessaire d'utiliser une énergie de torsion locale (voir Section 2.3), la variation d'énergie, et donc le taux de rejet des mouvements, dépend de la manière de répartir la variation de twist dans les cylindres. De plus, la méthode choisie doit vérifier la relation de bilan détaillé afin de ne pas introduire de biais dans la méthode de Monte Carlo.

Conservation du linking number

Une méthode de conservation du linking number repose sur la possibilité de calculer, lors de toute rotation, ou bien la variation de writhe, ou bien la variation de twist. Par construction, la détermination d'une de ces quantités compense exactement l'autre. Grâce à la méthode décrite ci-dessus et résumée par la figure 3.2, la variation totale du twist au cours de la rotation d'un bloc est simplement égale à la somme des variations aux extrémités du bloc et implique

donc au plus les quatre cylindres pivots, suivant la méthode de répartition choisie. La conservation du linking number peut alors se vérifier en calculant explicitement le writhe. La Fig. 3.3 montre effectivement que l'écart relatif entre le linking number initial du polymère et la somme $Tw + Wr$ après un nombre quelconque de pas Monte Carlo est extrêmement faible (les fluctuations autour de zéro étant dues à l'imprécision des calculs sur les nombres à virgule flottante, qui introduisent nécessairement quelques arrondis et propagations d'erreur).

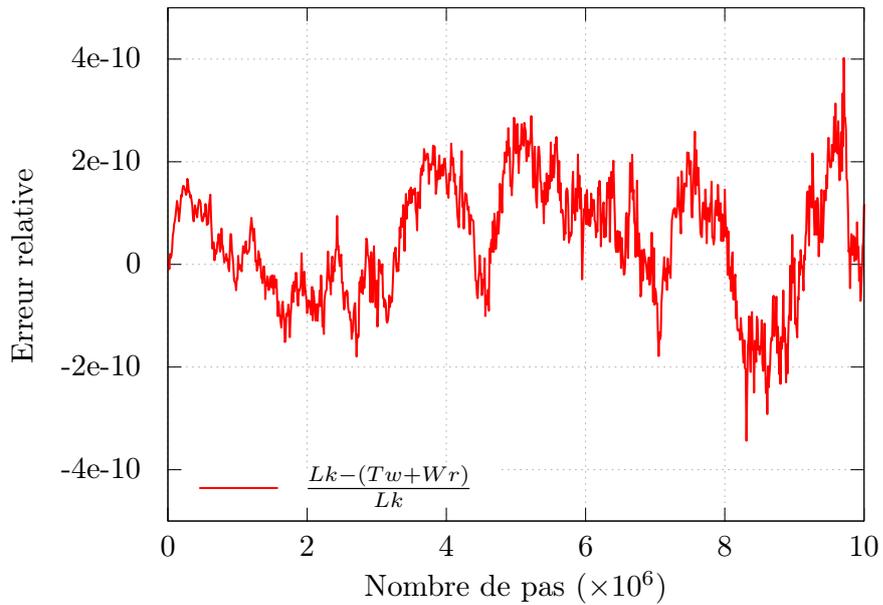


FIGURE 3.3: Erreur relative sur la valeur du linking number au cours d'une simulation. Le linking number Lk est un entier fixé au début de la simulation, le twist Tw est mis à jour à chaque pas par la méthode décrite dans la section 3.1 et le writhe Wr est calculé par la double intégrale usuelle [91] (Section 2.3).

Molécules linéaires

Le linking number n'étant défini que pour des courbes fermées, les molécules linéaires ne sont pas soumises en principe à la même contrainte topologique que les molécules circulaires. De fait, il est nécessaire d'utiliser des outils tels que des pinces magnétiques pour maintenir un niveau de surenroulement non nul dans une molécule d'ADN linéaire.

Cylindres virtuels

Il est cependant possible de modéliser l'action des pinces magnétiques à l'aide de cylindres virtuels, attachés aux extrémités du polymère. L'idée est de rendre la molécule circulaire « à l'infini » (en pratique à l'extérieur de la boîte dans laquelle le polymère est astreint à se mouvoir, voir Section 3.2), ce qui permet de connecter entre elles les deux extrémités du brin et donc de définir

et conserver le linking number. la méthode décrite précédemment (Section 3.1) pour assurer la continuité du brin s'applique alors aussi aux cylindres virtuels, en particulier aux extrémités de la molécule. Les cylindres virtuels peuvent ainsi échanger du twist avec la molécule.

La manière la plus simple d'imposer un surenroulement σ , dans le cas d'une énergie globale, consiste à décaler le minimum de l'énergie de torsion (Section 2.3) :

$$E_T = 2\pi^2 k_B T \frac{C}{L} (\text{Tw} - \text{Tw}_0 + \sigma \text{Lk})^2 \quad (3.2)$$

où σLk est le nombre de tours ajoutés (non nécessairement entier) et Tw est le twist total du système (molécule et cylindres virtuels).

Mais ces cylindres virtuels permettent de définir un réservoir de twist, mimant ainsi le piège de torsion utilisé dans les expériences de pinces magnétiques mesurant directement le couple [75]. Il suffit de scinder le twist total du système en deux composantes : $\text{Tw} = \text{Tw}_M + \text{Tw}_R$, où Tw_M est le twist de la molécule et Tw_R est le twist du réservoir. Le surenroulement est maintenant pris en compte dans l'énergie E_R associée au réservoir et non plus dans l'énergie de torsion globale E_T :

$$E_T = 2\pi^2 k_B T \frac{C}{L} (\text{Tw}_M - \text{Tw}_0)^2 E_R = 2\pi^2 k_B T k_R (\delta \text{Tw})^2 \quad (3.3)$$

avec k_R la constante de raideur du réservoir et $\delta \text{Tw} = \text{Tw}_R + \sigma \text{Lk}$. Dans le cas de l'énergie locale, il est même nécessaire d'utiliser ce réservoir, sans quoi le twist des cylindres virtuels ne serait pas pris en compte dans l'énergie et ceux-ci pourraient ainsi stocker indéfiniment le surenroulement ajouté au système.

Ce réservoir, qui représente donc la composante de torsion des pinces magnétiques utilisées expérimentalement, va subir une réaction de la part de la molécule, caractérisée par le couple $\Gamma = 2\pi k_R \langle \delta \text{Tw} \rangle$. Les cylindres virtuels, associés au calcul local du twist, permettent donc de déterminer le couple de la molécule (Section 4.1), ce qui n'est pas possible avec un algorithme calculant le twist global à partir du writhe.

Disposition

Plusieurs contraintes sont à prendre en compte pour mettre en place les cylindres virtuels. Tout d'abord, le linking number devant être conservé, la molécule ne doit pas passer au travers des cylindres virtuels (et ceux-ci doivent également s'éviter entre eux). Cette contrainte est assez facilement respectée en interdisant aux cylindres de la molécule de dépasser suivant z (la direction de la force \vec{f}) les limites définies par ses extrémités : $z(0)$ et $z(L)$, représentées par les parois rouges sur la figure 3.4. Ces deux parois représentent la surface et la bille auxquelles la molécule est attachée dans les expériences [41].

La disposition des cylindres virtuels est indiquée sur la figure 3.4, et permet de simplifier les échanges de twist. À chaque extrémité de la molécule, une série de trois cylindres virtuels est attachée, chacun selon une direction fixe et perpendiculaires deux à deux, ce qui permet aux extrémités de se déplacer librement dans toutes les directions, sans que les cylindres virtuels n'échangent de twist entre eux (ils s'étirent et se contractent, chacun selon sa direction, pour accommoder les déplacements des extrémités). Un septième cylindre, fixe (le numéro 4), permet de fermer la chaîne afin de calculer son writhe, par

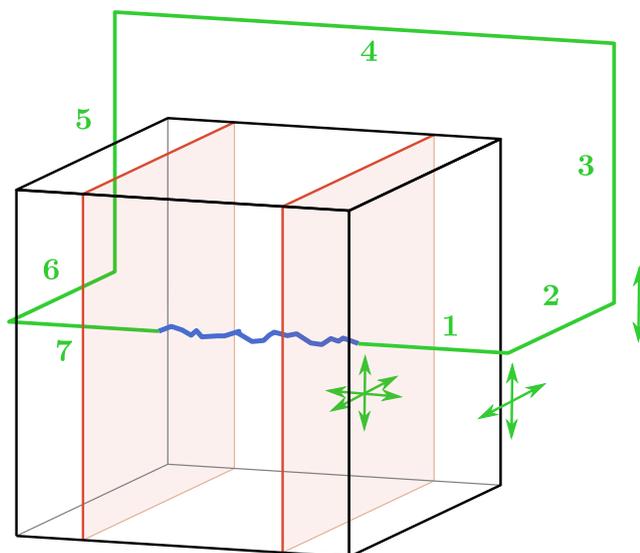


FIGURE 3.4: Disposition des cylindres virtuels (en vert) permettant de fermer la molécule linéaire (en bleu) à l'extérieur de la boîte (en noir). Les degrés de liberté des joints sont représentés (d'un seul côté) par les flèches vertes; le cylindre 4 est fixe. Les extrémités de la molécule définissent deux parois impénétrables (en rouge), perpendiculaires à la force de traction.

exemple, à des fins de vérifications. Grâce à cette disposition, en pratique, seuls deux cylindres virtuels sont véritablement utiles pendant les simulations (le premier et le septième), car seuls ces deux-ci échangent du twist avec la molécule.

3.2 Mouvements invalides

Au cours d'un mouvement élémentaire, avant même de calculer l'énergie de la nouvelle conformation, il est nécessaire de vérifier que cette dernière est valide du point de vue des contraintes géométriques et topologiques appliquées. Trois phénomènes peuvent donner lieu à une conformation invalide : les conditions aux bords, une collision entre deux cylindres ou le croisement de la chaîne par un cylindre (qui peuvent provoquer des nœuds ou faire varier le linking number).

Conditions aux bords

Les conditions aux bords utilisées ici sont dures, c'est-à-dire que le polymère est astreint à se déplacer à l'intérieur d'une boîte parallélépipédique. Outre la simplification d'un certain nombre de calculs, notamment celui du voisinage des cylindres (pour tester les autres conditions de rejet), du barycentre de la molécule, ou encore du writhe, ces conditions aux bords peuvent permettre d'étudier l'effet du confinement sur le repliement de l'ADN surenroulé.

Pour une molécule linéaire, le long de la direction de la force de tension, les parois sont mobiles et attachées aux extrémités de la chaîne afin d'empêcher un quelconque cylindre de passer à travers l'un des cylindres virtuels (Section 3.1).

Ainsi, lorsque le polymère a dérivé jusqu'au bord de son volume autorisé, une rotation peut amener un ou plusieurs cylindres de l'autre côté de la paroi. Ce mouvement sera alors automatiquement rejeté.

Auto-évitement de la chaîne

Les collisions entre cylindres doivent être détectées et rejetées afin de respecter la contrainte d'auto-évitement représentant la répulsion électrostatique entre deux segments d'ADN, modélisée par un potentiel de cœur dur de rayon r_e .

Plus exactement, on veut tester l'intersection des volumes formés par les cylindres auxquels sont ajoutés deux hémisphères aux extrémités. En effet, considérer des cylindres arrondis est à la fois un peu plus simple que des cylindres tronqués à la perpendiculaire, et permet également d'avoir une modélisation plus régulière du polymère (on obtient un tube au lieu d'une chaîne de cylindres entrecoupée de « trous » abrupts risquant de piéger d'autres cylindres).

Nécessité d'une grille

Une approche naïve consistant à tester, après chaque rotation, tous les couples de cylindres n'est pas du tout efficace : le nombre de couples étant proportionnel à N^2 (où N est le nombre de cylindres), les efforts accomplis pour s'affranchir du calcul du writhe auraient été inutiles.

Même en ne considérant que les couples dont les membres ont bougé l'un par rapport à l'autre (c'est-à-dire les couples composés d'un cylindre appartenant au bloc qui a tourné et d'un cylindre n'y appartenant pas ; les autres couples ne pouvant pas être entrés en collision durant cette rotation-ci), le problème reste entier : la taille moyenne d'un bloc étant proportionnelle à N , de même que la taille moyenne du complémentaire du bloc, cette optimisation ne fait que diminuer le préfacteur et pas l'exposant.

Puisque seul le voisinage d'un cylindre importe, la solution consiste à découper le volume total en une grille de cases cubiques aux dimensions judicieusement choisies.

Étant donné un cylindre i dont le centre \vec{c}_i se retrouve dans la case \mathcal{C} , il faut au moins tester les cylindres se trouvant dans cette case ainsi que dans toutes ses (vingt-six) plus proches voisines (c'est-à-dire un cube de trois cases de côté centré sur \mathcal{C}). En effet, \vec{c}_i pouvant *a priori* être aussi proche que possible d'une face, d'une arête ou d'un sommet de \mathcal{C} , le cylindre i est susceptible d'intercepter un cylindre dont le centre se trouve dans une case adjacente, y compris en diagonale. Il est possible d'utiliser des cases suffisamment grandes pour qu'il ne soit pas nécessaire de regarder les voisines plus éloignées de \mathcal{C} . C'est le cas si la largeur des cases est supérieure à $l + 2r_e$.

Utiliser des cases plus grandes que cette taille minimale semble, à première vue, réduire l'efficacité de l'algorithme puisque chaque case contiendra plus de cylindres en moyenne, ce qui résultera en un plus grand nombre de tests de collision à chaque pas. Il faut cependant noter que plus les cases sont petites, plus elles sont nombreuses. Le gain en temps de calcul se paye donc au prix d'une

plus grande utilisation de la mémoire, ce qui peut devenir problématique pour de très grandes molécules (les dimensions de la boîte augmentant typiquement comme N^3).

Pour résoudre ce dilemme, il serait possible d'utiliser, par exemple, des cases de taille variable, capables de fusionner ou de se séparer au gré des fluctuations de leur contenu : la boîte serait majoritairement composée d'un petit nombre de grandes cases vides, mais le polymère resterait réparti dans autant de petites cases que nécessaire.

Toutefois, la nécessité de détecter les croisements (Section 3.2) limite fortement la nécessité d'avoir des cases aussi petites que possibles, et donc l'intérêt de mettre en place une grille plus complexe.

Collisions entre cylindres

Ainsi, pour chaque cylindre i ayant subi une rotation, il suffit de déterminer la case dans laquelle ce cylindre aboutit, puis de dresser la liste de tous les cylindres j avec lesquels i est susceptible d'être entré en collision : il s'agit des cylindres du voisinage n'ayant pas subi la rotation. Il faut noter que les cylindres trop proches le long de la chaîne, qui se recouvrent nécessairement, doivent également être ignorés. En particulier, deux cylindres consécutifs se recouvrent toujours autour de leur extrémité commune. Sont également impliqués tous les cylindres dont le diamètre effectif est plus grand que la longueur. Plus précisément, il faut ignorer tous les k -ièmes voisins tels que $kl < 2r_e < (k+1)l$ ($k \in \mathbb{N}$).

Pour savoir si deux cylindres i et j , non voisins, se recouvrent effectivement, l'idée est de se placer dans le plan \mathcal{P} défini par \vec{t}_i et \vec{t}_j , passant par \vec{r}_i . L'axe du cylindre i appartient donc à ce plan, et l'axe du cylindre j lui est parallèle. Si ce cylindre est trop éloigné du plan, il ne peut y avoir de collision. S'il est suffisamment proche, en revanche, il faut chercher à savoir si sa projection sur ce plan coupe le cylindre i (Fig. 3.5), ou bien si la projection d'une de ses extrémités est suffisamment proche de i (ou, de façon symétrique, si les extrémités de i sont suffisamment proche de la projection de j).

Plus précisément, en notant $r_{ij}^{\vec{r}} = \vec{r}_j - \vec{r}_i$ la position de j par rapport à i , et $\vec{n} = \vec{t}_i \times \vec{t}_j$ le vecteur normal au plan \mathcal{P} , la distance de l'axe de j au plan vaut $r_{ij}^{\vec{r}} \cdot \vec{n}$. Les deux axes sont alors trop éloignés l'un de l'autre pour qu'il y ait recouvrement si $r_{ij}^{\vec{r}} \cdot r_{ij}^{\vec{r}} > 2r_e$.

Sinon, il faut projeter le cylindre j dans le plan \mathcal{P} . Sa position devient $\vec{\rho} = r_{ij}^{\vec{r}} - (r_{ij}^{\vec{r}} \cdot r_{ij}^{\vec{r}}) r_{ij}^{\vec{r}}$, tandis que son orientation \vec{t}_j n'est, par construction, pas modifiée. La question est de déterminer la position de l'intersection I des deux axes, pour savoir si celle-ci est à l'intérieur des segments i et j . En appelant u la distance entre \vec{r}_i et I et v la distance entre r_{ho} et I , le problème est donc de savoir si $0 < u < l$ et $0 < v < l$. En notant θ l'angle (\vec{t}_i, \vec{t}_j) et en définissant le vecteur unitaire $\vec{\tau} = \frac{\vec{t}_i - (\vec{t}_i \cdot \vec{t}_j) \vec{t}_j}{\sqrt{1 - (\vec{t}_i \cdot \vec{t}_j)^2}}$, perpendiculaire à \vec{t}_j , il est possible d'écrire la relation trigonométrique suivante :

$$\begin{aligned} u \sin \theta &= u \sqrt{1 - (\vec{t}_i \cdot \vec{t}_j)^2} \\ &= \vec{\rho} \cdot \vec{\tau} \\ &= r_{ij}^{\vec{r}} \cdot \vec{\tau} \end{aligned} \tag{3.4}$$

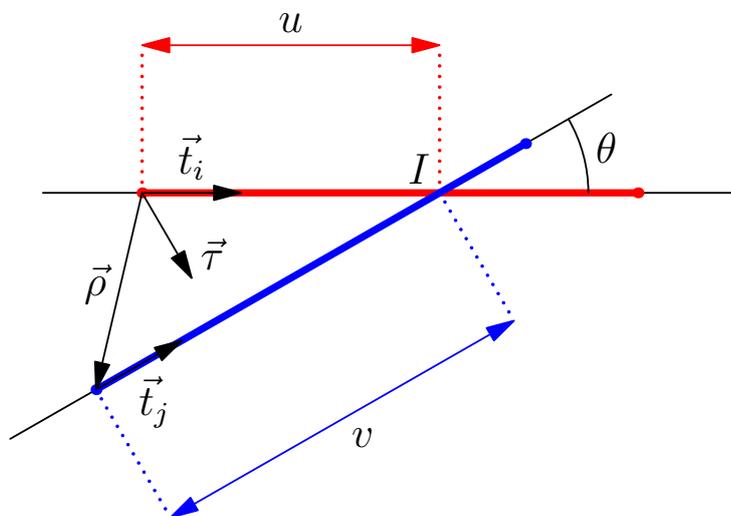


FIGURE 3.5: Projection d'un recouvrement dans le plan \mathcal{P} formé par les axes des deux cylindres et passant par \vec{r}_{ij} . Les cylindres i et j se recouvrent si l'intersection I de i (en rouge) et la projection de j sur \mathcal{P} (en bleu) se recouvrent est à l'intérieur des deux segments, ou si une extrémité d'un segment est trop proche de l'autre segment. Les notations sont décrites dans le texte.

D'où on déduit :

$$u = r_{ij} \cdot \frac{\vec{t}_i - (\vec{t}_i \cdot \vec{t}_j) \vec{t}_j}{1 - (\vec{t}_i \cdot \vec{t}_j)^2} \quad (3.5)$$

De même, en changeant r_{ij} en $-r_{ij}$ et en échangeant les rôles de i et j :

$$v = -r_{ij} \cdot \frac{\vec{t}_j - (\vec{t}_i \cdot \vec{t}_j) \vec{t}_i}{1 - (\vec{t}_i \cdot \vec{t}_j)^2} \quad (3.6)$$

Ainsi, si $0 < u < l$ et $0 < v < l$, les deux cylindres se coupent. Sinon, il est encore possible que l'extrémité de l'un soit suffisamment proche du deuxième. Il faut donc calculer les distances des deux extrémités de chaque cylindre à l'axe de l'autre cylindre et les comparer à $2r_e$.

Enfin, si les deux cylindres sont quasiment alignés, au point que \vec{n} est indéfini (ou que la précision des calculs risque d'être compromise), ce dernier test sur la distance entre les extrémités suffit.

Détection des croisements

Bien que l'amplitude des rotations soit limitée, certains cylindres peuvent parcourir une grande distance au cours d'un seul mouvement (notamment les cylindres situés au milieu d'un long bloc). Il arrive alors parfois qu'un cylindre passe à travers un autre (Fig. 3.6). Ces croisements peuvent donner lieu à des nœuds ou à une variation du linking number (de la même manière qu'une topoisomérase de type II qui casse le double-brin pour lui faire traverser un autre segment d'ADN).

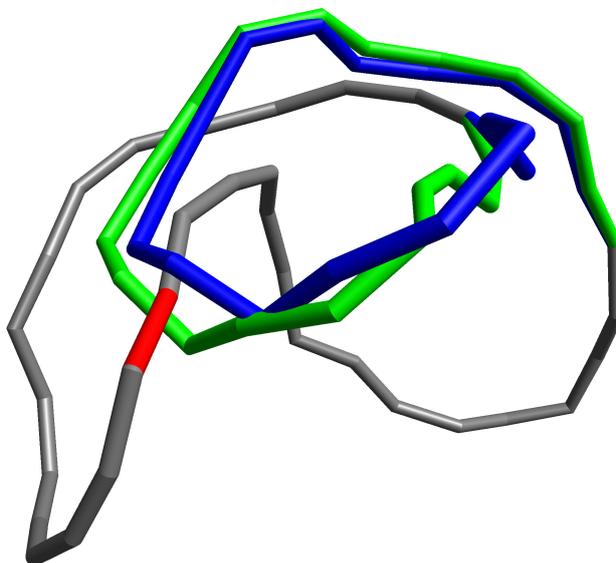


FIGURE 3.6: Exemple de croisement. Entre la position initiale (en bleu) et la position finale (en vert) du bloc subissant une rotation, ce dernier passe à travers le cylindre rouge. Dans cet exemple, le croisement est particulièrement problématique puisqu'il provoque un nœud.

Il est donc nécessaire de détecter ces événements afin de rejeter les conformations ne respectant pas les contraintes topologiques imposées à la molécule.

L'objectif ici est de voir si un cylindre i , au cours d'une rotation d'axe $\vec{\omega}$ et d'angle Ω , traverse un cylindre immobile j . On cherche donc les intersections entre l'axe de j et la surface balayée par l'axe de i (Fig. 3.7). Toutefois, si les deux extrémités du cylindre i ont parcouru moins de deux diamètres effectifs ($4r_e$), la détection des croisement n'est pas nécessaire car le mouvement aura déjà été rejeté par le test de recouvrement.

Surface balayée

La surface balayée est exprimée dans le repère cylindrique \mathcal{R}' associé à la rotation des deux extrémités du cylindre i (Fig. 3.8). L'origine O' de \mathcal{R}' est placée sur l'axe de rotation, par exemple sur l'un des pivots du bloc. L'axe \vec{z}' est défini par l'axe de rotation : $\vec{z}' = \vec{\omega}$. L'axe \vec{x}' est choisi comme le projeté du centre du cylindre \vec{c}_i orthogonal à \vec{z}' , soit, en notant $\vec{c}_i^* = \vec{c}_i - OO'$ et $\vec{c}'_i = \vec{c}_i^* - (\vec{c}_i^* \cdot \vec{z}')\vec{z}'$, alors $\vec{x}' = \frac{\vec{c}'_i}{|\vec{c}'_i|}$. Enfin, $\vec{y}' = \vec{z}' \wedge \vec{x}'$.

Pour caractériser l'appartenance d'un point à la surface, il est judicieux d'utiliser ses coordonnées cylindriques (ρ, θ, z) dans \mathcal{R}' . Au cours de la rotation, les points de l'axe du cylindre i parcourent une trajectoire à z et $\rho(z)$ constants, tandis que θ passe de $\theta_i(z)$ à $\theta_i(z) + \Omega[2\pi]$ (Fig. 3.8). Ainsi, un point $P = (\rho, \theta, z)$

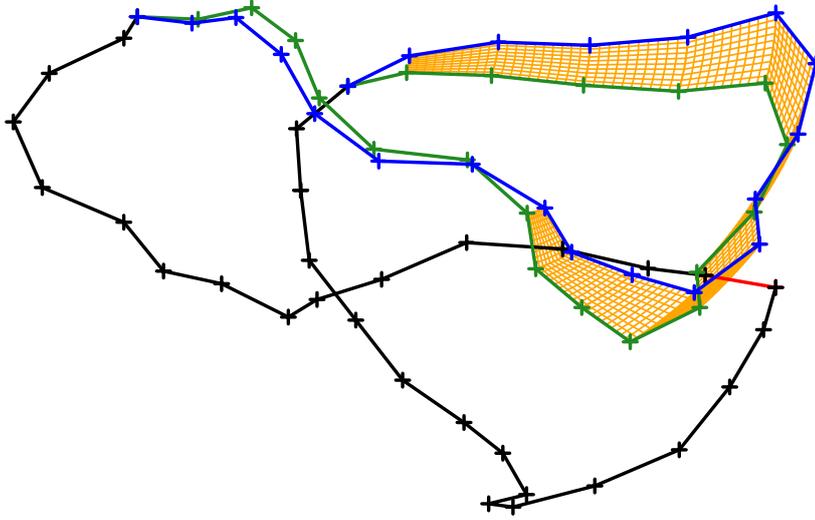


FIGURE 3.7: Le croisement de la Fig. 3.6 vu sous un autre angle. En jaune, la surface balayée par le bloc au cours de sa rotation (elle n'a pas été calculée pour les segments ayant suffisamment peu bougé pour qu'un croisement soit impossible, ce qui arrive généralement vers les bords du bloc). Cette surface est traversée par le segment rouge.

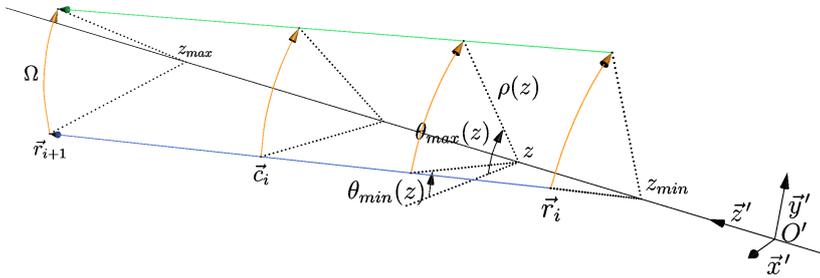


FIGURE 3.8: Schéma du repère \mathcal{R}' associé à la rotation du cylindre i (flèches orange). La surface balayée par l'axe du cylindre au cours d'une rotation d'angle Ω (du segment bleu au vert) est délimitée par les coordonnées z_{min} et z_{max} des extrémités. Pour un z donné dans cet intervalle, cette surface possède un ρ constant et θ varie de $\theta_{min}(z)$ à $\theta_{max}(z) = \theta_{min}(z) + \Omega$.

appartient à la surface balayée aux conditions suivantes :

$$\begin{aligned}
 & z_{min} < z < z_{max} \\
 & \rho = \rho(z) \\
 & \theta_{min}(z) < \theta < \theta_{max}(z) \quad \text{si } \theta_{min}(z) < \theta_{max}(z) \\
 & \theta < \theta_{min}(z) \text{ ou } \theta > \theta_{max}(z) \quad \text{sinon}
 \end{aligned} \tag{3.7}$$

Où z_{min} et z_{max} sont les coordonnées des extrémités $\vec{r}'_k, \vec{r}'_{k+1}$ du cylindre, avec $z_{min} < z_{max}$ et $\theta_{min}(z)$, et $\theta_{max}(z)$ sont définis suivant le sens de la rotation :

- Si $\Omega > 0$, $\theta_{min}(z) = \theta_i(z)$ et $\theta_{max}(z) = \theta_i(z) + \Omega[2\pi]$.
- Si $\Omega < 0$, $\theta_{min}(z) = \theta_i(z) + \Omega[2\pi]$ et $\theta_{max}(z) = \theta_i(z)$.

Cases impliquées

Le but est de trouver tous les cylindres suffisamment proches de la surface pour que le test de croisement soit utile. Pour cela, il est utile de considérer le critère suivant : un cylindre j dont le centre se situe à une distance supérieure à l de l'arc de cercle formé par la rotation du centre du cylindre i , ne peut en aucun cas traverser la surface balayée par le cylindre i . Il faut donc considérer toutes les cases susceptibles de contenir un cylindre proche de cet arc de cercle, c'est-à-dire dont la distance à l'arc est inférieure à $l + r_{case}$, et faire un test de croisement avec tous les cylindres de ces cases (sauf ceux qui appartiennent au bloc ou qui sont trop proches voisins du cylindre i le long de la chaîne).

Malheureusement, la taille des blocs augmentant avec N , la distance moyenne parcourue par les cylindres augmente également, ainsi que la surface balayée et le nombre de cases et de cylindres à tester. C'est en fait la détection des croisements qui est la partie la plus coûteuse de l'algorithme (Section 3.2).

Traitement des croisements

Dans \mathcal{R}' , l'équation de la droite supportant l'axe du cylindre j s'écrit, en coordonnées cylindriques :

$$\begin{aligned}\rho_j(z) \cos \theta_j(z) &= x'_j + \frac{t'_{jx}}{t'_{jz}}(z - z'_j) \\ \rho_j(z) \sin \theta_j(z) &= y'_j + \frac{t'_{jy}}{t'_{jz}}(z - z'_j)\end{aligned}\tag{3.8}$$

Ce cylindre coupe la surface balayée par i en deux points au plus, dont les coordonnées z sont données par l'équation $\rho_i(z) = \rho_j(z)$, que l'on peut écrire (en utilisant l'équation 3.8) :

$$\begin{aligned}\rho_j^2(z) - \rho_i^2(z) &= (x'_j + \frac{t'_{jx}}{t'_{jz}}(z - z'_j))^2 + (y'_j + \frac{t'_{jy}}{t'_{jz}}(z - z'_j))^2 \\ &\quad - (x'_i + \frac{t'_{ix}}{t'_{iz}}(z - z'_i))^2 - (y'_i + \frac{t'_{iy}}{t'_{iz}}(z - z'_i))^2 \\ &= 0\end{aligned}\tag{3.9}$$

Une solution z^* de cette équation (si elle existe), donne les coordonnées cylindriques de l'intersection ($\rho^* = \rho_i(z^*) = \rho_j(z^*)$, $\theta^* = \theta_j(z^*)$, z^*). Après avoir calculé cette solution, il faut encore vérifier qu'elle remplit les conditions 3.7, ainsi que la condition d'appartenance au cylindre j : $z_j < z^* < z_{j+1}$ (ou $z_{j+1} < z^* < z_j$ si $z_{j+1} < z_j$). Si c'est le cas, un croisement a effectivement eu lieu et vient d'être détecté. Si les deux solutions de l'équation vérifient ces conditions, cependant, cela est équivalent à ne pas avoir fait de croisement du tout, et le programme n'en tient pas compte.

Lorsqu'un croisement est détecté, il n'est pas nécessairement synonyme de nœud ou de variation du linking number. Il est possible de tester explicitement l'apparition d'un nœud (Section 3.2), quoique cette routine soit coûteuse. En comparant les angles relatifs des axes des cylindres avant et après le croisement, au niveau de l'intersection, il est possible de connaître l'éventuelle variation de writhe due au croisement (qui, si elle est non nulle, vaudra ± 2 , voire un multiple

de ± 2 dans le cas de plusieurs croisements simultanés). Si cette variation est non nulle, elle peut être compensée en modifiant le twist dans certains cylindres (tirés au hasard par exemple) afin de conserver le linking number, ou bien le mouvement peut tout simplement être rejeté.

Il est aussi possible (et surtout plus simple) de rejeter systématiquement un mouvement si un croisement a lieu. Cette méthode a l'inconvénient de refuser des conformations pouvant être légitimes, mais l'avantage d'éviter de fastidieuses vérifications de la topologie du système, et présente peu de risques d'introduire un biais puisque si un croisement a lieu entre une configuration A et une configuration B , il aura également lieu lors du mouvement de B vers A . Ainsi, j'ai pu vérifier que la méthode choisie n'a pas d'effet sur les résultats.

Détection des nœuds

Si les croisements ne sont pas systématiquement rejetés, il devient nécessaire de vérifier que les configurations générées ne forment pas de nœud. Ce problème se résume au calcul d'un invariant, une quantité indépendante de la géométrie exacte du nœud (invariante par déformation du nœud). Bien qu'aucun invariant absolu ne soit connu pour distinguer avec certitude deux nœuds différents, ni même pour distinguer le nœud trivial (équivalent à un cercle) des nœuds plus complexes, il existe cependant plusieurs invariants « faibles » utiles en pratique [94].

En particulier, le polynôme d'Alexander [95] permet de distinguer le nœud trivial d'un grand nombre des nœuds les plus simples. En effet, le premier nœud non-trivial ayant le même polynôme d'Alexander que le nœud trivial possède onze croisements [96, 97], ce qui correspond à un enchevêtrement difficile à obtenir au cours d'une simulation. D'autres nœuds plus simples partagent un même polynôme d'Alexander, mais le problème à résoudre ici étant seulement de distinguer le nœud trivial des autres nœuds, cet invariant partiel est tout-à-fait suffisant.

Ce polynôme se calcule à partir d'un diagramme du nœud considéré (une projection du nœud dans un plan). Les croisements formés par ce diagramme servent à construire une matrice de polynômes, la matrice d'Alexander, dont le déterminant permet d'obtenir le polynôme d'Alexander, qui se résume au monôme 1 dans le cas du nœud trivial.

Plus précisément, pour savoir si une configuration donnée est nouée (Fig. 3.9), il faut tout d'abord projeter les cylindres dans le plan (O, \vec{y}, \vec{z}) pour former un diagramme du nœud (Fig. 3.10). Les croisements de ce diagramme sont alors répertoriés en cherchant, pour chaque cylindre i , tous les cylindres d'indice $j > i$ dont la projection coupe celle de i . Il suffit pour cela de ne regarder que les cases alignées selon \vec{x} avec celle contenant i , ainsi que leurs plus proches voisines suivant \vec{y} et \vec{z} . Les positions u et v du croisement le long des droites supportants les axes de i et j , respectivement, sont données par :

$$u = \frac{(y_j - y_i)t_{jz} - (z_j - z_i)t_{jy}}{t_{iy}t_{jz} - t_{iz}t_{jy}} \quad (3.10)$$

$$v = \frac{(y_j - y_i)t_{iz} - (z_j - z_i)t_{iy}}{t_{iy}t_{jz} - t_{iz}t_{jy}} \quad (3.11)$$

Les projections des cylindres i et j se coupent si et seulement si $0 < u < l$ et $0 < v < l$. Lors de la projection, le cylindre i se retrouve au-dessous de j si

$\vec{r}_i + u\vec{t}_i < \vec{r}_j + v\vec{t}_j$, et au-dessus sinon. On repère alors le croisement par le cylindre situé au-dessous et par la position du croisement le long de ce cylindre (u ou v), et on range les croisements par position croissante le long de la chaîne. On note N_c le nombre total de croisements. On appellera également « région a » le morceau de chaîne compris entre les croisements $a - 1$ et a (modulo N_c).

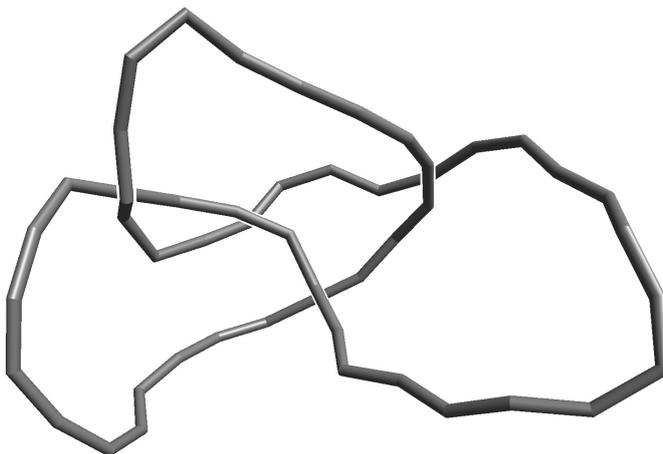


FIGURE 3.9: Nœud formé par la chaîne si l'on autorise le croisement de la Fig. 3.6. Il s'agit d'un nœud de trèfle, le plus simple des nœuds non-triviaux. Son polynôme d'Alexander vaut $x^2 - x + 1$.

Il faut ensuite construire la matrice d'Alexander \mathcal{M} et calculer son déterminant (Fig. 3.11). Les colonnes de cette matrice correspondent aux régions, et les lignes aux croisements. Pour chaque croisement a , si la région passant au-dessus est a ou $a + 1$, la ligne ne contient que deux éléments : $\mathcal{M}_{aa} = -1$ et $\mathcal{M}_{aa+1} = 1$. Sinon, elle en contient trois : $\mathcal{M}_{ab} = t - 1$ (où b est la région du dessus), et, suivant le type de croisement :

- $\mathcal{M}_{aa} = 1$ et $\mathcal{M}_{aa+1} = -x$ si $t_{iy}t_{jz} - t_{iz}t_{jy} > 0$ (a est à gauche de b)
- $\mathcal{M}_{aa} = -x$ et $\mathcal{M}_{aa+1} = 1$ si $t_{iy}t_{jz} - t_{iz}t_{jy} < 0$ (a est à droite de b)

La matrice d'Alexander consiste en une sous-matrice de taille $(N_c - 1) \times (N_c - 1)$ de la matrice décrite ci-dessus : une ligne et une colonne sont à retirer (le choix de la ligne ou de la colonne n'a pas d'effet sur le polynôme d'Alexander).

Le polynôme obtenu en calculant le déterminant de la matrice d'Alexander est alors égal au polynôme d'Alexander à un facteur $\pm x^{-k}$ près ($k \in \mathbb{N}$) : le polynôme d'Alexander possède par définition un terme d'ordre 0 non-nul et positif [97].

Cet algorithme étant coûteux (recherche des croisements sur le diagramme de nœud et calcul du déterminant d'une matrice de polynômes), il n'est véritablement utile que pour des petites molécules (quelques milliers de paires de bases au maximum). Les nœuds apparaissant relativement rarement, il est possible de ne faire ce calcul qu'une fois de temps en temps (par exemple tous les dix mille pas), et de revenir en arrière au besoin, en annulant tous

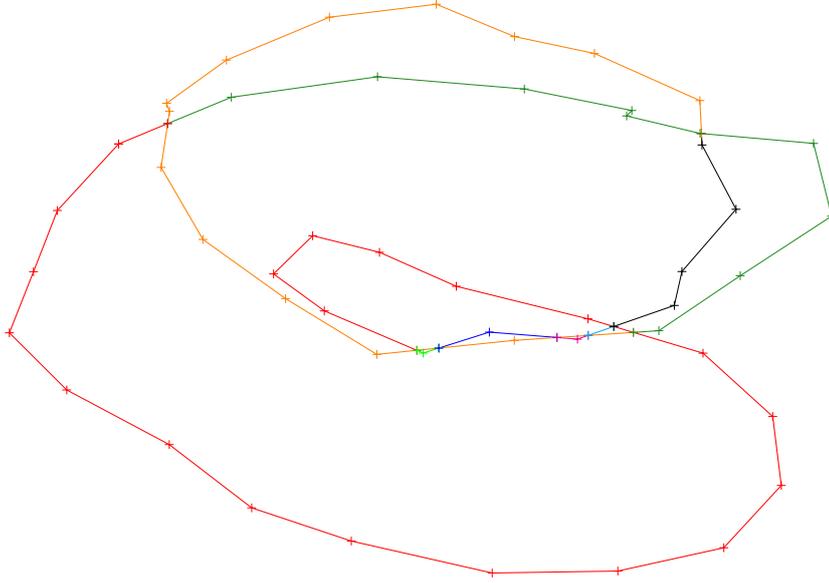


FIGURE 3.10: Diagramme du nœud de la Fig. 3.9. Chaque couleur représente une région différente, délimitée par deux croisements successifs.

$$\begin{pmatrix} 1 & -x & 0 & 0 & 0 & 0 & x-1 \\ 0 & -x & 1 & 0 & 0 & 0 & x-1 \\ 0 & 0 & 1 & -x & 0 & 0 & x-1 \\ 0 & 0 & 0 & -x & 1 & 0 & x-1 \\ x-1 & 0 & 0 & 0 & 1 & -x & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -x \\ x-1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

FIGURE 3.11: Matrice d'Alexander du nœud de la Fig. 3.9. Son déterminant vaut $x^5 - x^4 + x^3$, ce qui donne le polynôme d'Alexander $x^2 - x + 1$ du nœud de trèfle.

les mouvements effectués depuis la dernière vérification. Il est aussi beaucoup plus efficace de rejeter systématiquement les croisements, ceux-ci étant les seuls responsables de l'apparition des nœuds.

3.3 Performances

Complexité algorithmique

Afin de tester les performances de mon nouvel algorithme, j'ai mesuré la durée moyenne des simulations en fonction de la taille L de la molécule. À titre de comparaison, j'ai également mesuré cette durée pour l'algorithme basé sur le calcul du writhe et développé par A. Vologodskii, grâce à un programme aimablement fourni par G. Charvin. La figure 3.12 confirme que l'algorithme

basé sur le calcul du writhe résulte en une durée de simulation proportionnelle à L^2 (à cause du calcul du writhe nécessitant une double somme sur les cylindres, voir Section 2.3), tandis que mon approche locale conduit à un exposant 1,2.

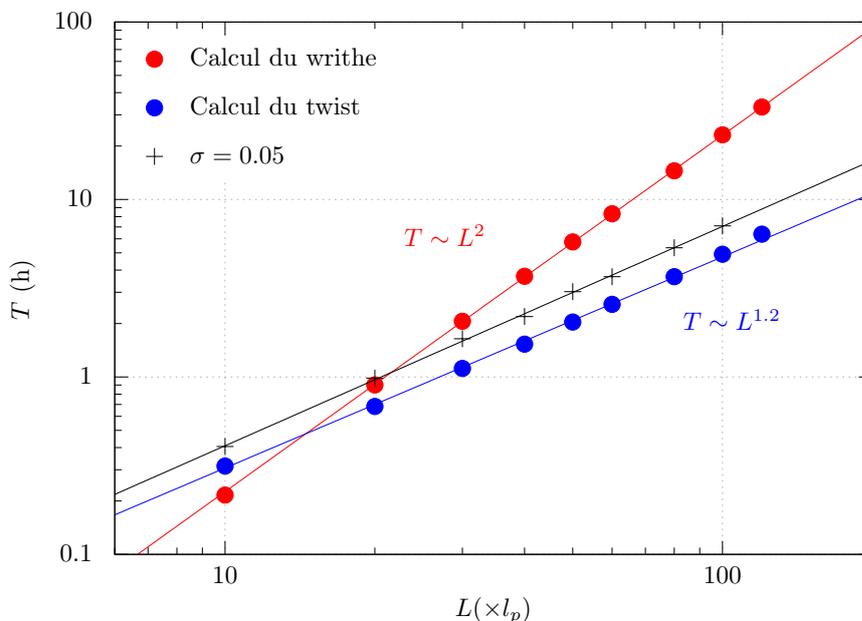


FIGURE 3.12: Durée moyenne T nécessaire pour simuler 10^7 pas en fonction de la longueur L de la molécule. Les droites correspondent aux meilleurs fits des points. L'algorithme calculant le writhe est quadratique en L , tandis que le calcul direct du twist (bleu) réduit l'exposant à 1,2. La simulation de molécules surenroulées (noir) est plus longue, mais ne modifie pas l'exposant.

Facteur limitant

C'est en fait la détection des croisements qui est responsable de cet exposant supérieur à 1. En effet, toutes les autres parties de l'algorithme (rotation des cylindres, test d'intersection, calcul de l'énergie, etc.) sont locales, au sens où elles n'utilisent que les propriétés du cylindre considéré ou de ses plus proches voisins (le long de la chaîne ou bien, à trois dimensions, à l'intérieur de la case dans laquelle celui-ci se trouve et des cases adjacentes). Si la détection des croisements n'était pas nécessaire, cet algorithme s'exécuterait donc en un temps proportionnel à L .

La détection des croisements, en revanche, nécessite de s'intéresser à tous les cylindres susceptibles de croiser la surface balayée au cours de la rotation. Or la distance typique parcourue par un cylindre, et donc la surface balayée par ce cylindre, augmente avec L : lorsque le nombre de cylindres augmente, la longueur moyenne des blocs subissant des rotations fait de même, ce qui augmente alors la distance typique d'un cylindre du bloc à l'axe de rotation, et par conséquent la distance qu'il parcourt au cours de son mouvement.

Il serait possible de limiter la distance parcourue par les cylindres, par exemple en limitant la taille maximale des blocs ou en réduisant l'angle des rotations, cependant l'évolution de la molécule s'en trouverait ralentie. Le gain en temps de calcul risquerait alors d'être réduit ou annulé par la nécessité d'augmenter le nombre de pas pour atteindre l'équilibre.

Variabilité dans le temps de calcul

De nombreux autres facteurs peuvent influencer le temps de calcul. La figure 3.12 montre notamment que les simulations sont plus longues lorsque la molécule est surenroulée. En effet, la formation de plectonèmes augmente la densité des cylindres et par conséquent la durée des tests de recouvrement et de croisements, tous deux coûteux. D'autres paramètres, tels que le rayon des cylindres ou la taille des cases de la grille, sont également susceptibles d'avoir une influence dans certaines situations. Toutefois, dans tous les cas, ces paramètres ne contribueront à la durée des simulations que d'un facteur multiplicateur, et ne modifient pas l'exposant.

Parallélisation

Afin d'accélérer encore plus les simulations, j'ai également développé une version du programme capable de tirer parti d'une architecture multi-cœurs (à l'aide de la bibliothèque *openMP* [98]). Cet algorithme se prête relativement bien à la parallélisation car les routines qui demandent le plus de ressources traitent un grand nombre de cylindres indépendamment les uns des autres. En effet, à chaque pas, une fois les paramètres de la rotation choisis, le même traitement s'applique à chaque cylindre du bloc : Rotation, puis test de recouvrement et de croisement avec des cylindres immobiles.

Dans sa version parallèle, le programme lance, à chaque pas, plusieurs processus (threads) qui traitent chacun une partie du bloc. Lorsque tous les processus ont terminé leur travail, et si le mouvement est valide, le programme calcule la variation de twist aux extrémités du bloc, la variation d'énergie et applique la méthode de Metropolis, puis prépare le mouvement suivant.

Les processus sont également capables de communiquer de façon basique entre eux : dès que l'un d'eux détecte une intersection ou un croisement, non seulement il s'arrête sans traiter les cylindres suivants, mais il signale également aux autres processus en cours de calcul qu'ils doivent eux aussi s'arrêter immédiatement. De cette manière, aucun processus ne continuera à traiter des cylindres pour un mouvement qui sera de toutes façons rejeté.

L'effet de la parallélisation est visible sur la figure 3.13. Le bénéfice augmente avec la taille de la molécule : pour $N = 50$ cylindres, utiliser six processus divise environ par deux le temps de calcul (de 15,4 min. à 6,7 min.), tandis que pour $N = 500$ cylindres, utiliser six processus divise ce temps par trois (de 3,64 h à 1,21 h).

Bien sûr, utiliser k processus ne divise pas par k le temps de calcul, du fait des parties non-parallèles du code ainsi que de l'interférence entre les processus (compétition pour accéder aux mêmes données en mémoire, par exemple). Par conséquent, lorsque de nombreuses simulations doivent être effectuées (par exemple pour caractériser un système de façon exhaustive), il est préférable de lancer simultanément plusieurs instances non-parallèles du programme plutôt

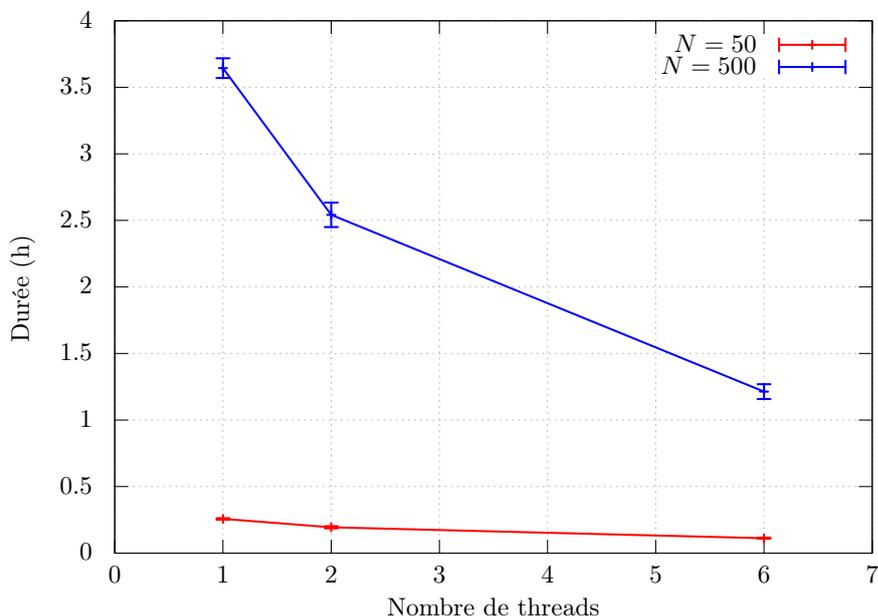


FIGURE 3.13: Durée moyenne nécessaire pour simuler 10^7 pas en fonction du nombre de processus utilisés, pour deux chaînes de $N = 50$ et $N = 500$ cylindres. Toutes ces simulations ont été réalisées sur un ordinateur possédant six cœurs physiques (douze cœurs logiques via l’Hyper-threading).

que de lancer l’une après l’autre des instances parallèles. Toutefois, la version parallèle est très utile pour obtenir rapidement une idée grossière des résultats, avant de lancer une étude plus précise nécessitant de nombreuses répliques.

Elle permet également d’entrevoir la possibilité de simuler un chromosome entier : celui de *Mycoplasma genitalium*, par exemple, long de 580 kb, nécessiterait environ 2 000 cylindres (avec $n = 5$ cylindres par l_p), soit autour de 7 h pour effectuer 10^7 pas avec six processus, tandis que le programme d’A. Vologodskii nécessiterait autour de deux mois et demi (de l’ordre de 2 000 h).

Implémentation

Afin d’implémenter cet algorithme, mon choix s’est porté sur le langage C++ pour plusieurs raisons :

- ce langage, comme le C, est réputé rapide et performant tout en restant relativement simple d’utilisation,
- il est très populaire et utilisé par de nombreux développeurs, ce qui signifie qu’il est assez facile d’obtenir de l’aide et de la documentation, et de trouver sur Internet des réponses à ses questions techniques.
- Toujours du fait de sa popularité, de nombreuses bibliothèques existent (sous licence libre pour beaucoup d’entre elles) pour ajouter des fonctionnalités à la bibliothèque standard, ou en faciliter l’utilisation. À ce titre, j’ai notamment utilisé la bibliothèque *Eigen* [99], spécialisée dans l’algèbre linéaire, pour manipuler la géométrie des cylindres et les ma-

trices de rotations associées aux mouvements élémentaires, ainsi que plusieurs des bibliothèques de la famille *boost* [100]. En particulier, *random* est utile pour la génération de nombres pseudo-aléatoires selon diverses distributions, et permet d'utiliser différents générateurs afin de s'assurer qu'un comportement inattendu des simulations n'est pas dû à un biais statistique de l'algorithme de génération utilisé. La *serialization* permet assez simplement de sauvegarder et de restaurer l'état d'un objet quelconque (une configuration particulière du polymère pour s'en servir de condition initiale dans une autre simulation, ou bien l'état d'une simulation dans son ensemble à un instant donné, sauvegardé sur le disque dur à intervalles régulier pour éviter d'avoir à recommencer une simulation de zéro si un problème survient : processus tué, machine redémarrée, etc.). J'ai également utilisé les *shared_pointers* qui simplifient la gestion des pointeurs, *filesystem* pour une manipulation facile et multi-plateformes des fichiers, et *program_options* et *ptree* pour la gestion de l'interface utilisateur et de la configuration des simulations. Enfin, comme évoqué précédemment, la parallélisation a été réalisée à l'aide de la bibliothèque *openMP* [98].

- Le C++ est un langage orienté objet dont l'une des principales caractéristiques, l'héritage des classes, m'a permis de simplifier certaines parties du code. Ainsi, les polymères circulaires et linéaires possédant leurs spécificités propres (par exemple, seul un polymère linéaire peut subir des rotations de ses extrémités autour d'un axe aléatoire), ils doivent être représentés par deux classes distinctes. Cependant, puisqu'ils partagent aussi de nombreux comportements et propriétés (par exemple, dans les deux cas, une fois choisie la rotation, il faut appliquer celle-ci à tous les cylindres du bloc et traiter la continuité du brin aux extrémités de ce bloc), avoir deux classes totalement indépendantes nécessiterait de recopier une grande partie des attributs et des fonctions de l'une vers l'autre, ce qui alourdirait le code source et augmenterait le risque d'erreur. L'héritage permet alors de définir une classe « polymère » de base, contenant le code commun aux polymères, et dont héritent les classes « polymère linéaire » et « polymère circulaire », chacune implémentant ses spécificités. De même, les cylindres virtuels permettant de clore une chaîne linéaire héritent de la classe « cylindre », dont ils partagent plusieurs propriétés (position, axe, longueur, twist, etc.), mais rédéfinissent la manière dont ils se comportent au cours d'un mouvement (étirement et translation plutôt que rotation).

Enfin, s'il est impossible d'affirmer qu'un programme composé de plusieurs milliers de lignes de code est exempt de bugs, j'ai néanmoins consacré un temps considérable à soumettre le mien à une quantité de tests qui m'ont permis de corriger un nombre d'erreurs que je préfère garder secret.

Chapitre 4

Retour sur les expériences de molécule unique

CET CHAPITRE est consacré à l'étude de molécules linéaires étirées et surenroulées. J'y compare les résultats de mes simulations à ceux des expériences de molécule unique décrites précédemment (Section 2.2). Ces simulations servent aussi bien à tester la validité du modèle que j'utilise et de son implémentation, qu'à apporter un éclairage nouveau sur les questions encore débattues à propos du couple exercé par la molécule ou de la nature des structures et du nombre de plectonèmes formés.

4.1 Étirement de l'ADN surenroulé

Le modèle Worm-Like Chain non surenroulé

Je commence par vérifier la validité de mes simulations en les comparant, à surenroulement nul, au modèle Worm-Like Chain (Fig. 4.1), et en particulier à des relations prédites pour ce modèle entre la force f et l'extension relative χ . Pour la première, Bouchiat *et. al.* [81] ont obtenu calculé numériquement une valeur de l'extension pour certaines forces, puis utilisé une formule d'interpolation établie par Bustamante *et. al.* [101], afin d'obtenir l'approximation suivante :

$$f = \frac{k_B T}{l_p} \left(\frac{1}{4(1-\chi)^2} - \frac{1}{4} + \chi + \sum_{i=2}^7 \alpha_i \chi^i \right) \quad (4.1)$$

avec $\alpha_2 = -0.5164228$, $\alpha_3 = -2.737418$, $\alpha_4 = 16.07497$, $\alpha_5 = -38.87607$, $\alpha_6 = 39.49944$ et $\alpha_7 = -14.17718$.

La seconde formule est un cas particulier, à surenroulement nul, d'une formule établie par Moroz et Nelson [83] pour l'extension d'une molécule faiblement surenroulée :

$$\chi = 1 - \frac{1}{2} \left(\frac{l_p f}{k_B T} - \frac{1}{32} \right)^{-1/2} + \frac{f}{K_0} + \frac{k_B T}{L f} \quad (4.2)$$

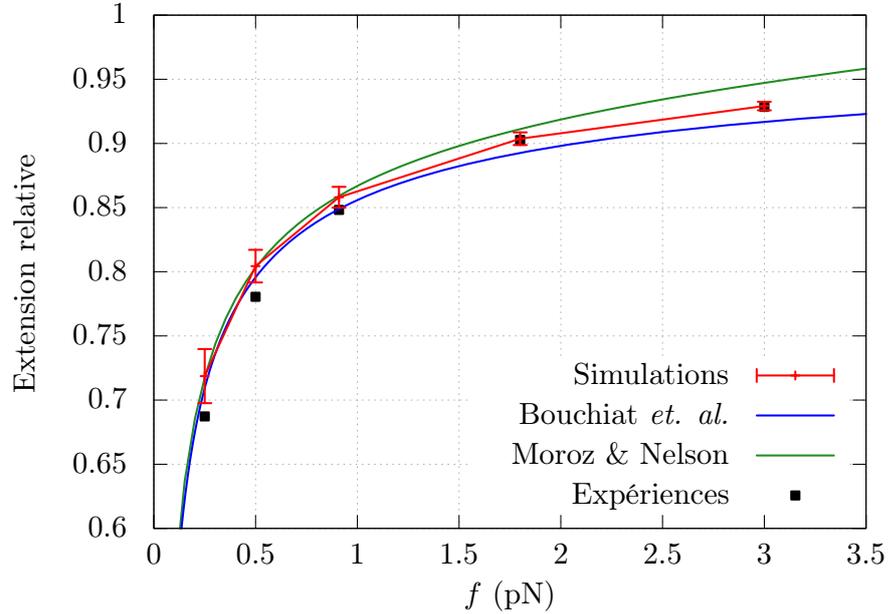


FIGURE 4.1: Courbe force-extension d'une molécule de 16 kb non surenroulée ($[\text{NaCl}] = 100 \text{ mM}$). En rouge : moyenne et écart-type (barres d'erreur) de l'extension obtenus par simulation pour différentes forces ($l_p = 50 \text{ nm}$). En noir : données expérimentales [74] aimablement fournies par F. Mosconi. En bleu : prédiction de Bouchiat et Mézard [81]. En vert : prédiction de Moroz et Nelson [83].

où K_0 est le module d'élasticité linéaire de l'ADN (de l'ordre de 1100 pN [80]) et L est la longueur de la molécule (5333 nm ici). Les deux derniers termes, qui tiennent compte respectivement de l'élasticité de la molécule et des effets de taille finie, n'ont que très peu d'influence dans ce régime de forces (leur contribution à l'extension χ est de l'ordre d'un millièème).

La figure 4.1 montre que les simulations sont tout-à-fait en accord avec le modèle Worm-Like Chain, ce qui est particulièrement visible à faible force ($f < 1 \text{ pN}$), lorsque les deux approximations sont elles-mêmes parfaitement superposées. L'accord avec les données expérimentales est lui aussi très bon, en particulier à force élevée.

Le modèle Worm-Like Chain surenroulé

En faisant varier le surenroulement $\sigma = \frac{Lk - Lk_0}{Lk_0}$, il est possible de discuter les expériences de surenroulement qui ont été effectuées sur des molécules dont la longueur varie entre 2,2 et 21 kb. Tout d'abord, les simulations permettent de reproduire les courbes d'extension mesurées par Mosconi *et al.* [74] (Fig. 4.2), et en particulier la transition de flambage. Cette transition, particulièrement prononcée à force importante ($f > 1 \text{ pN}$), correspond au niveau de surenroulement à partir duquel la contrainte de torsion est tellement grande qu'il devient plus favorable thermodynamiquement pour la molécule de se courber

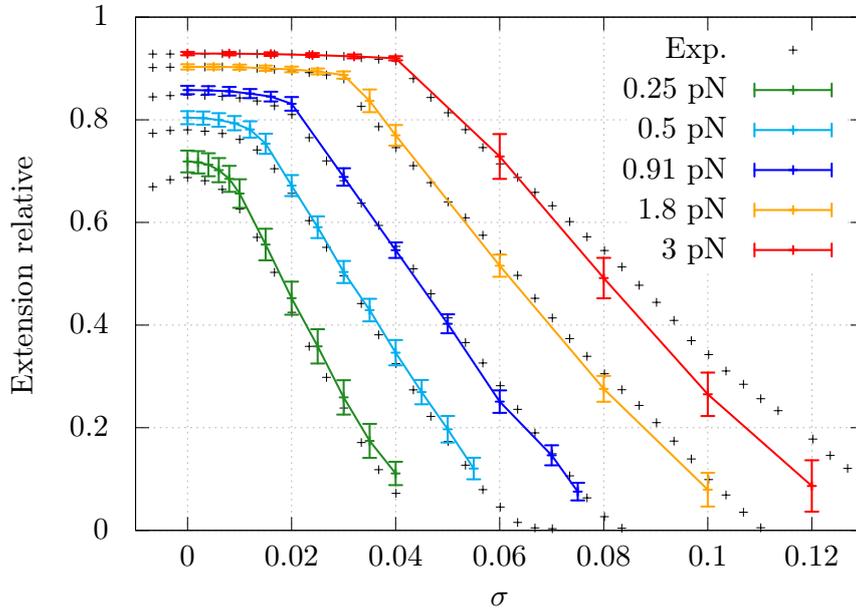


FIGURE 4.2: Extension d'une molécule de 16 kb en fonction du surenroulement σ ($[\text{NaCl}] = 100 \text{ mM}$), pour différentes forces. En couleur : moyenne et écart-type de l'extension obtenus par simulation. En noir : données expérimentales [74].

pour former des boucles et des plectonèmes, ce qui a pour effet de diminuer son extension.

L'accord entre les simulations et les mesures expérimentales est très bon, mis à part une extension légèrement sous-estimée à large force (3 pN) et large surenroulement ($\sigma > 0,1$). Cette différence peut s'expliquer par l'approximation de cœur dur du potentiel électrostatique de la double-hélice, qui est susceptible d'avoir un effet notable lorsque le surenroulement est très important et que de nombreuses parties de la molécule se trouvent alors en contact les unes avec les autres au sein des plectonèmes. Il est également possible que des termes non harmoniques de l'énergie de courbure ou de torsion deviennent non négligeables, ou qu'un couplage apparaisse entre ces termes.

En mesurant l'extension de molécules beaucoup plus petites (2200 paires de bases), Forth *et. al.* [71] ont observé une transition de flambage beaucoup plus abrupte : l'extension saute brutalement, suggérant que la formation du premier plectonème absorbe d'un seul coup une part non négligeable de la contrainte de torsion. La figure 4.3 montre que les simulations sont capables de reproduire cet effet, du moins à la concentration saline utilisée dans cette expérience ($[\text{NaCl}] = 150 \text{ mM}$). En revanche, à faible concentration en sel ($[\text{NaCl}] = 20 \text{ mM}$), cet effet disparaît et la transition est beaucoup plus douce. Ce résultat est en accord avec d'autres mesures effectuées par l'équipe de R. Seidel [102] [103] sur une molécule de 1900 paires de bases : le saut en extension, très prononcé à 320 mM, est plus difficilement visible à 60 mM et disparaît totalement à 20 mM. Ce saut abrupt est encore visible sur une molécule de 10,9 kb, quoique

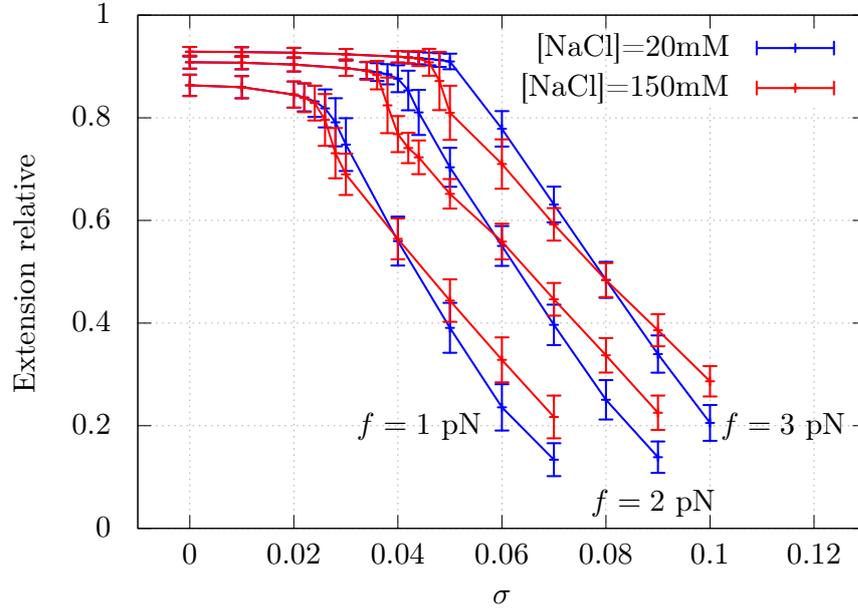


FIGURE 4.3: Effet de la concentration en sel : extension (moyenne et écart-type) d'une molécule de 2,2 kb en fonction du surenroulement, pour différentes forces et pour deux conditions salines différentes.

bien plus petit, ce qui est attendu si la taille du premier plectonème qui se forme à la transition est constante : plus la molécule est longue, moins un saut de taille fixe (typiquement quelques dizaines de nanomètres) sera visible sur la courbe de l'extension relative.

De plus, lorsque la concentration en sel diminue, la transition de flambage survient plus tard (c'est-à-dire avec un surenroulement plus élevé) et la pente de l'extension après la transition est plus forte ; deux effets que l'on observe également avec les simulations. De fait, une diminution de la concentration en sel se traduisant par une augmentation de la répulsion électrostatique de l'ADN (et une augmentation du rayon effectif dans mes simulations), il est intuitif de penser qu'une énergie plus importante est nécessaire pour former le premier plectonème.

Couple

Le calcul local du twist, associé aux cylindres virtuels jouant le rôle d'un réservoir d'hélices permet de déterminer le couple exercé par la molécule (Section 3.1), ce qui n'avait jamais pu être réalisé dans le cadre de simulations de Monte-Carlo.

Grandes molécules

La figure 4.4 présente ce couple pour une molécule de 16 kb dans les mêmes conditions que la figure 4.2. Là encore, la transition de flambage est clairement

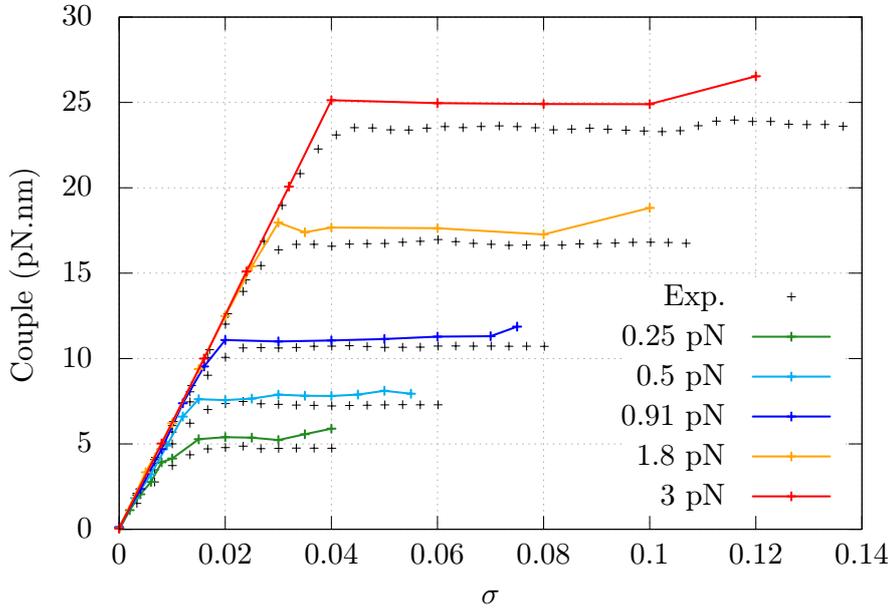


FIGURE 4.4: Couple exercé par une molécule de 16 kb en fonction du surenroulement ($[\text{NaCl}] = 100 \text{ mM}$), pour différentes forces. En noir : valeurs obtenues indirectement (Section 2.2) à partir des données expérimentales [74].

visible, et d'autant plus abrupte que la force est élevée. Avant la transition, le couple augmente linéairement avec le surenroulement, le surplus d'énergie étant intégralement stocké sous forme de torsion. Une fois les premières structures formées, en revanche, le couple atteint un plateau et reste constant par la suite, ce qui traduit le fait que tous les tours ajoutés au delà de la transition sont transformés en writhe.

Le couple semble augmenter à nouveau légèrement à très fort surenroulement, typiquement lorsque l'extension devient inférieure à 0,1. Cela peut être interprété comme l'effet de l'auto-évitement de la chaîne qui gêne le repliement de la molécule quand la densité d'ADN devient très élevée (encombrement stérique). Une partie du twist supplémentaire ne peut alors plus être transformée en writhe.

Les simulations sont, comme dans le cas de l'extension, en très bon accord avec l'expérience. En particulier, l'hypothèse du couple constant après la transition, utilisée par Mosconi *et al.* pour obtenir ces valeurs à partir des courbes d'extension, est vérifiée ici. Ce résultat va à l'encontre des conclusions d'Emanuel *et al.* [104] qui prédisent une augmentation continue du couple à faible force, due selon eux à la formation de plusieurs plectonèmes.

D'un point de vue quantitatif, la valeur du couple de flambage (hauteur du plateau) diffère légèrement de la valeur déduite des expériences. Cette différence est probablement due aux détails de la répulsion électrostatique de la molécule. Comme le montre la figure 4.5, le couple de flambage est très sensible au rayon effectif r_e des cylindres. Il est donc raisonnable de penser qu'il est également sensible à la forme exacte du potentiel de répulsion électrostatique entre les

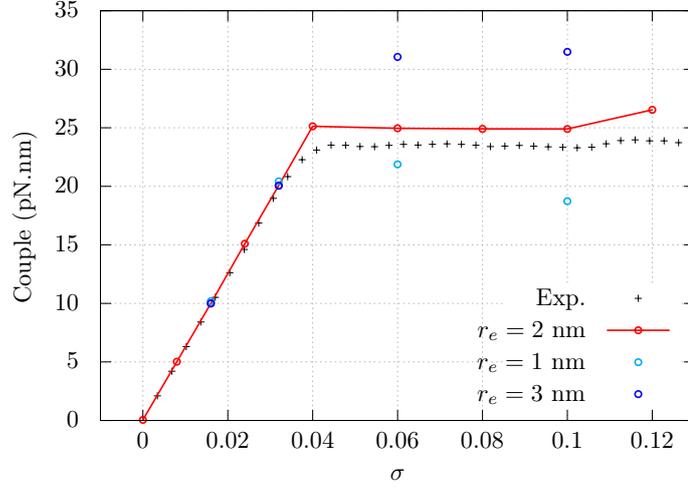


FIGURE 4.5: Couple exercé par une molécule de 16 kb (mêmes conditions que la figure 4.4, pour $f = 3$ pN et différents rayons effectifs r_e).

cylindres. L'approximation de cœur dur est donc peut-être trop grossière dans cette situation, en particulier lorsque les cylindres entrent fréquemment en contact au sein des plectonèmes.

Le couple exercé par la molécule en réponse au surenroulement n'est pas censé dépendre de la raideur du réservoir de twist ; ce que j'ai pu vérifier en faisant varier k_r de deux ordres de grandeurs (Fig. 4.6). Dans toutes les autres simulations, j'ai utilisé la même valeur $k_R = 100$ pN.nm, qui est du même ordre de grandeur que celle utilisée par Lipfert *et. al.* [75] pour mesurer directement le couple (Section 2.2), et qui permet de limiter les fluctuations du twist dans le réservoir sans pour autant figer le système à cause de trop de rejets.

La pente des courbes du couple en fonction du surenroulement avant la transition dépend de la force, ce qui est visible en zoomant sur la figure 4.4 (Fig. 4.7). Cette pente peut s'exprimer en termes d'un module de torsion effectif C_s :

$$\Gamma(\sigma) = 2\pi k_B T \frac{C_s}{h} \sigma \quad (4.3)$$

où h est la longueur d'une hélice de l'ADN (3,47 nm). C_s correspond intuitivement au module de torsion C que devrait posséder un tube parfaitement rigide pour exercer le même couple Γ . Du fait des fluctuations dues à la flexibilité de l'ADN, C_s est toujours inférieur à C (et tend vers C quand $f \rightarrow \infty$, c'est-à-dire quand la molécule se rapproche d'une droite).

La figure 4.8 présente les modules de torsion effectifs obtenus à partir des régressions linéaires de la figure 4.7. Ces résultats sont en bon accord avec ceux de Mosconi *et. al.* [74], ainsi qu'avec les prédictions théoriques de Moroz et Nelson [83], qui ont utilisé un modèle d'ADN faiblement surenroulé tenant compte de la courbure, de la torsion et du couple pour établir l'approximation

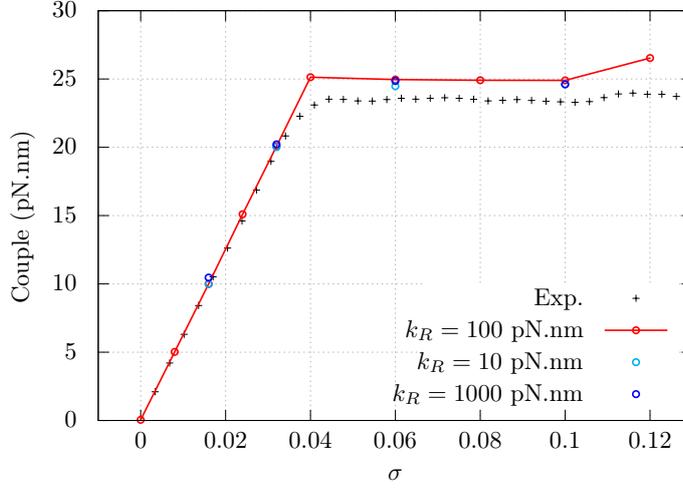


FIGURE 4.6: Couple exercé par une molécule de 16 kb (mêmes conditions que la figure 4.4, pour $f = 3$ pN et différentes raideurs k_r du réservoir de twist.

suivante :

$$C_s = C \frac{4l_p \sqrt{\frac{l_p f}{k_B T}}}{C + 4l_p \sqrt{\frac{l_p f}{k_B T}}} \quad (4.4)$$

La différence que l'on peut observer à faible force tient au moins en partie au fait que la transition de flambage est définie bien moins précisément : le couple évolue de manière beaucoup plus douce de la phase linéaire à la phase constante (Fig. 4.7). Il faudrait donc disposer de plus de points expérimentaux (et de plus de simulations) à surenroulement très faible pour augmenter la précision des régressions.

Quoi qu'il en soit, l'accord entre simulations et expériences confirme que $C = 86$ nm est une valeur raisonnable pour le module de torsion de l'ADN dans les conditions salines correspondant à $[\text{NaCl}] = 100$ mM ($l_p = 50$ nm et $r_e = 2$ nm).

Petites molécules

À l'instar des résultats concernant l'extension, le couple exercé par des molécules plus petites présente un comportement différent au niveau de la transition de flambage (Fig. 4.9) : une diminution rapide juste après la transition, avant de se stabiliser. Ici encore, c'est le relâchement soudain d'une part non négligeable de la contrainte de torsion par les premiers repliements en super-structures qui est visible sur ces courbes. De même que le saut de l'extension, la diminution du couple est très largement réduite à faible concentration en sel ($[\text{NaCl}] = 20$ mM).

Curieusement, cet effet n'a pas été observé par Forth *et. al.* [71], alors que leurs mesures montrent clairement un saut dans l'extension, et que les résultats d'Emanuel *et. al.* [104] (par ailleurs en très bon accord avec ces expériences) le

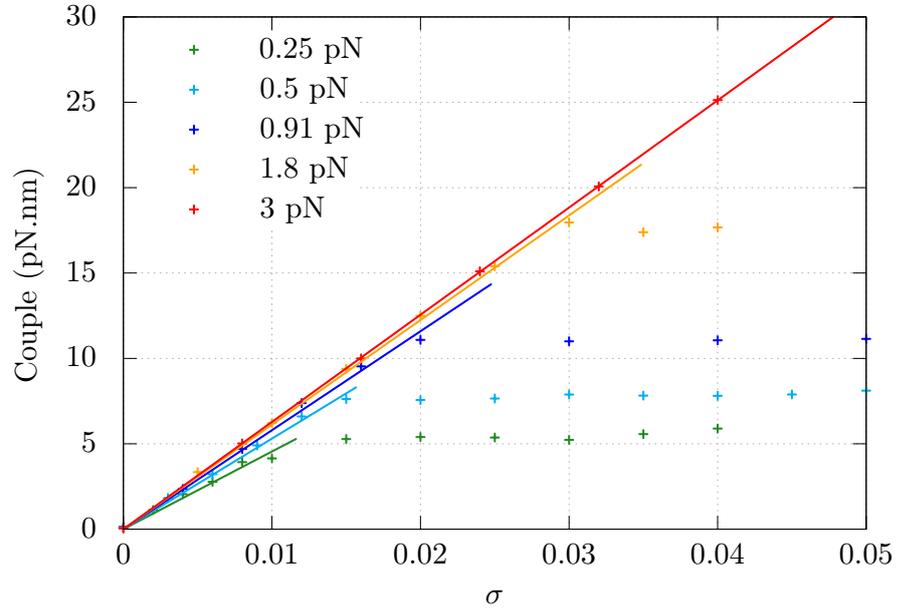


FIGURE 4.7: Zoom de la figure 4.4 à faible surenroulement. Les points représentent les valeurs obtenues par simulation et les segments correspondent aux régressions linéaires des points situés avant la transition.

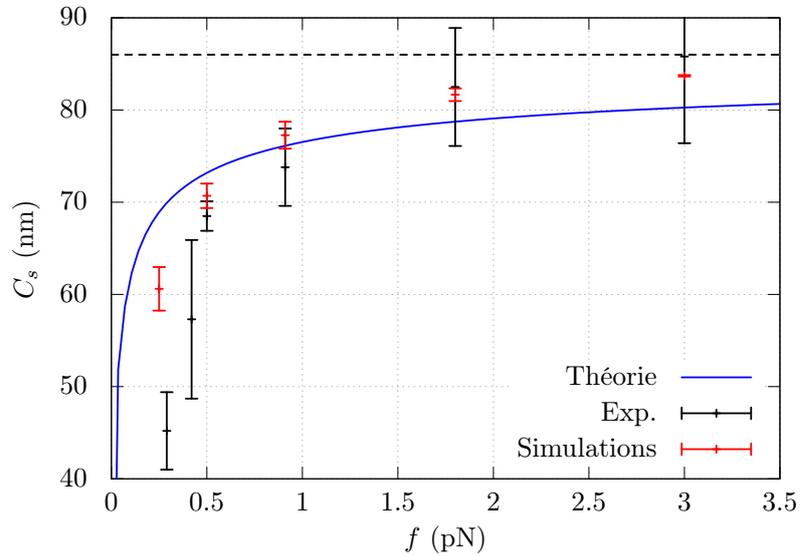


FIGURE 4.8: Module de torsion effectif en fonction de la force. En rouge : valeurs obtenues par les fits des couples de la figure 4.7. En noir : valeurs obtenues à partir des données expérimentales [74]. En bleu : prédiction de Moroz et Nelson [83] pour $l_p = 50$ nm et $C = 86$ nm (ligne en pointillés).

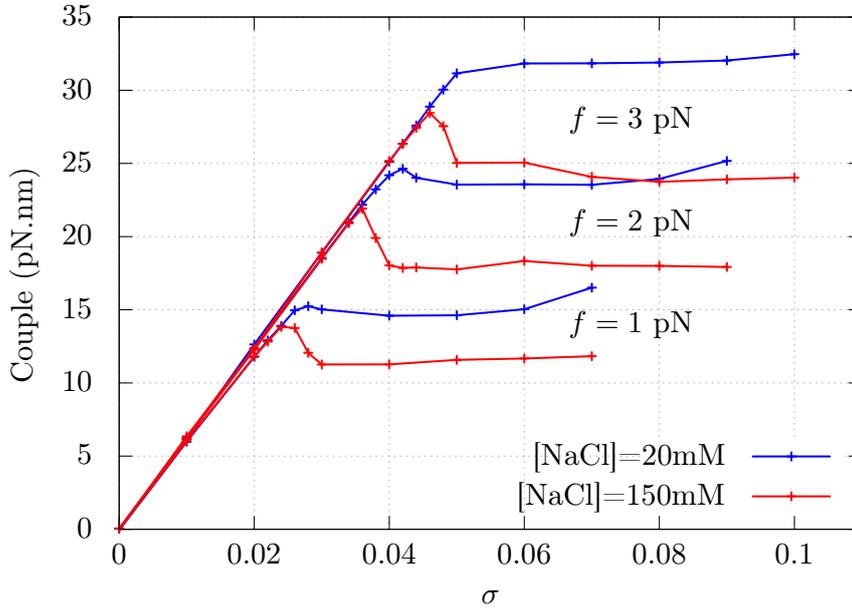


FIGURE 4.9: Effet de la concentration en sel : couple exercé par une molécule de 2,2 kb en fonction du surenroulement, pour différentes forces et pour deux conditions salines différentes.

prédisent également, du moins à force élevée. De plus, Maffeo *et. al.* [103] ont effectivement observé un telle « bosse » avec une molécule de 1,9 kb.

Effet de la discrétisation

Toutes les simulations de ce chapitre ont été réalisées avec un niveau de discrétisation de $n = 5$ cylindres par longueur de persistance, valeur également utilisée par Vologodskii *et. al.* [87]. Afin de s'assurer que ce niveau est suffisant pour obtenir des résultats pertinents et raisonnablement précis, j'ai étudié l'effet de la discrétisation sur les courbes d'extension et de couple.

La figure 4.10 montre, pour $f = 3$ pN, la convergence de l'extension quand n augmente, en particulier dans la phase plectonémique. Cette convergence est encore plus nette avec le couple (Fig. 4.11) et montre bien que les résultats deviennent indépendants de n à partir de $n = 5$. La figure 4.12 montre une convergence très rapide des résultats à plus faible force ($f = 0,25$ pN), puisque deux cylindres par l_p semblent suffisants dans ce cas.

En conclusion, il apparaît que le choix de cinq cylindres par l_p est effectivement un bon compromis entre rapidité de calcul et précision des résultats, même dans le cas extrême d'une force importante. Toutefois, je discuterai à nouveau de la discrétisation plus loin (Section 4.2), où cette question se posera à nouveau pour l'étude des structures créées, du fait du niveau de détail souhaité.

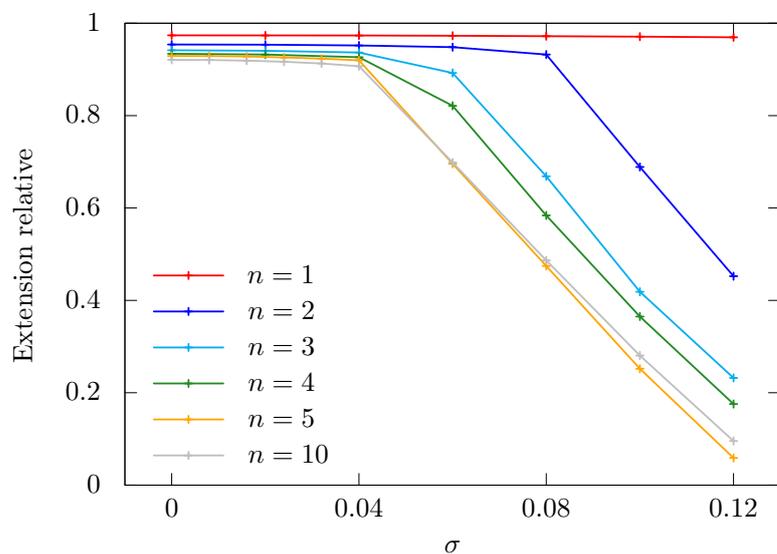


FIGURE 4.10: Extension relative d'une molécule de 16 kb en fonction du nombre n de cylindres par longueur de persistence l_p ($f = 3$ pN).

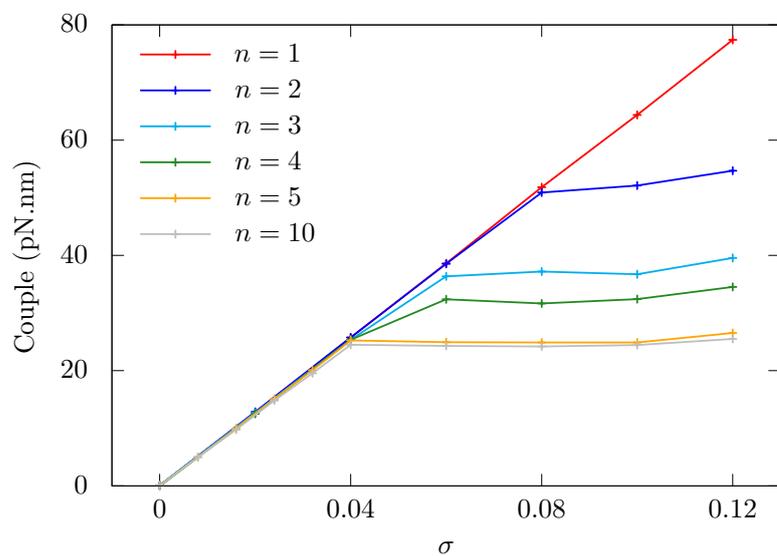


FIGURE 4.11: Couple d'une molécule de 16 kb en fonction du nombre n de cylindres par longueur de persistence l_p ($f = 3$ pN).

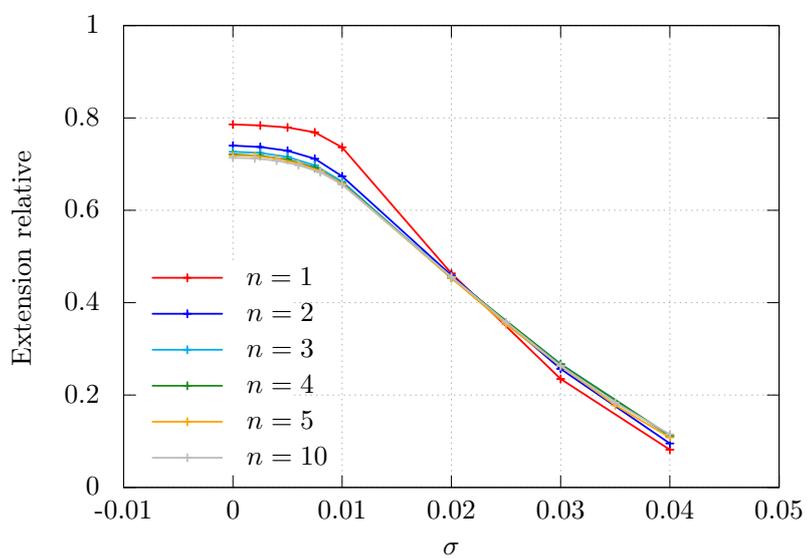


FIGURE 4.12: Extension relative d'une molécule de 16 kb en fonction du nombre n de cylindres par longueur de persistance l_p ($f = 0,25$ pN).

Énergie de torsion globale ou locale ?

Comme évoqué précédemment (Section 2.3), l'énergie de torsion peut être calculée de manière locale (en sommant les énergies de torsion de chaque cylindre) ou globale (en utilisant le twist total), ces deux méthodes étant censées donner les mêmes résultats [41]. J'ai pu vérifier ce fait en comparant, à paramètres identiques, des simulations utilisant l'une ou l'autre des formules pour calculer E_t (Fig. 4.13 pour l'extension et Fig. 4.14 pour le couple).

Cependant, l'évolution du système est beaucoup plus lente avec une énergie locale car le taux de rejet des mouvements est plus important. Ainsi, l'énergie locale impose une répartition bien plus homogène du twist le long de la molécule et en limite grandement les fluctuations (Fig. 4.15). De plus, lorsque le surenroulement est très important, l'énergie de torsion locale de la condition initiale est tellement grande (le surenroulement étant toujours nécessairement introduit sous forme de twist) que la molécule commence par se replier brutalement en une pelote très dense, avant de se réorganiser extrêmement lentement.

Afin de minimiser le temps nécessaire pour atteindre l'équilibre thermodynamique, il est donc préférable d'utiliser l'énergie globale de torsion. Cependant, une approche locale s'avérera nécessaire pour étudier des domaines indépendants (Chapitre 5.2), les propriétés des interactions avec des protéines ou les effets de séquence, pas exemple.

Convergence des simulations

Dans la plupart des simulations, la condition initiale est assez éloignée d'une configuration typique de l'état d'équilibre : la molécule est totalement étirée,

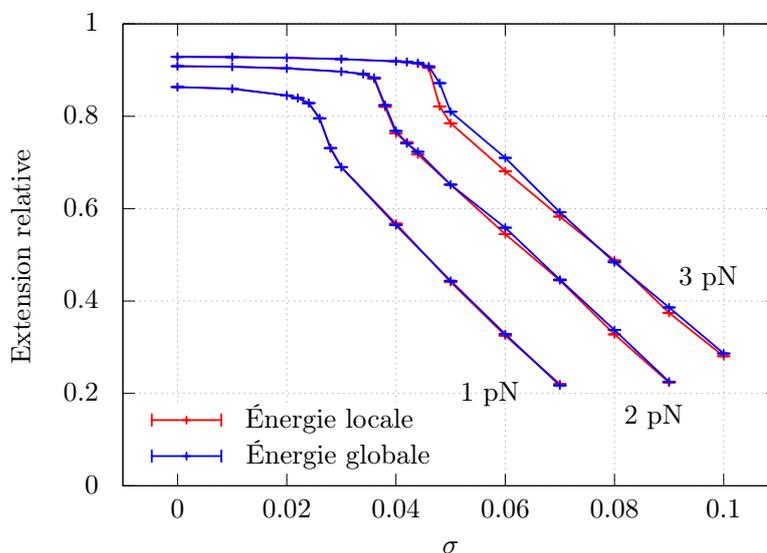


FIGURE 4.13: Extension relative d'une molécule de 2,2 kb en fonction du surenroulement, pour différentes forces et pour un calcul local ou global de l'énergie de torsion.

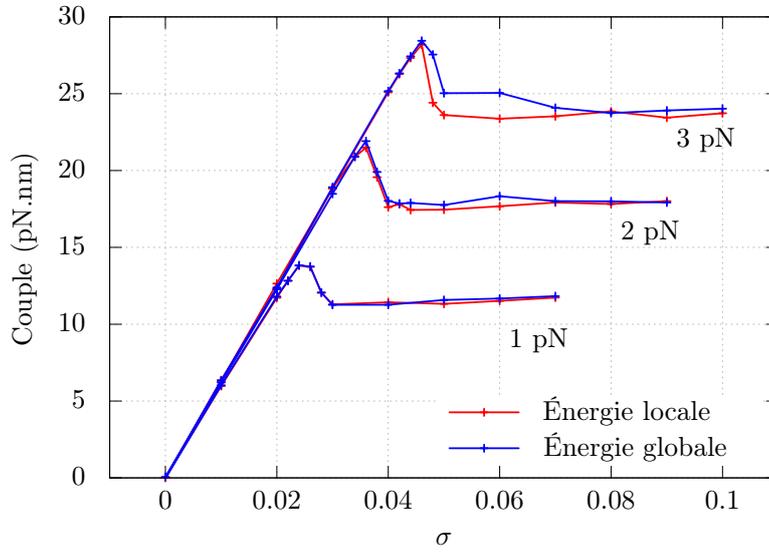


FIGURE 4.14: Couple d'une molécule de 2,2 kb en fonction du surenroulement, pour différentes forces et pour un calcul local ou global de l'énergie de torsion.

avec une extension relative $\chi = 1$ et un writhe $Wr = 0$. Il est difficile de prouver rigoureusement que le système atteint l'équilibre thermodynamique au cours d'une simulation (à moins de disposer d'un temps infini, ce qui n'est pas tout-à-fait mon cas). En revanche, il est raisonnable de penser que l'équilibre est atteint lorsque les diverses propriétés observables du système (énergie, extension, couple, writhe, etc.) n'évoluent plus en dehors de fluctuations autour d'une valeur moyenne, et que plusieurs simulations indépendantes (utilisant différentes séquences de nombres aléatoires) produisent les mêmes moyennes. J'ai pu vérifier cette condition à de nombreuses reprises en suivant l'évolution de ces observables au cours de différentes simulations (la figure 4.16 montre l'exemple du writhe et de l'extension). En pratique, toutes les moyennes présentées ici ont été effectuées sur au moins quatre simulations indépendantes de plusieurs millions à plusieurs dizaines de millions de pas Monte-Carlo (suivant la taille de la molécule), correspondant à des milliards de déplacements possibles de cylindres individuels.

Le nombre de pas nécessaires pour que le système atteigne l'équilibre thermodynamique dépend de nombreux facteurs ; le principal étant le nombre de cylindres dans la chaîne (lui-même dépendant de la taille de la molécule et du niveau de discrétisation). Si une molécule de 16 kb (avec $n = 5$) nécessite quelques millions de pas pour que l'extension et le writhe se stabilisent, comme le montre la figure 4.16, la convergence est beaucoup plus rapide pour les molécules de 2,2 kb également étudiées. L'amplitude des rotations, la température, les paramètres énergétiques du polymère et, plus généralement, tout ce qui influe sur le taux de rejet des mouvements, peuvent affecter la durée de la thermalisation.

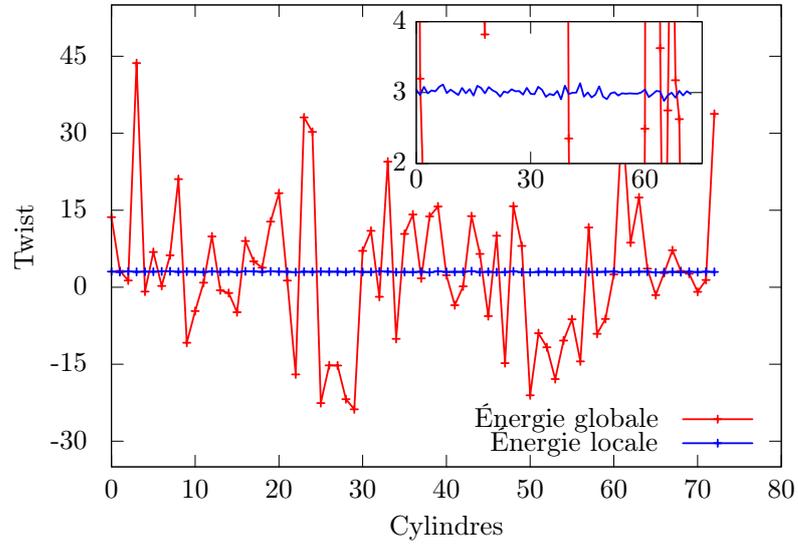


FIGURE 4.15: Répartition du twist d'une molécule de 2,2 kb ($f = 3 pN$, $\sigma = 0,06$) entre les cylindres, à un instant donné d'une simulation. Miniature : zoom permettant de distinguer les fluctuations du twist pour l'énergie locale. L'inhomogénéité de la répartition du twist dans le cas de l'énergie globale augmente avec le nombre de pas, seul le twist total étant contraint.

Comparaison avec les simulations existantes

Un dernier test consiste à comparer mes simulations à celles d'A. Vologodskii. La figure 4.17 montre que les résultats des deux programmes sont en bon accord, en particulier à faible surenroulement.

La source de la différence visible à surenroulement élevé est délicate à déterminer car je n'ai pas pu avoir le contrôle sur tous les paramètres du programme d'A. Vologodskii ; je ne peux donc pas assurer que les deux simulations ont été réalisées exactement dans les mêmes conditions.

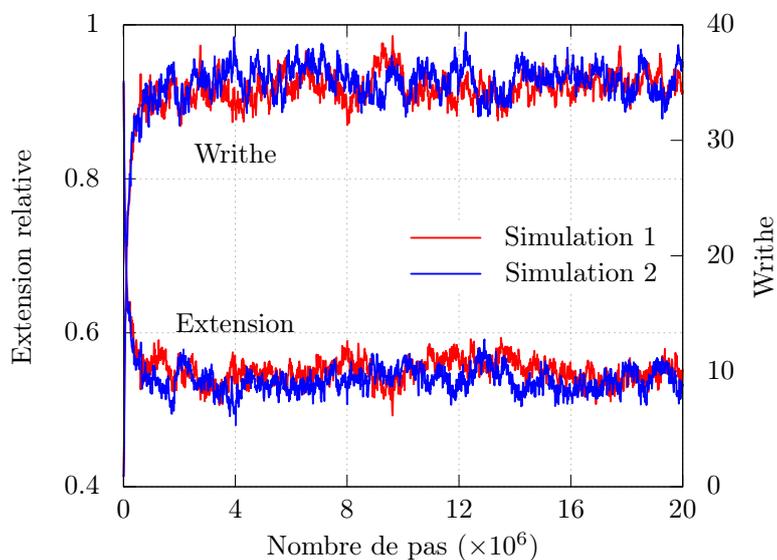


FIGURE 4.16: Extension et writhe d'une molécule de 16 kb ($f = 0.91$ pN, $\sigma = 0.04$) au cours de deux simulations différentes.

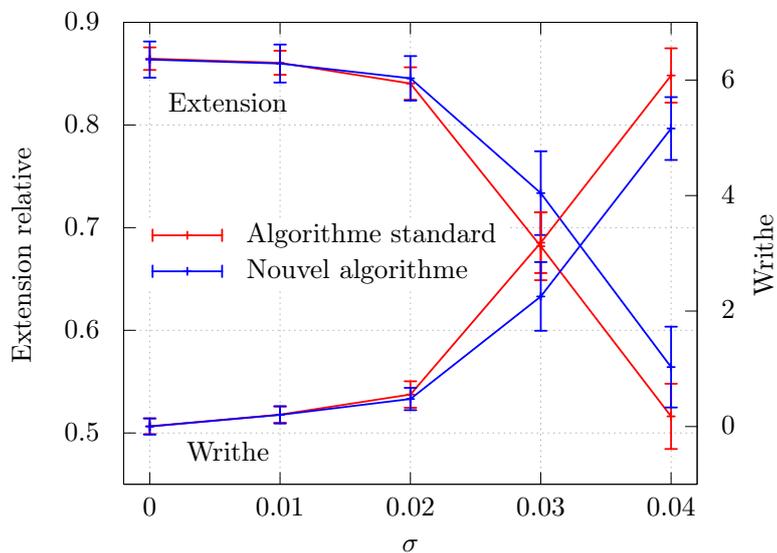


FIGURE 4.17: Comparaison de l'extension et du writhe d'une molécule de 3 kb ($f = 1$ pN) obtenus par mon programme (bleu) et celui d'A. Vologodskii (rouge).

4.2 Structures créées

Visualisation des structures

Mes simulations permettent également d'étudier les plectonèmes et les autres structures créés par le surenroulement de grandes molécules étirées, pour lesquelles aucune information directe n'a encore été obtenue. Quelques exemples de configurations sont montrés sur la figure 4.18 et permettent déjà de constater que même à force élevée (3 pN), c'est-à-dire lorsque les fluctuations sont réduites, le repliement d'une molécule de 16 kb est relativement complexe, avec plusieurs plectonèmes de tailles différentes distribués le long de l'axe z de la force f , certains étant branchés.

Afin d'étudier le comportement de ces structures au cours des simulations, j'ai développé un outil permettant de visualiser la densité d'ADN le long de l'axe z et son évolution. Cet outil s'inspire de la méthode expérimentale utilisée par Van Loenhout *et. al.* [69] pour visualiser les plectonèmes (voir Section 2.2), et me permettra donc également de comparer mes résultats à ces observations. La figure 4.19 illustre cette méthode : il suffit de réaliser, pour une configuration donnée, un histogramme du nombre de cylindres le long de z . Les plectonèmes apparaissent alors comme des zones localisées de haute densité. En construisant cet histogramme régulièrement au cours des simulations, on peut suivre l'évolution des plectonèmes sur plusieurs millions de pas Monte-Carlo.

Plusieurs remarques peuvent d'ores et déjà être faites sur cette figure. Premièrement, les plectonèmes semblent très stables : ils sont présents tout au long des 10^8 pas de cette simulation. Ils sont également très peu mobiles et ne diffusent pas le long de l'axe z , au contraire des observations expérimentales (il est toutefois possible d'augmenter leur mobilité, voir plus loin la Section 4.2). Enfin, la densité fluctue autour de cette position fixe des plectonèmes.

En revanche, une autre simulation utilisant exactement les mêmes paramètres que la première, à l'exception de la séquence de nombres pseudo-aléatoires, peut conduire à des plectonèmes à d'autres positions, et même en nombre différent. La figure 4.20 permet de constater cette variabilité des traces de densité : bien que l'extension soit identique pour toutes les simulations, le nombre et la position des plectonèmes varie grandement. De plus, étant réalisées à force plus faible que pour la figure 4.19, ces simulations montrent que certains plectonèmes peuvent effectivement apparaître ou disparaître au cours d'une seule simulation, quoique les plus grosses structures restent très stables.

La stabilité du nombre et de la position des plectonèmes, observée dans les simulations comme dans les expériences (en particulier à force importante), pose alors la question des transitions entre ces différentes configurations et de la brisure de l'ergodicité du système (qui n'a toutefois de véritable sens que dans la limite de taille infinie). Cette question ne peut être traitée qu'à l'aide d'une visualisation directe des plectonèmes, puisque la valeur moyenne de l'extension est la même pour toutes les simulations.

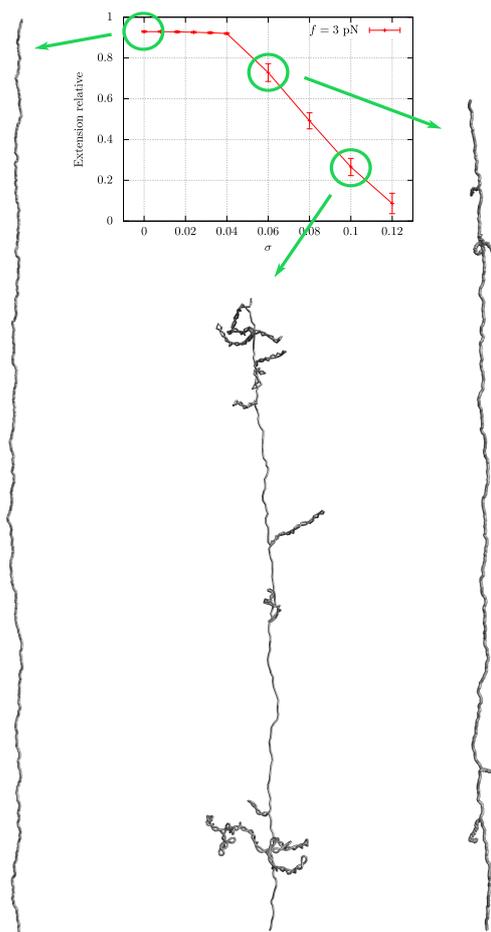


FIGURE 4.18: Exemples de configurations d'une molécule de 16 kb ($f = 3$ pN, $[\text{NaCl}] = 100$ mM) pour différentes valeurs du surenroulement. Les trois configurations ne sont pas à la même échelle (celle correspondant à $\sigma = 0.1$ est en réalité trois fois moins étendue que celle correspondant à $\sigma = 0$).

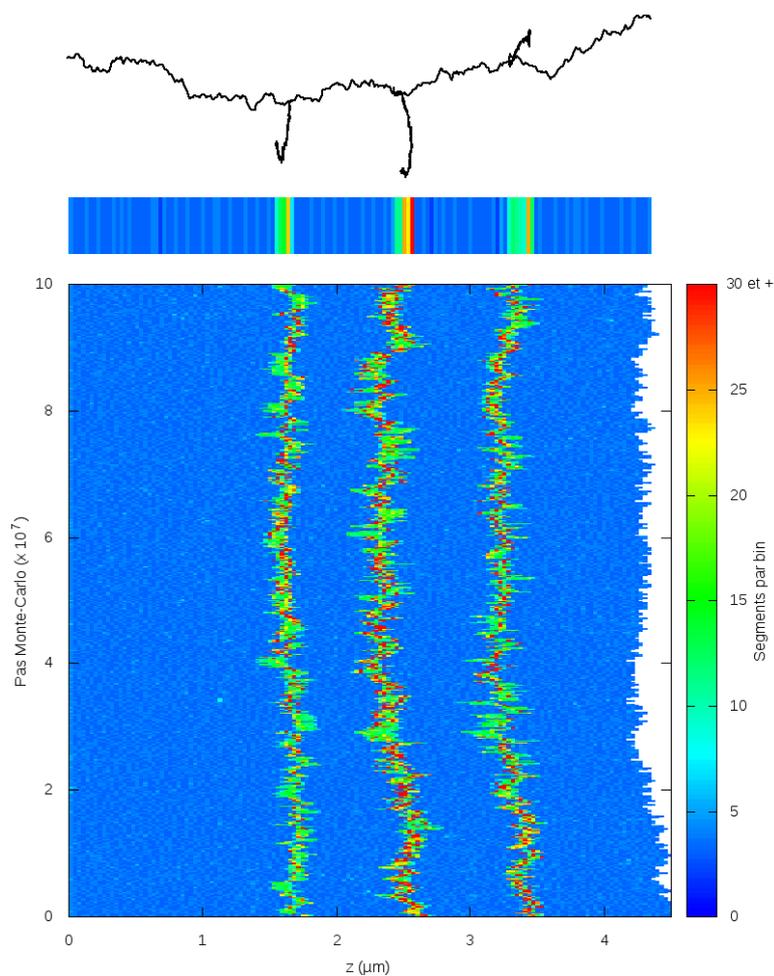


FIGURE 4.19: Construction d'une matrice de densité. En haut : exemple d'une molécule de 21 kb surenroulée ($\sigma = 0.045$) et étirée suivant z (axe horizontal, $f = 1,6$ pN), présentant trois plectonèmes. Au dessous : densité d'ADN le long de z (en nombre de cylindres par classe ; une classe mesurant environ 31 nm). Les plectonèmes se détachent nettement des zones de faible densité (en bleu). En bas : évolution de cette densité au cours des simulations.

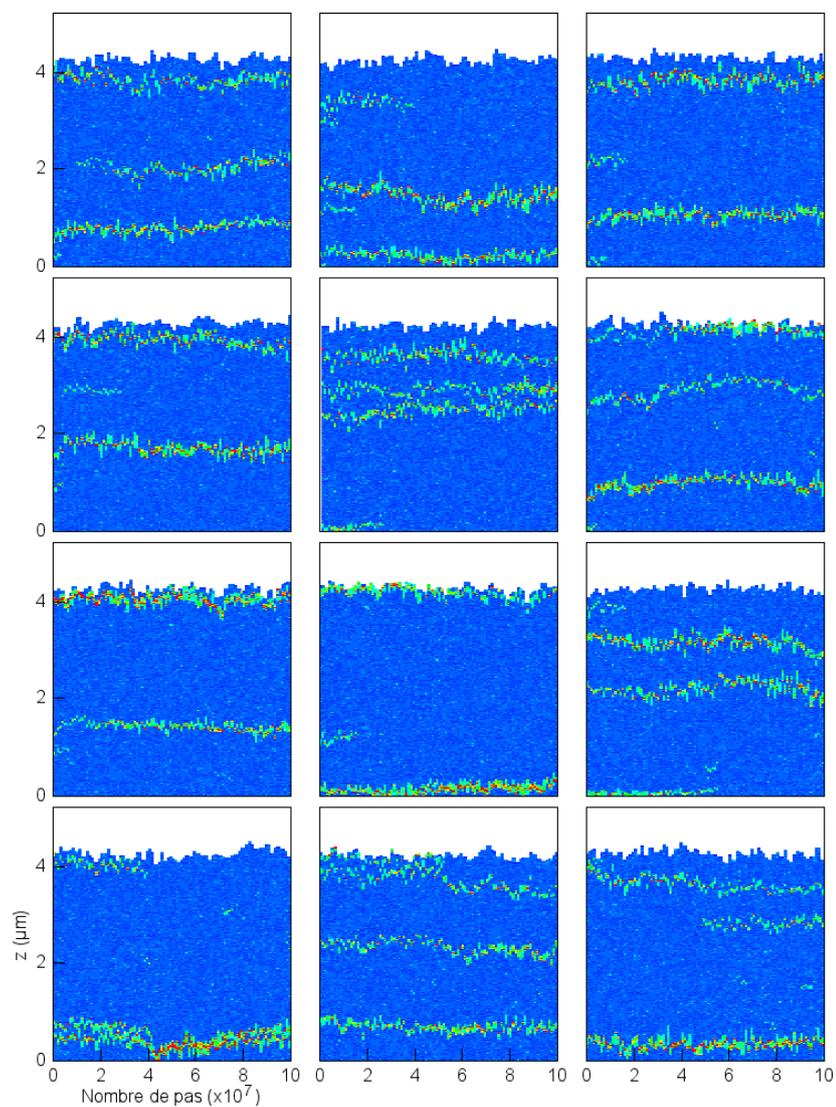


FIGURE 4.20: Traces obtenues par douze simulations indépendantes d'une molécule de 21 kb ($f = 0,8$ pN, $\sigma = 0,033$).

Problème de la stabilité des plectonèmes

Numériquement, l'absence de diffusion peut donc être un problème important pour échantillonner correctement l'espace des configurations. En effet, la position de la base du plectonème n'ayant aucun rôle dans l'énergie totale (la séquence n'étant pas prise en compte ici), on s'attend à observer une diffusion des plectonèmes, voire des sauts à longue distance tels qu'observés par Van Loenhout *et. al.* [69], ainsi que dans les simulations de petites molécules (Fig. 4.21).

De plus, cette fixité des plectonèmes a une autre conséquence : ceux-ci ont tendance à se trouver aux extrémités du polymère. Cet effet est visible sur plusieurs traces de la figure 4.20, en particulier celle située ligne 3, colonne 2, qui est assez caricaturale pour cet aspect.

Deux explications sont possibles pour expliquer ce phénomène. Premièrement, il est possible que des effets de bords favorisent la formation des premiers plectonèmes aux extrémités (par exemple, du fait de la plus grande liberté de mouvement aux bords et parce que les axes des rotations y ont des directions aléatoires, contrairement au cœur du polymère où l'axe de rotation, qui joint deux cylindres, aura tendance à être aligné selon z). Deuxièmement, même en supposant que la nucléation des plectonèmes peut se produire n'importe où, uniformément suivant z , l'accroissement d'un plectonème va englober petit-à-petit les plus proches cylindres de chaque côté du bourgeon, jusqu'à atteindre sa taille d'équilibre λ . Ainsi, tous les plectonèmes qui commencent à se former à une distance inférieure à $\sim \lambda/2$ d'une extrémité vont, au cours de leur croissance, englober les cylindres jusqu'à cette extrémité. Dans les deux cas,

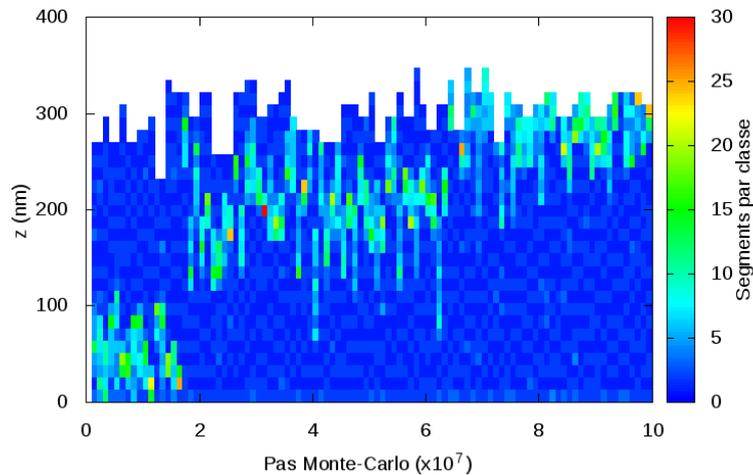


FIGURE 4.21: Trace de densité d'une molécule de 2,2 kb ($f = 1$ pN, $\sigma = 0,05$). Le plectonème saute régulièrement d'une position à une autre.

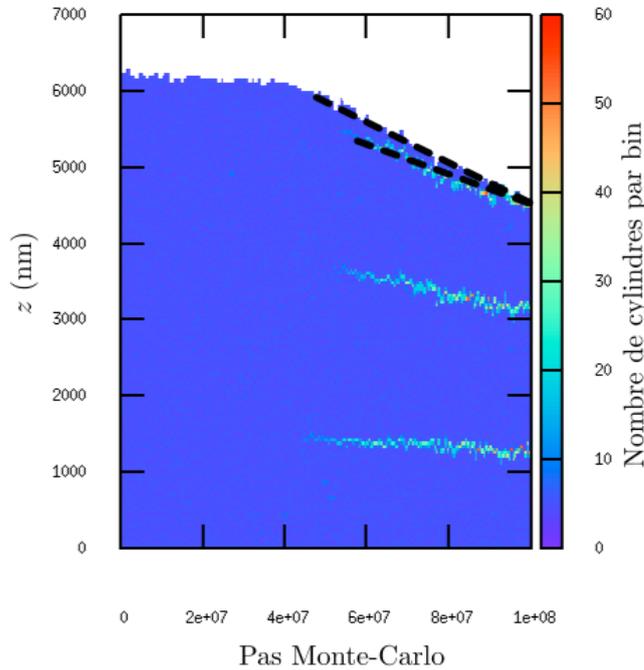


FIGURE 4.22: Trace de densité d’une molécule de 21 kb au cours d’une simulation où le surenroulement est ajouté progressivement. Le système n’est ici jamais à l’équilibre thermodynamique, mais la croissance progressive des plectonèmes est suffisamment lente pour être suivie à l’oeil nu, et permet de constater l’attraction effective des bords (par le rapprochement des deux lignes en pointillés correspondant à l’extension et à la trajectoire d’un plectonème).

l’immobilité des plectonèmes aura pour conséquence un biais de leur position vers les extrémités de la molécule.

La seconde explication est d’ailleurs illustrée par la figure 4.22, qui montre l’évolution de la densité au cours d’une simulation où le surenroulement augmente petit-à-petit. Au fur et à mesure de sa croissance, le plectonème qui s’est formé non loin d’un bord s’en rapproche jusqu’à y adhérer tout-à-fait.

Afin de savoir si la stabilité des plectonèmes bloque le système dans des états métastables et l’empêche d’atteindre l’équilibre thermodynamique, j’ai réalisé des simulations de recuit simulé pour étudier l’effet d’un refroidissement lent sur les résultats statistiques. Partant de $\beta = 1/k_B T = 0$ au début des simulations, β est augmenté progressivement, à chaque pas, jusqu’à atteindre sa valeur finale β_f (correspondant à $T_f = 300$ K) après un temps t_f de l’ordre du temps nécessaire pour stabiliser l’énergie, l’extension, etc. dans les simulations précédentes, soit 10^7 pas. La figure 4.23 permet de constater, pour différents niveaux de surenroulement, que le refroidissement est suffisamment lent : l’évolution de l’extension au cours du refroidissement n’est pas modifiée si la vitesse de refroidissement est divisée par deux ($t_f = 2 \cdot 10^7$).

La figure 4.24 compare les résultats obtenus de cette manière à ceux obtenus

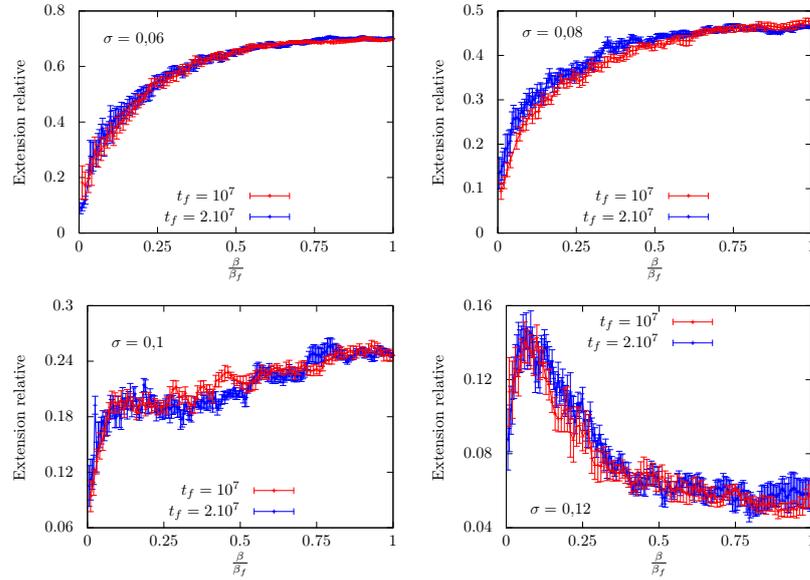


FIGURE 4.23: Extension (moyenne et écart-type) d’une molécule de 16 kb ($f = 3$ pN) en fonction de β au cours de simulations de recuit simulé, pour différentes valeurs de surenroulement. L’évolution de l’extension est identique pour les deux vitesses de refroidissement (t_f est le nombre de pas après lequel $\beta = \beta_f$).

lors des simulations précédentes ; l’extension et le couple sont identiques. Si le nombre moyen de plectonèmes d’une molécule de 21 kb est lui aussi inchangé, l’écart-type est en revanche légèrement plus grand pour le recuit simulé, en particulier pour $f = 3$ pN. La taille moyenne des plectonèmes est également légèrement plus faible à force importante. Les résultats sont cependant indépendants de la vitesse de refroidissement, ce qui conforte l’idée que ce dernier est suffisamment lent.

Possibles solutions pour la stabilité des plectonèmes

Il est assez intuitif de penser que les rotations de blocs utilisées dans les simulations, bien qu’efficaces pour former les plectonèmes (ou les défaire si la contrainte de surenroulement est relâchée), ne permettent pas vraiment de déplacer ces plectonèmes le long de l’axe : la plupart des tentatives de déplacement d’une partie du plectonème (et non pas de l’ensemble) se soldera par un croisement ou une intersection de cylindres.

La solution au problème de la stabilité des plectonèmes peut donc passer par l’utilisation d’autres types de mouvements élémentaires adaptés à l’évolution des structures formées.

En particulier, la diffusion des plectonèmes le long de l’axe de la force d’étirement nécessite plutôt une forme de translocation de l’ADN : les cylindres doivent « glisser » les uns sur les autres pour décaler l’ensemble de la structure par rapport à l’axe principal. Un tel mouvement élémentaire peut être défini comme illustré sur la figure 4.25. Un cylindre i pris au hasard est déplacé à une

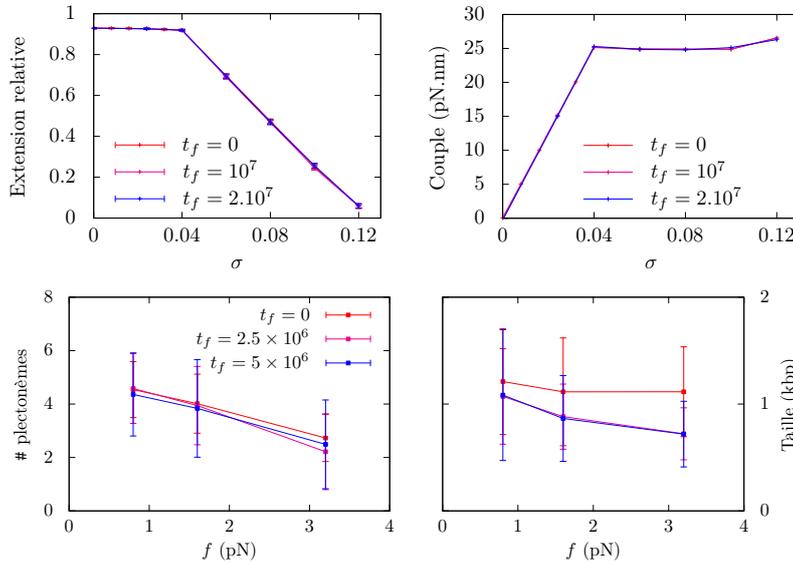


FIGURE 4.24: En haut : extension et couple d’une molécule de 16 kb ($f = 3$ pN) en fonction du surenroulement pour différentes vitesses de refroidissement. En bas : nombre et taille des plectonèmes d’une molécule de 21 kb en fonction de la force (σ correspondant aux expériences de Van Loenhout *et. al.*, voir Section 2.2).

autre position aléatoire dans la chaîne, et le bloc compris entre ces deux positions subit une translation de $\pm l \vec{t}_i$ (le signe dépendant de sa position relative à i avant le mouvement). Cette translocation est inspirée d’un mouvement similaire proposé par Vologodskii *et. al.* [87] impliquant le déplacement de plusieurs cylindres de part et d’autre du bloc, ce qui complique, dans mon algorithme, le traitement de la continuité du brin et le calcul de la variation de twist. Par conséquent, un seul cylindre se déplace ici. À chaque pas Monte-Carlo, le mouvement sera alors tiré au sort : rotation ou translocation.

L’effet de la translocation est visible sur la figure 4.26. Les plectonèmes diffusent maintenant rapidement et parcourent tout l’espace qui leur est disponible suivant z . Cependant, ce nouveau mouvement ne parvient toujours pas à véritablement changer la structure des plectonèmes (par exemple, un plectonème branché le restera) ni leur nombre, les seules translocations acceptées étant celles qui se contentent de déplacer des plectonèmes entiers. Il est donc toujours nécessaire de réaliser plusieurs simulations en parallèle pour obtenir un échantillon représentatif des structures qui peuvent se former. Sachant que les translocations ne sont pas très efficaces pour former les plectonèmes et que leur présence divise mécaniquement le nombre de rotations par deux, les simulations nécessitent deux fois plus de pas pour atteindre l’équilibre (bien que les translocations elles-mêmes soient assez rapides car, pour peu que les cylindres soient suffisamment petits, elles ne nécessitent pas de tester si un croisement a eu lieu). Enfin, le taux de rejet des translocations étant particulièrement élevé (généralement autour de 90%), j’ai choisi de ne pas utiliser ce mouvement pour

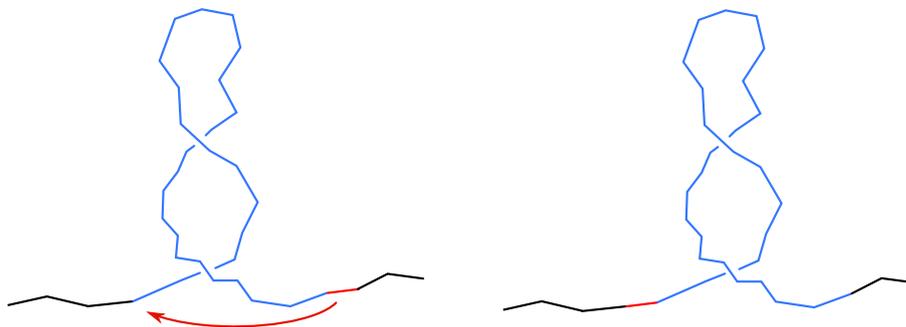


FIGURE 4.25: Schéma illustrant un mouvement élémentaire de translocation. Un segment (en rouge) est choisi aléatoirement et déplacé à une nouvelle position le long de la chaîne. Le bloc (en bleu) situé entre les positions initiale et finale de ce cylindre est décalé en conséquence.

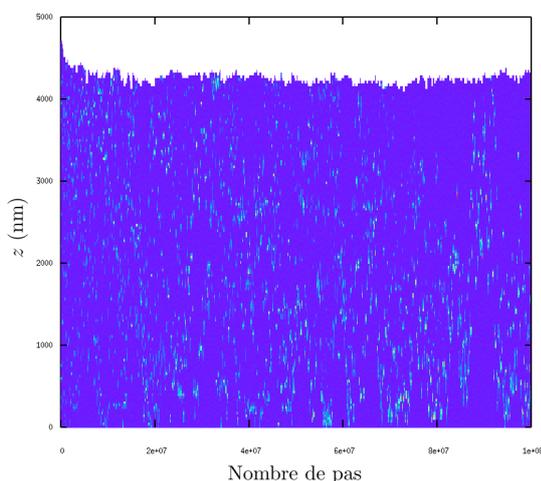


FIGURE 4.26: Densité d'une molécule de 21 kb subissant des mouvements de rotation et de translocation.

obtenir les résultats que je présente ici.

Il est possible d'imaginer d'autres types de mouvements conçus spécifiquement pour déformer les plectonèmes afin d'en créer ailleurs, de les faire échanger des supertours entre eux ou de les fusionner, mais la multiplication des mouvements élémentaires n'est peut-être pas la manière la plus efficace de faire évoluer le système.

En revanche, cette évolution pourrait éventuellement bénéficier d'une modélisation plus douce de la répulsion électrostatique. En effet, le potentiel de cœur dur est responsable de nombreux rejets dus aux collisions entre cylindres, qui surviennent principalement au sein des plectonèmes. Un potentiel plus doux serait donc susceptible d'autoriser quelques mouvements à l'intérieur des plec-

tonèmes, et donc de favoriser la reptation des cylindres par un effet de lubrification, diminuant ainsi la stabilité des plectonèmes. Le risque est cependant de ralentir les simulations car un grand nombre de paires de cylindres seraient susceptibles de contribuer à l'énergie de répulsion électrostatique de la molécule, ce qui pourrait réduire voire annuler l'avantage d'un taux de rejet plus faible.

Enfin, il est possible d'avoir recours à des techniques plus générales (non spécifiques au problème du repliement de l'ADN), utilisées dans des simulations de Monte-Carlo ou de dynamique moléculaire, telles que le « *parallel tempering* » [105], où le système est plongé, au cours d'une simulation, dans des bains thermiques à différentes températures, ce qui facilite la traversée des barrières énergétiques. La métadynamique [106], en pénalisant les états déjà visités par le système au fur et à mesure de son évolution, pourrait également permettre de réduire la stabilité des plectonèmes, à condition d'utiliser un jeu de variables collectives adapté. Par exemple, une quantité inspirée du nombre de coordination, en rendant compte des contacts entre cylindres au sein d'un plectonème, pourrait forcer ce dernier à se dérouler, provoquant la formation ou l'élongation d'autres plectonèmes.

Nature des structures

Lorsque la force et le surenroulement ne sont pas trop élevés, les plectonèmes ne sont pas les seules structures à apparaître. De nombreuses boucles isolées peuvent être observées, ainsi que des régions hélicoïdales (Fig. 4.27). Contrairement aux plectonèmes, ces structures sont transitoires, et même généralement très éphémères.

Comme le montre la figure 4.28, ces structures apparaissent rarement lorsque la concentration en sel est trop élevée (de l'ordre de 100 mM), le faible rayon électrostatique favorisant l'existence de longs plectonèmes. Lorsque la salinité diminue, les plectonèmes diminuent en taille et en nombre, jusqu'au point où seules de simples boucles et des hélices sont présentes (en dessous de 10 mM). Le writhe n'est alors plus du tout localisé et la densité d'ADN le long de z devient homogène (Fig. 4.28C).

Les boucles sont également les premières structures détectables à l'aide de la densité à la transition de flambage (Fig. 4.29). Très vite, cependant, un plectonème stable apparaît, et le nombre de plectonèmes augmente rapidement, tout comme la taille moyenne de ceux-ci.

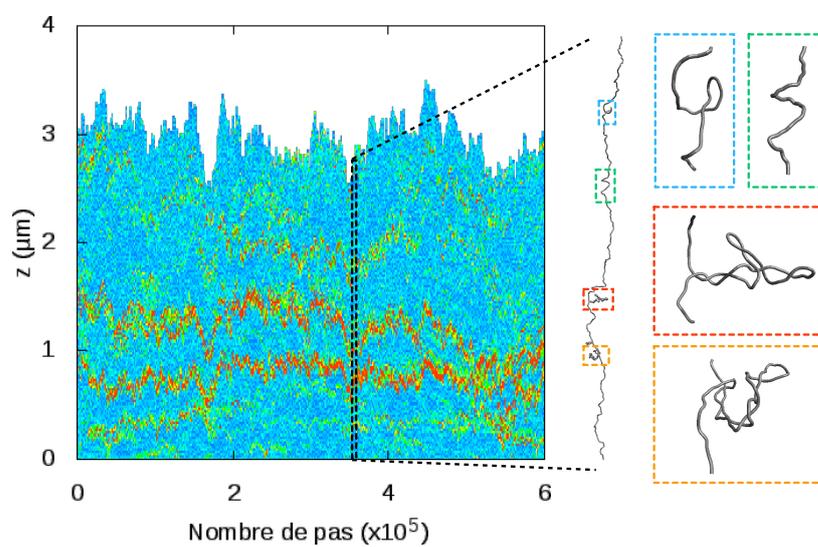


FIGURE 4.27: Exemple de trace de densité d'une molécule de 21 kb ($f = 0,25$ pN, $\sigma = 0,02$, $[\text{NaCl}] = 100$ mM). Une configuration particulière extraite de ces simulations, à droite, présente les différentes structures observées dans les simulations : boucles (en bleu), hélices (en vert) et plectonèmes branchés (en rouge) ou non (en jaune).

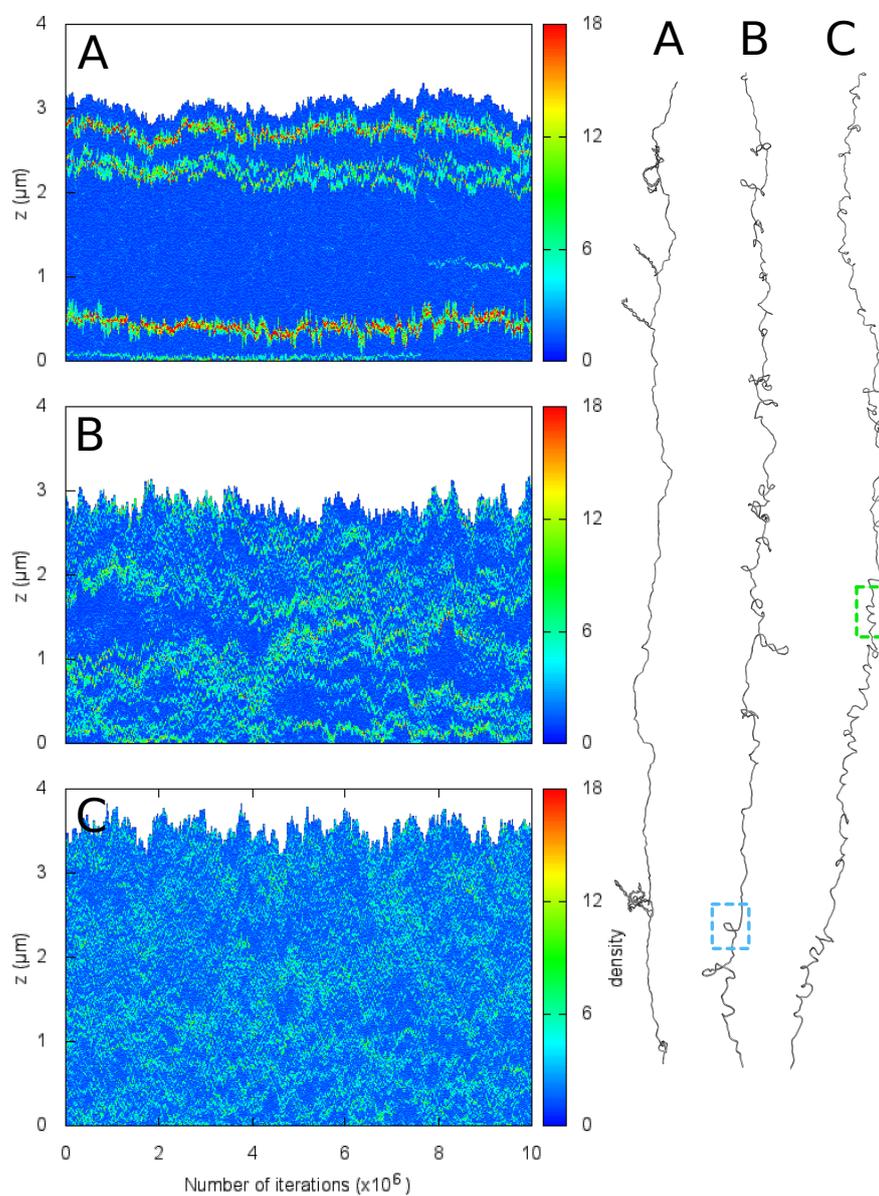


FIGURE 4.28: Effet du rayon effectif sur les traces (à gauche) et les configurations typiques (à droite) d'une molécule de 21 kb ($f = 1$ pN, $\sigma = 0,05$). A : $r_e = 2$ nm ($[\text{NaCl}] = 100$ mM). B : $r_e = 5$ nm ($[\text{NaCl}] = 10$ mM). C : $r_e = 8$ nm ($[\text{NaCl}] = 3$ mM). Plus le rayon augmente, moins on observe de plectonèmes et plus on observe de boucles (bleu), puis d'hélices (vert).

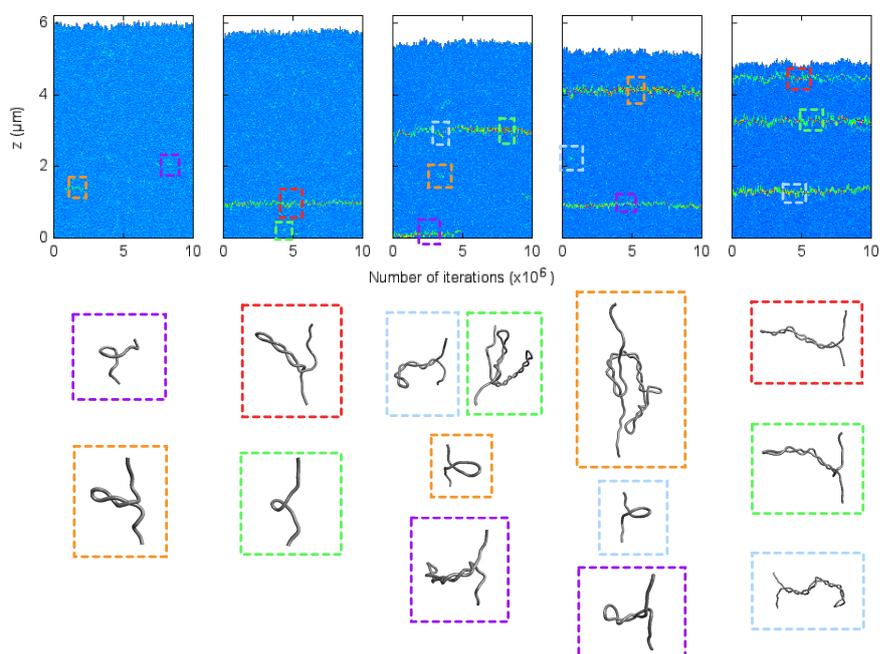


FIGURE 4.29: Apparition des structures au cours de la transition de flambage d'une molécule de 21 kb ($f = 1$ pN). De gauche à droite : $\sigma = 0, 0,02, 0,0225, 0,025, 0,0275, 0,03$. En bas : zooms sur les structures repérées dans les traces de densité.

Nombre de plectonèmes

Afin de réaliser des prédictions un peu plus quantitatives et de pouvoir notamment comparer celles-ci avec les résultats expérimentaux de Van Loenhout *et. al.* [69], j'ai implémenté un algorithme de détection automatique des plectonèmes. L'idée est de parcourir la chaîne en cherchant à savoir, pour chaque cylindre i , si ce cylindre correspond au début d'un plectonème. Comme illustré sur la figure 4.30, c'est le cas s'il existe un cylindre éloigné de i le long de la chaîne mais proche dans l'espace.

Plus précisément, en définissant λ la longueur minimale, le long de l'ADN, des plectonèmes qui seront détectés (typiquement $\lambda = 200$ nm, soit 600 paires de bases [69]) et w la largeur supposée des plectonèmes (quelques nanomètres, voir plus loin), il faut chercher parmi les segments situés à une distance supérieure à λ en « aval » de i (par rapport au sens de parcours de la molécule) s'il en existe dans la sphère de rayon w centrée en i . Si c'est le cas, un nouveau plectonème vient d'être trouvé. En considérant qu'il se termine au cylindre le plus en aval (cylindre j sur la figure 4.30), il est possible de calculer sa longueur. La recherche reprend alors à partir du cylindre $j + 1$ (ou $i + 1$ si aucun plectonème n'a été trouvé).

L'efficacité de cette méthode peut être constatée sur la figure 4.31. Elle est bien sûr limitée au cas particulier des molécules linéaires sous tension, pour laquelle aucun risque n'existe de voir se rapprocher deux points distants de la molécule autrement que par l'intermédiaire d'un plectonème. Cet outil peut être adapté à la détection de petits plectonèmes et de boucles en adaptant la résolution λ , mais il faut alors faire attention aux phénomènes purement statistiques qui introduiront inévitablement du bruit dans les résultats.

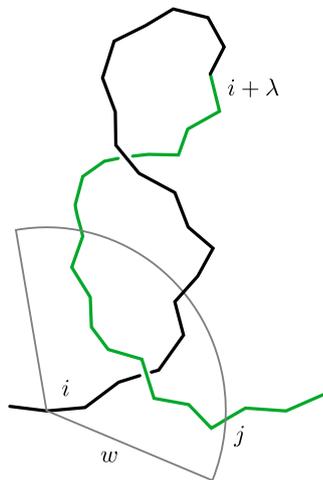


FIGURE 4.30: Schéma illustrant l'algorithme de détection des plectonèmes. Le cylindre i est le début d'un plectonème si, parmi les cylindres distants d'au moins λ le long de la chaîne (segments verts), au moins un se trouve à une distance de i inférieure à w (cercle gris). Le plectonème s'étend jusqu'au plus éloigné de ces cylindres le long de la chaîne (cylindre j).



FIGURE 4.31: Application de la détection automatique des plectonèmes à une molécule de 21 kb. Les différents plectonèmes trouvés sont notés ici par des couleurs différentes.

La largeur typique w des plectonèmes peut être estimée, en minimisant l'énergie requise pour ajouter une boucle de diamètre w à la molécule [62], à $w \approx 2\sqrt{\frac{l_p k_B T}{f}}$, ce qui correspond à $w = 16$ nm pour $f = 3,2$ pN, $w = 23$ nm pour $f = 1,6$ pN et $w = 32$ nm pour $f = 0,8$ pN. Toutefois, comme le montre la figure 4.32, la valeur précise de w n'a aucun effet sur le nombre ou la longueur des plectonèmes détectés par cet outil.

L'automatisation du comptage des plectonèmes permettant d'obtenir des moyennes assez précises, il est maintenant possible de tester quantitativement si le niveau de discrétisation influe sur ces mesures. Alors que $n = 5$ cylindres par l_p est un niveau clairement suffisant pour obtenir une extension et un couple corrects (Section 4.1), la figure 4.33 montre qu'il est insuffisant pour rendre compte avec précision du nombre et de la taille des plectonèmes. Cet effet peut se comprendre en réalisant que, pour $n = 5$, les cylindres mesurent

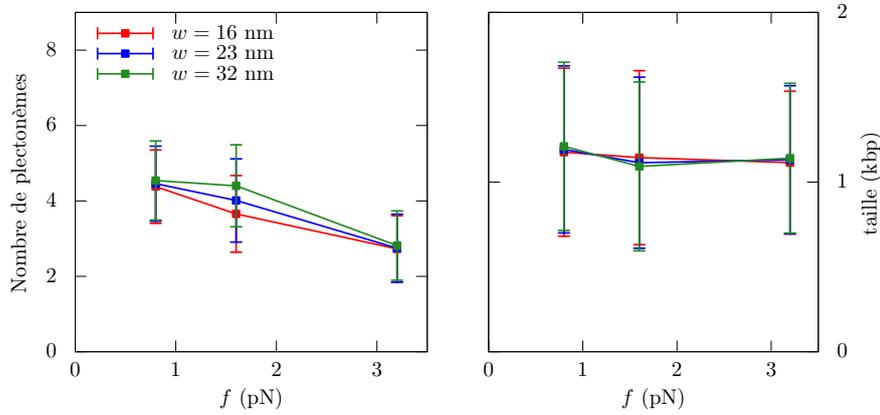


FIGURE 4.32: Effet de la largeur présumée w des plectonèmes sur le nombre et la taille des structures détectées automatiquement.

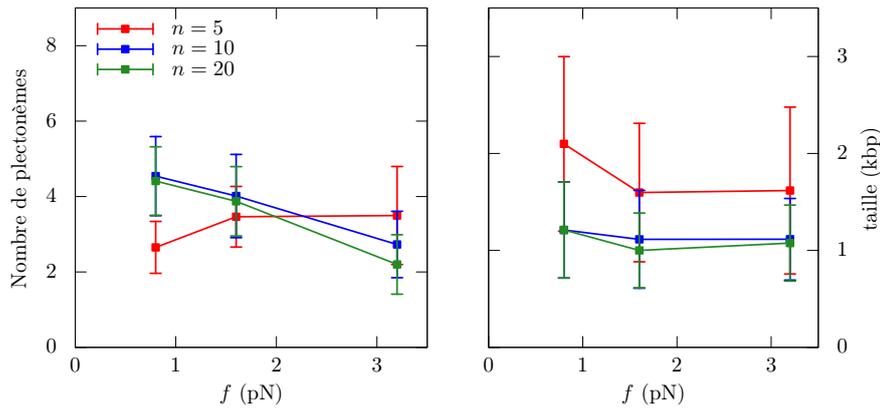


FIGURE 4.33: Effet de la discrétisation sur le nombre et la taille des structures détectées automatiquement (n est le nombre de cylindres par longueur de persistance).

10 nm de long tandis que les boucles et les extrémités des plectonèmes mesurent typiquement de 15 à 30 nm, ce qui veut dire qu'elles ne sont composées que d'une poignée de segments. Il est donc nécessaire d'utiliser une discrétisation plus fine ($n = 10$), ce que j'ai fait pour tous les résultats présentés ici.

La figure 4.34 montre que le nombre de plectonèmes obtenu par simulation est toujours supérieur à celui mesuré par Van Loenhout *et. al.* [69], y compris lorsque la résolution de la détection des plectonèmes est artificiellement adaptée à celle des expériences (500 nm, correspondant à la taille des taches fluorescentes observées) en fusionnant les plectonèmes proches. De plus, alors que ce nombre semble relativement constant dans les expériences, il diminue lorsque la force augmente dans les simulations.

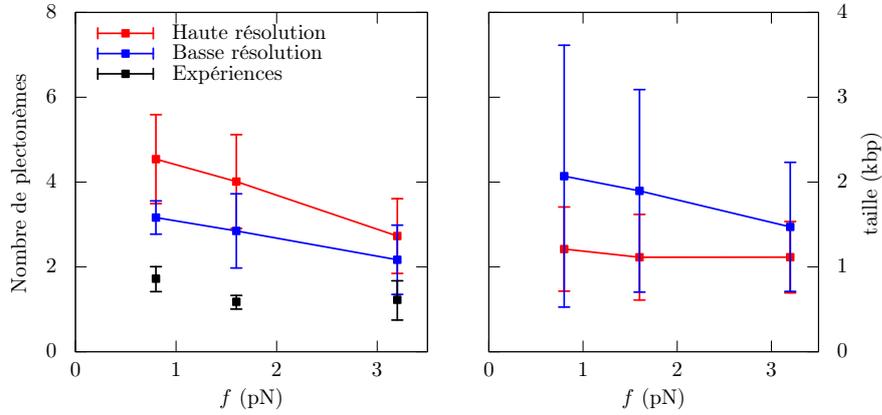


FIGURE 4.34: Nombre et taille des plectonèmes d’une molécule de 21 kb sur-enroulée de manière à diminuer l’extension de 25 % par rapport à une absence de surenroulement. Haute résolution (en rouge) : toutes les structures détectées (> 200 nm de long) sont comptées individuellement. Basse résolution (en bleu) : les plectonèmes séparés de moins de 500 nm sont fusionnés. En noir : résultats expérimentaux de Van Loenhout *et. al.* [69].

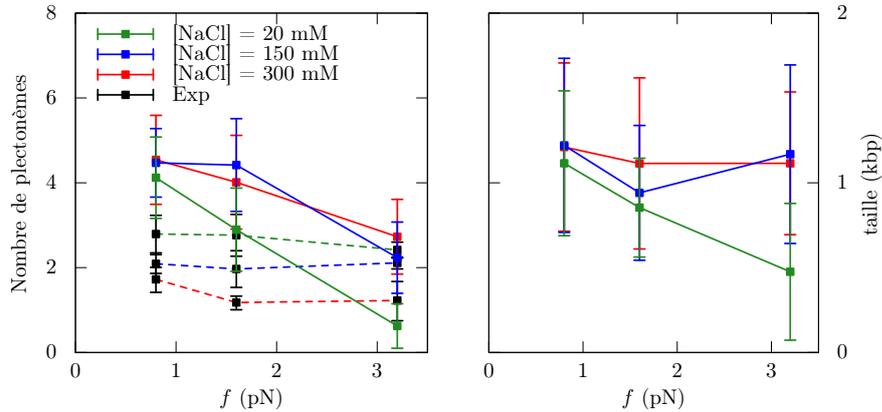


FIGURE 4.35: Effet de la concentration en sel sur le nombre et la taille des plectonèmes. Points noirs reliés par des lignes en pointillés : résultats expérimentaux de Van Loenhout *et. al.* [69].

Ces différences entre les simulations et l’expérience sont également visibles sur la figure 4.35 : dans les simulations, le nombre et la taille des plectonèmes ne varient pas lorsque la concentration en sel passe de 300 mM à 150 mM, puis diminuent, à force élevée, pour $[\text{NaCl}] = 20$ mM, contrairement aux résultats de Van Loenhout *et. al.* qui observent une augmentation du nombre de plectonèmes quand la concentration saline diminue.

Ces différences peuvent également être vues en comparant la densité d’ADN

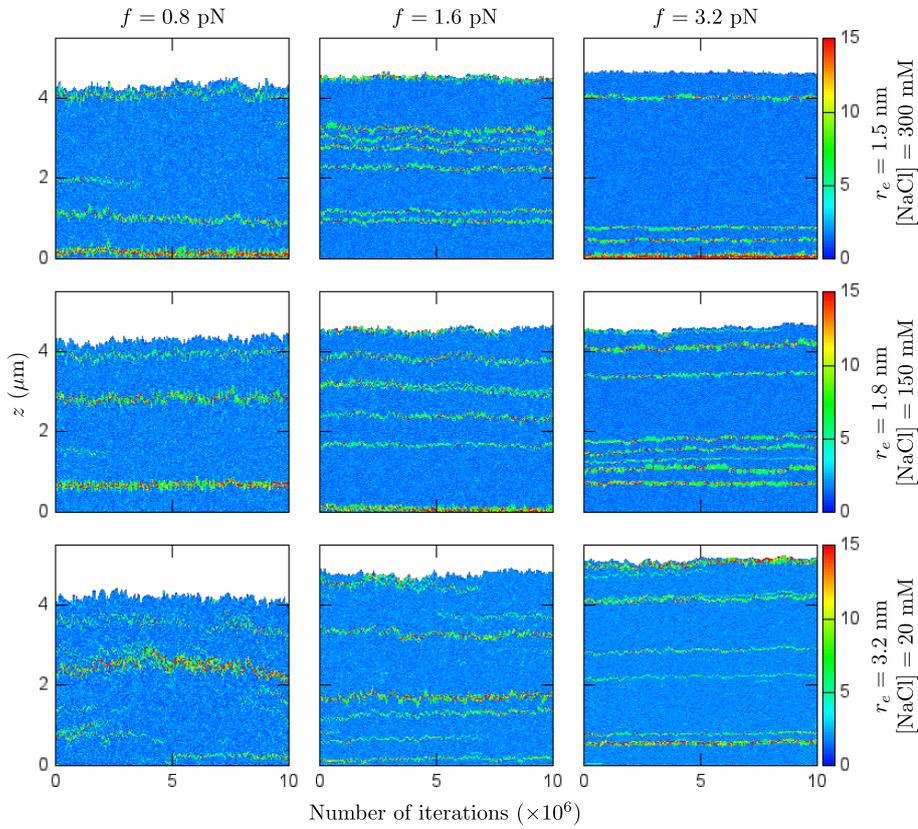


FIGURE 4.36: Traces obtenues pour les mêmes conditions que Van Loenhout *et. al.* [69] (molécule de 21 kb, surenroulement ajusté de manière à réduire l’extension de 25 %), pour différentes forces et concentrations en sel.

mesurée expérimentalement, visible sur la figure 4.36, à celle obtenue par des simulations dans les mêmes conditions (de forces et de concentrations en sel) : la dynamique mise à part, on constate qu’un plus grand nombre de plectonèmes est visible, en particulier à force élevée. Cependant, ces traces présentent un point commun, qualitatif, avec celles de Van Loenhout *et. al.* : lorsque la force et la salinité diminuent, les structures deviennent moins nettes, plus étalées, et se détachent moins nettement du bruit de fond.

L’explication la plus plausible à ces différences réside dans l’effet de séquence, avancé par les auteurs eux-mêmes pour justifier de la distribution non-homogène des plectonèmes le long de la molécule. En effet, la molécule est modélisée à l’aide d’une longueur de persistance supposée constante le long de la molécule, alors que la rigidité de l’ADN, à petite échelle, dépend de la séquence exacte des paires de bases. Ces défauts dans la rigidité de la molécule réelle sont alors probablement capables de piéger les plectonèmes à certaines positions précises, et peut-être de favoriser l’allongement de ceux-ci plutôt que la création de nouvelles structures.

Une autre explication possible tient au protocole expérimental. La molécule est en effet maintenue contre la surface transparente à travers laquelle les images

sont capturées, ce qui peut perturber grandement la manière dont les structures se forment (en réduisant notamment fortement l'entropie associée à la direction des plectonèmes, puisque tout un demi-espace leur est interdit).

Chapitre 5

Interactions à longue portée

L'IMPORTANCE de la formation de boucles à différentes échelles dans la régulation de la transcription motive l'étude des interactions entre sites distants le long de l'ADN et de l'effet du surenroulement et de l'existence de domaines topologiques indépendants. Ce chapitre présente les possibilités offertes par mon nouvel algorithme pour aborder ces questions. Premièrement, à grande échelle, la simulation du repliement de grandes molécules permet d'étudier l'influence du surenroulement sur la probabilité de contact en fonction de la distance génomique, ce que j'aborde en considérant les propriétés de repliement de molécules de 40 kb. Je discute ensuite la question de l'isolation de domaines indépendants est abordée, la définition locale du twist implémentée dans mon algorithme se révélant cruciale pour réaliser cette isolation.

5.1 Fréquence de contact

Expérimentalement, la probabilité de formation des boucles et des contacts à longue distance a été étudiée, chez les eucaryotes, par des techniques de « Chromosome Conformation Capture » (3C) [107] (voir [108] pour une revue sur les aspects expérimentaux de la 3C et de ses multiples dérivations), mais aussi chez les procaryotes, via la mesure du niveau de répression de gènes codant pour un rapporteur fluorescent [27, 3], notamment à longue portée [44, 49]. Du point de vue de la modélisation, si de nombreux travaux impliquant de conséquentes simulations se sont attaqués au problème des probabilités de contact (aussi appelé J-facteur) pour les eucaryotes [109, 110, 111, 112], incluant parfois une notion de surenroulement [113], les résultats pour de l'ADN nu surenroulé sont limités à de faibles distances (3 kb) [28]. La difficulté de simuler de longues molécules surenroulées a, en effet, fortement limité l'étude de la probabilité de contact de sites distants le long du chromosome bactérien. Les résultats présentés ici pour une molécule de 40 kb, bien que préliminaires, montrent déjà un comportement intéressant à grande échelle.

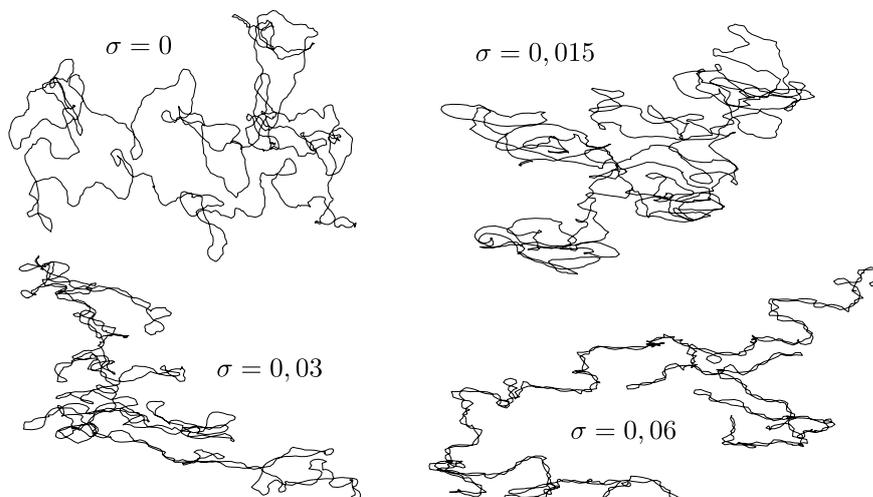


FIGURE 5.1: Exemples de configurations obtenues lors de simulations de molécules circulaires de 40 kb, pour différents niveaux de surenroulement.

Contacts entre sites distants

Afin d'étudier l'effet du surenroulement sur la fréquence de contact de sites distants le long de l'ADN, j'ai réalisé des simulations d'une molécule circulaire de 40 kb à différents niveaux de surenroulement (Fig. 5.1).

Étant donnée une configuration de la molécule, deux cylindres i et j sont considérés en contact si leur distance spatiale $|\vec{r}_j - \vec{r}_i|$ est inférieure à un seuil de contact Δ_c de l'ordre de quelques nanomètres ou dizaines de nanomètres, c'est-à-dire approximativement la taille d'une protéine ou d'un complexe susceptible de se fixer aux deux sites en question. En échantillonnant les configurations générées au cours d'une simulation, il est alors possible d'estimer la fréquence des contacts entre i et j (le nombre de configurations où ils sont en contact divisé par le nombre total de configurations). La figure 5.2 présente les matrices de contacts ainsi obtenues au cours d'une simulation individuelle pour les quatre niveaux de surenroulement étudiés ici.

Les structures transversales de la matrice de contact (perpendiculaires à la diagonale principale) apparaissant à surenroulement non nul correspondent aux différentes branches des plectonèmes visibles sur la figure 5.1, et sont d'autant mieux définies que le surenroulement est élevé : les plectonèmes étant très stables, de nombreuses configurations contribuent aux mêmes contacts, ce qui accentue le contraste de ces matrices. Pour $\sigma = 0,06$ par exemple, de nombreux contacts hors diagonale ont une fréquence égale à 1, ce qui signifie que les deux sites en question sont restés proches l'un de l'autre pendant toute la durée des simulations (en l'occurrence quinze millions de pas Monte-Carlo).

La figure 5.3 montre les matrices de contacts moyennes obtenues à partir de quatre simulations indépendantes pour chaque valeur du surenroulement. À faible surenroulement, le bruit est ainsi diminué, tandis qu'à surenroulement élevé, un plus grand nombre de structures apparaît, distribué plus uniformément le long de la molécule.

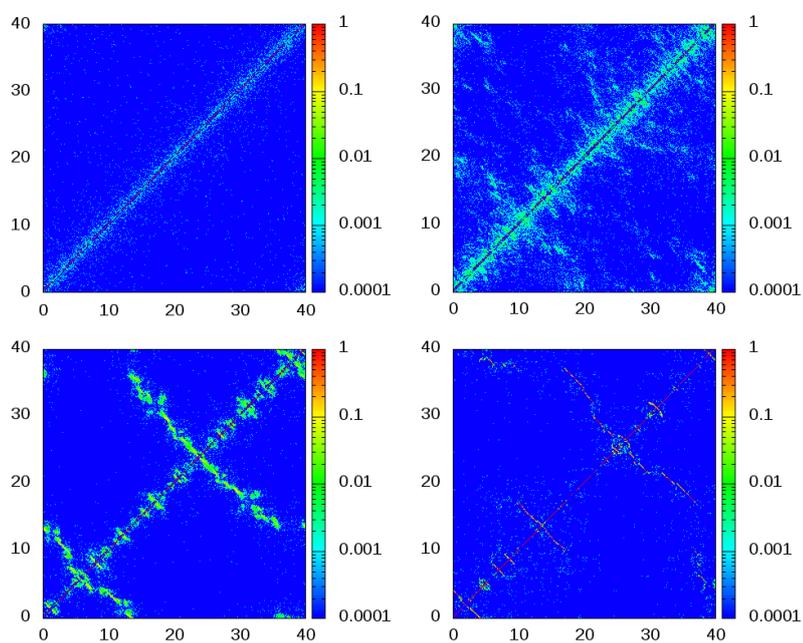


FIGURE 5.2: Matrices de contact obtenues pour différents niveaux de surenroulement. Le seuil de contact vaut ici $\Delta_c = 15$ nm.

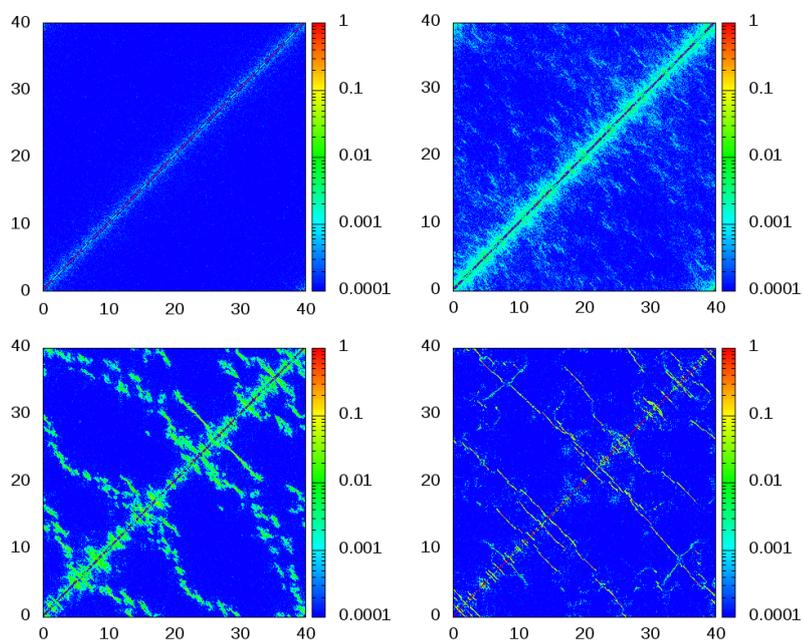


FIGURE 5.3: Matrices de contact moyennes obtenues à partir de quatre simulations indépendantes pour chaque valeur du surenroulement.

Contacts en fonction de la distance

À partir des données de contact entre paires de sites, il est facile de calculer la fréquence de contact P en fonction de la distance s entre sites. Le comportement de cette fonction dépend fortement de σ .

Faible surenroulement

À surenroulement nul ou faible, deux régimes sont visibles (Fig. 5.4). Lorsque la distance (1D) entre les deux sites est de l'ordre de la longueur de persistance, la fréquence de contact est réduite du fait de la rigidité du polymère. Cette fréquence atteint un maximum pour une distance d'environ 3 ou 4 l_p , avant d'entrer dans un régime où elle décroît suivant une loi de puissance de la forme :

$$P(s) \sim s^{-\alpha} \quad (5.1)$$

Afin d'estimer l'exposant α , il est nécessaire de n'utiliser, pour la régression, que les points situés au delà d'un certain seuil de distance s_0 . Comme le montre la figure 5.5, les exposants calculés sont indépendants de s_0 au delà de 8 à 10 l_p environ, convergeant vers une valeur d'environ 1,8 pour $\sigma = 0$ et 1,5 pour $\sigma = 0,015$.

L'exposant $\alpha = 1,8$ observé à surenroulement nul peut être obtenu par un argument d'échelle en considérant qu'à grande distance génomique s , le polymère peut être décrit par une chaîne gaussienne. Dans ce cas, la probabilité que l'extension de la chaîne soit trouvée dans l'intervalle $[r, r + dr]$ est donnée par [77] :

$$p(r)dr \sim \frac{r^2}{R^3} \exp\left(-\frac{3r^2}{2R^2}\right) dr \quad (5.2)$$

où $R = \langle r \rangle$ est l'extension moyenne. La probabilité $P(s)$ que les deux extrémités de la chaîne de longueur s soient en contact s'écrit :

$$P(s) = \int_{2r_e}^{\Delta_c} p(r)dr \quad (5.3)$$

où la borne inférieure de l'intégrale traduit le potentiel de cœur dur et la borne supérieure correspond au seuil de contact $\Delta_c \ll R$. $P(s)$ varie donc comme R^{-3} .

L'extension moyenne R dépend elle-même différemment de s suivant le système considéré. Par exemple, pour une chaîne idéale, $R(s) \sim s^{1/2}$ et $P(s) \sim s^{-3/2}$ [111]. Pour une chaîne compacte dans un globule à l'équilibre, $R(s) \sim s^{1/2}$ également, si $s < L^{2/3}$ [114, 115]. Pour un globule fractal, en revanche, $R(s) \sim s^{1/3}$ et $P(s) \sim s^{-1}$ [114, 115]. Enfin, pour un modèle Worm-Like Chain auto-évitant, $R(s) \sim s^{3/5}$ [77], soit $P(s) \sim s^{-9/5} = s^{-1,8}$.

Ainsi, si un faible niveau de surenroulement augmente de manière générale la fréquence de contact à toutes les échelles (Fig. 5.4), la diminution de l'exposant α signifie également que cet effet est d'autant plus important que la distance génomique est grande.

Enfin, la figure 5.6 confirme que le choix du seuil de contact Δ_c n'a pas d'influence sur la forme des courbes de la fréquence de contact en fonction de la distance, ni sur l'exposant α . Bien entendu, la fréquence de contact augmente avec Δ_c , mais, pour peu que ce seuil soit suffisamment élevé pour que les

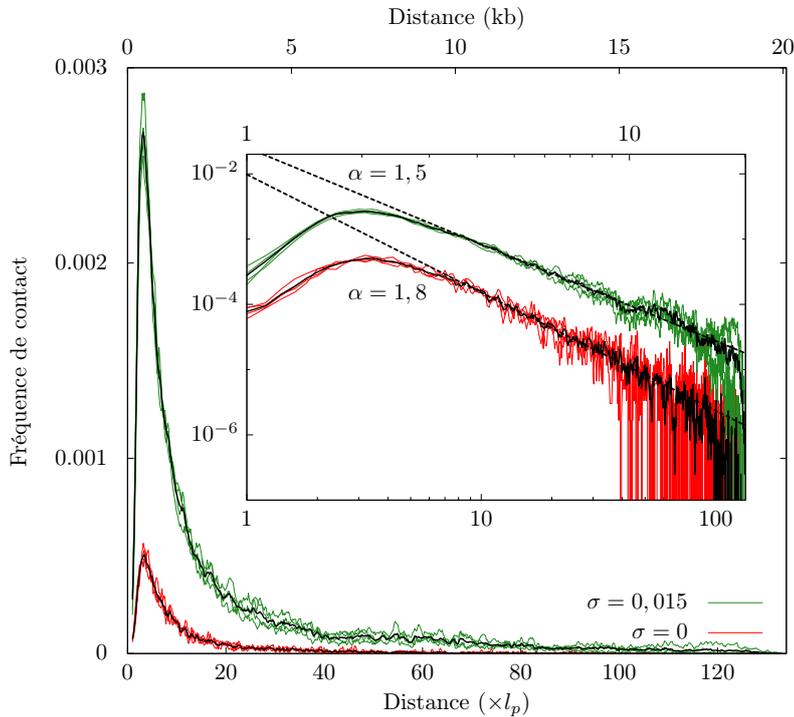


FIGURE 5.4: Fréquence de contact en fonction de la distance le long de la molécule (mesurée en nombre de l_p sur l'échelle du bas et en kb sur l'échelle du haut). Seuil de contact; $\Delta_c = 15$ nm. Pour chacun des deux niveau de surenroulement, quatre simulations indépendantes ont produit quatre courbes (de la même couleur) dont la moyenne est tracée en noir. Les droites de régression sont tracées en pointillés sur le graphe logarithmique, et les exposants α correspondants indiqués sur la figure. La distance minimale utilisée pour ces fits vaut $s_0 = 11l_p$.

courbes soient lisses ($\Delta_c \geq 10$ environ), l'atténuation de la fréquence de contact avec la distance s'effectue au même rythme.

Surenroulement élevé

À surenroulement élevé, la stabilité des structures se traduit par une grande variabilité entre différentes simulations, ce qui rend difficile l'interprétation des résultats. Cependant, la figure 5.7 semble indiquer une forte augmentation de la fréquence de contact à grande échelle, et suggèrent l'apparition de nouveaux régimes : après une phase de décroissance en loi de puissance, la fréquence de contact atteint un plateau, puis décroît très rapidement.

Ces résultats préliminaires laissent donc entrevoir un effet non trivial du surenroulement sur la probabilité de former des boucles à grande distance, et appellent une étude plus détaillée de ce phénomène, impliquant plus de simulations pour affiner les résultats, ainsi que des simulations de molécules de

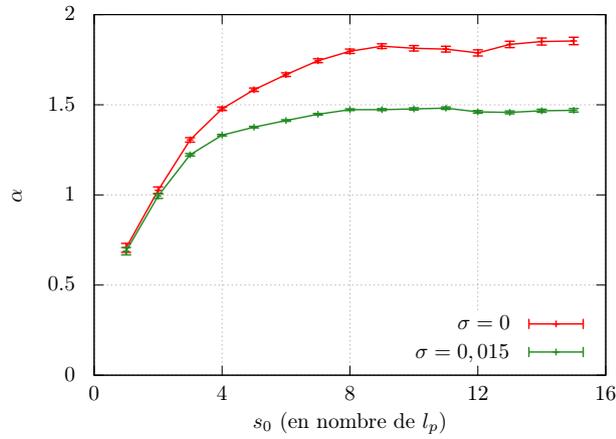


FIGURE 5.5: Valeurs de l'exposant α obtenues par régression des courbes moyennes de la figure 5.4 en fonction de la distance minimale s_0 des points utilisés dans la régression.

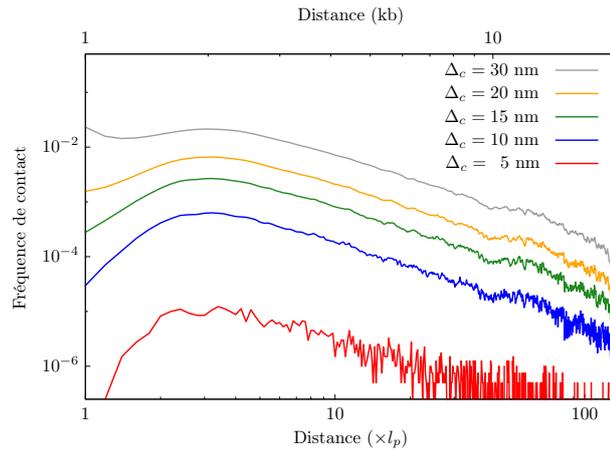


FIGURE 5.6: Fréquence de contact en fonction de la distance entre les deux sites le long de la molécule, pour différents seuils de contact ($\sigma = 0,015$).

plus grande taille encore. L'introduction de micro-domaines tels que présentés dans la section suivante permettra également d'étudier l'effet de ces derniers sur la probabilité de contact intra-domaine et inter-domaine.

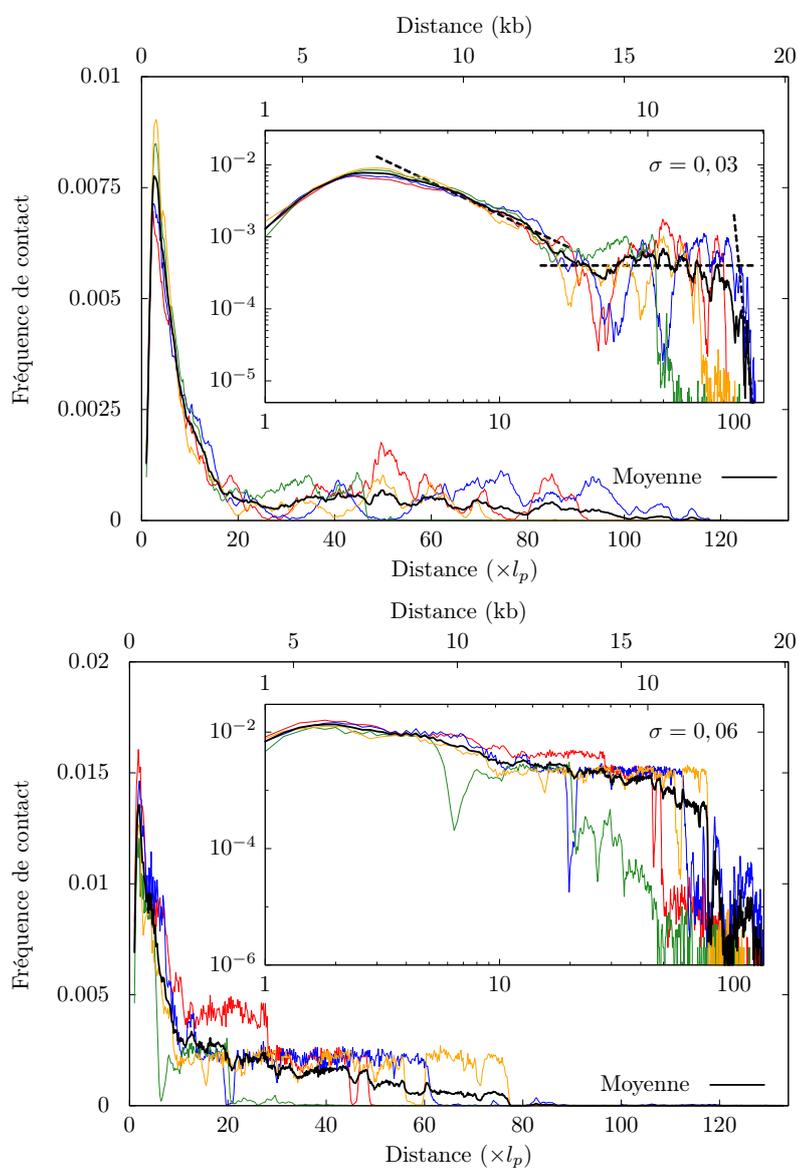


FIGURE 5.7: Fréquence de contact en fonction de la distance entre les deux sites le long de la molécule à surenroulement élevé. Les quatre courbes de couleur correspondent à quatre simulations indépendantes. La moyenne de ces quatre courbes est noire. Pour $\sigma = 0,03$ (en haut), des lignes en pointillés indiquent les trois régimes suggérés par les courbes.

5.2 Domaines indépendants

Modélisation d'un isolant topologique

LE FAIT de définir un twist local pour les cylindres permet d'utiliser la version locale de l'énergie de torsion (Section 2.3), et donc de simuler des domaines soumis à des niveaux différents de surenroulement. La question se pose alors de déterminer dans quelles conditions il est possible d'isoler deux domaines topologiquement indépendants, comme peuvent le faire certaines protéines, éventuellement sous forme de multimères [43] (Section 2.1).

Le pont créé par la protéine entre deux sites distants de la molécule permet de clore les deux domaines et donc de définir un linking number indépendant pour chacun d'eux (Fig. 5.8). L'isolation consiste alors à empêcher l'échange de linking number entre ces deux domaines. Ces domaines peuvent également être des fragments linéaires dont les extrémités sont contraintes topologiquement à la manière de pinces magnétiques, ces contraintes pouvant par exemple représenter l'ancrage à la membrane. Dans tous les cas, il est nécessaire d'empêcher le transfert de writhe entre les domaines en contraignant la géométrie du pont, mais également l'échange de twist, ce qui n'est possible qu'avec une définition locale de celui-ci dans les cylindres.

Afin d'avoir un modèle le plus simple et généraliste possible de la protéine, celle-ci est représentée par deux segments de longueur l (typiquement quelques nanomètres), faisant un angle α_0 entre eux, fixés chacun à un site le long de la molécule, et tendant à s'aligner avec la direction du brin au niveau de ces sites de fixation (Fig. 5.9). Au cours d'une simulation, ces deux « bras » vont s'étirer de manière à rester accrochés entre eux et à leurs sites de fixation, leur déformation étant restreinte en interdisant aux fluctuations de $l_1 - l$, $l_2 - l$, $\alpha - \alpha_0$, β_1 et β_2 de dépasser une amplitude donnée (quelques degrés pour les angles et quelques dixièmes de nanomètres pour les longueurs). Pour tester la possibilité de créer deux domaines indépendants, le puits d'énergie du pontage est considéré comme infiniment profond.

Dans ce contexte, pour réaliser l'isolation, il suffit ensuite de considérer les deux bras de la protéine comme deux cylindres autour desquels s'enroule le brin, comme dans le reste du domaine, et d'imposer la continuité de ce brin dans la

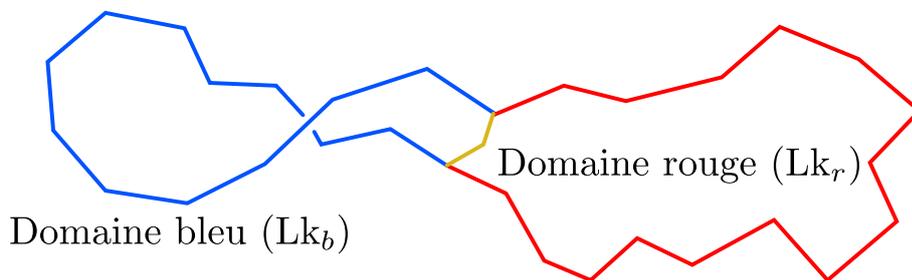


FIGURE 5.8: Schéma d'une molécule circulaire divisée en deux domaines, « bleu » et « rouge », par une protéine (en jaune) joignant deux sites de fixation. Chaque domaine, rendu circulaire par le pontage de la protéine, possède un linking number bien défini et indépendant.

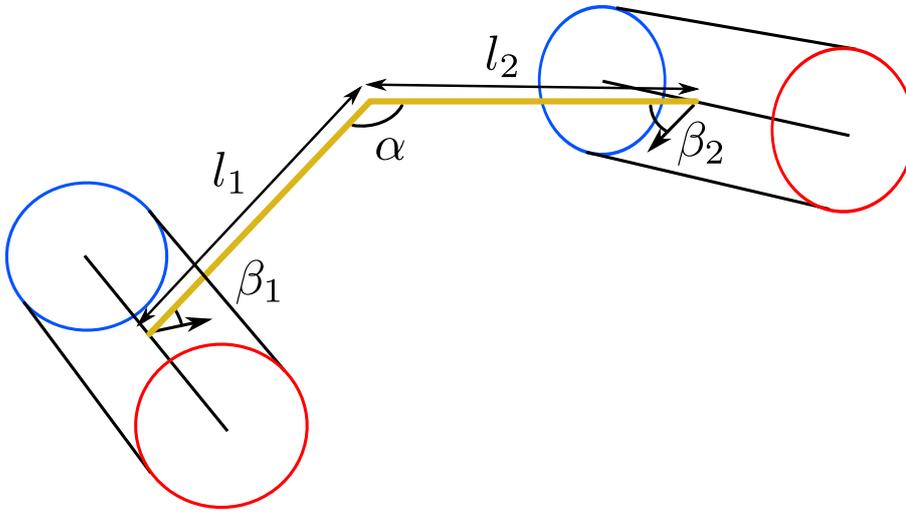


FIGURE 5.9: Schéma d'une « protéine » (en jaune) joignant deux sites situés sur deux cylindres différents. Les deux bras, de longueur l_1 et l_2 , font des angles β_1 et β_2 avec la direction du brin au niveau des sites de fixation (flèches noires), et un angle α entre eux. La protéine sépare la molécule en deux domaines, « bleu » (à gauche) et « rouge » (à droite).

protéine et au niveau de ses sites de fixation, pour conserver le linking number dans ce domaine. Le point crucial est alors de considérer deux brins s'enroulant de façon indépendante. Dans ce cas, et seulement dans ce cas, le linking number des deux domaines sera conservé, indépendamment l'un de l'autre (Fig. 5.10). En contraignant la torsion de chacun des deux brins à l'intérieur de la protéine, le surenroulement d'un domaine ne peut pas être absorbé par celle-ci sous forme de twist.

Isolation de deux domaines

L'isolation de deux domaines peut donc facilement être implémentée à l'aide de deux cylindres représentant les bras de la protéine et en considérant deux types de brin à l'intérieur de ces cylindres, l'un correspondant au domaine bleu et l'autre au domaine rouge, le principe de la simulation restant inchangé. Ainsi, partant d'une configuration initiale sans aucun surenroulement, il est possible de vérifier que l'introduction de surenroulement dans le domaine bleu, par exemple, a pour effet la formation d'un plectonème dans ce domaine, tandis que le domaine rouge n'est pas affecté (Fig. 5.11).

Dans sa version actuelle, cependant, l'algorithme n'est pas encore capable de détecter les croisements entre la protéine et les cylindres (principalement parce que le mouvement des bras de la protéine au cours d'un pas Monte-Carlo est différent de celui des cylindres, tout en étant simultané). Comme le montre la figure 5.12, ces croisements peuvent modifier les linking numbers des domaines rouge et bleu, compromettant de fait leur indépendance. La solution temporaire à ce problème est de calculer les writhes des deux domaines afin de vérifier la conservation de Lk_r et Lk_b , ce qui limite pour le moment les

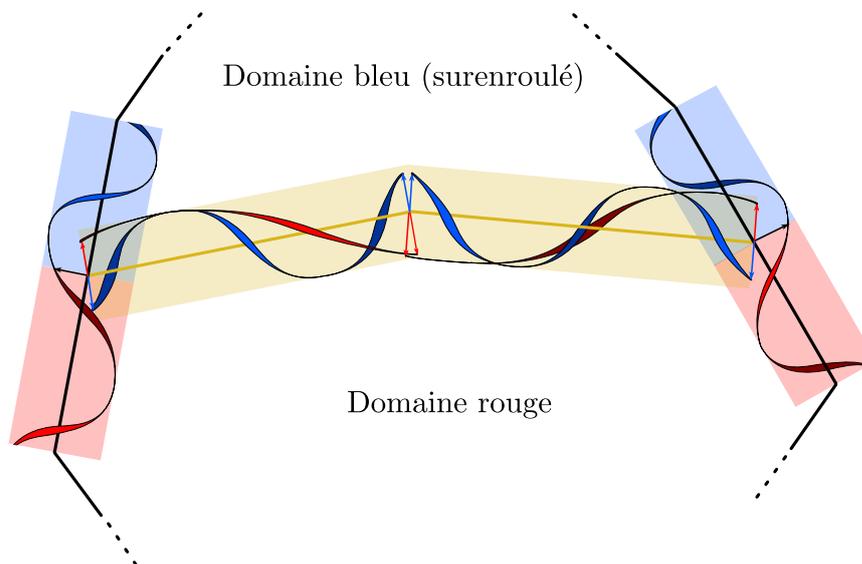


FIGURE 5.10: Schéma du principe de continuité du brin (Section 3.1) appliqué à la protéine (en jaune) joignant deux sites situés sur deux cylindres différents (en gris). Pour conserver le linking number du domaine rouge, le brin rouge doit rester continu, en particulier aux points d'attache des bras sur les cylindres. En ces points, la direction du brin dans le cylindre (flèche noire) doit tourner « à plat » autour du pivot, ce qui la transforme en la flèche rouge. La continuité doit également être respectée à l'intersection des deux bras (au centre). Indépendamment, la continuité doit aussi être respectée pour le brin bleu (flèches bleues) afin de conserver le linking number du domaine bleu. Ce dernier étant surenroulé, un twist assez important a été transféré à la protéine (hélice bleue). Ce twist ne peut toutefois pas être communiqué au domaine rouge car celui-ci n'échange que les hélices rouges avec la protéine.

simulations à de petites molécules. Les exemples expérimentaux auxquels sont confrontées les simulations [43] correspondent eux-mêmes à des molécules de petite taille (plasmides de quelques kb), pour lesquelles le calcul du writhe est rapide.

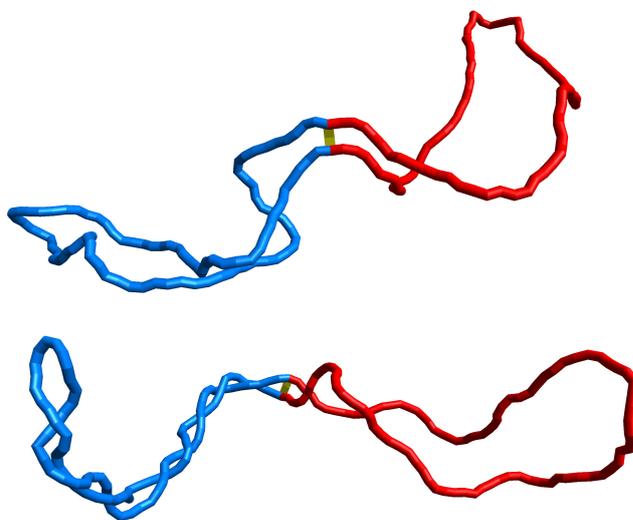


FIGURE 5.11: Exemple de configuration d'une molécule de 3 kb divisée en deux domaines (rouge et bleu) par une protéine (jaune), avant (en haut) et après (en bas) l'introduction d'un surenroulement $\Delta Lk_b = 10$ dans le domaine bleu. Le domaine rouge n'est pas affecté : les domaines sont topologiquement isolés.

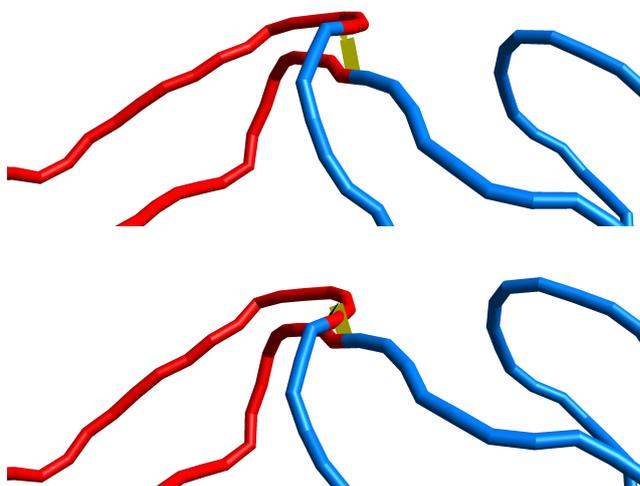


FIGURE 5.12: Exemple de croisement de la protéine par un cylindre au cours d'une rotation (en haut : avant la rotation, en bas : après). Un tel croisement ne change pas le linking number total de la molécule (puisque aucun croisement n'a eu lieu entre les cylindres représentant l'ADN, qu'ils soient rouges ou bleus), mais modifie le linking number du domaine rouge (un cylindre rouge est passé à travers un bras jaune de la protéine).

Chapitre 6

Conclusion

LE TRAVAIL présenté ici permet, pour la toute première fois, d'envisager la simulation du repliement d'un chromosome bactérien entier en prenant en compte sa superhélicité.

En effet, ce nouvel algorithme est basé sur un calcul local du twist de la molécule, ce qui présente deux avantages. D'une part, cette méthode permet de s'affranchir du calcul coûteux du writhe. L'impact sur les performances des simulations est sans équivoque : l'influence de la taille de la molécule sur le temps de calcul passe d'un exposant 2 à un exposant 1,2. Ainsi, des tailles de l'ordre de la dizaine, voire de la centaine de kb deviennent facilement accessibles aux simulations, ce qui ouvre la voie à l'étude du repliement de longues régions du chromosome, par exemple formées de plusieurs microdomaines, voire tout un macrodomaine (par l'utilisation de la parallélisation si nécessaire). Il est donc envisageable de simuler un chromosome entier (le chromosome de *M. genitalium* [32], long de 580 kb, a une taille comparable à un seul macrodomaine d'*E. coli*).

D'autre part, l'utilisation d'un twist local augmente les possibilités des simulations vis-à-vis des contraintes topologiques. Ainsi, cet algorithme nous a permis de définir un réservoir de twist pour déterminer le couple de la molécule (et son module de torsion effectif) et de constater le très bon accord du modèle Worm-Like Chain surenroulé avec les résultats expérimentaux ; ce qu'aucune simulation de Monte-Carlo n'avait pu réaliser jusqu'ici, à ma connaissance. La forte dépendance du couple de flambage au rayon effectif de la molécule met cependant en lumière une possible limite du modèle utilisé ici pour représenter la répulsion électrostatique de l'ADN, et suggère de tester l'effet d'autres types de potentiels, plus détaillés que le cœur dur, sur les propriétés de la molécule ; en particulier lorsque le repliement en plectonèmes provoque de nombreuses collisions.

Il est d'ailleurs à noter que les simulations effectuées ici supposent la molécule plongée dans une solution d'ions monovalents (chlorure de sodium), ce qui correspond à la situation des expériences *in vitro* évoquées. La présence d'ions multivalents a cependant un effet important sur les propriétés physiques

de la molécule (notamment son rayon effectif [116] et sa longueur de persistance [89]), et est donc susceptible de modifier significativement les structures formées par l'ADN surenroulé.

Les expériences de molécule unique révèlent également une autre limite du modèle : le comportement de l'ADN étiré n'est pas symétrique vis-à-vis du surenroulement. En effet, pour des forces de l'ordre de 1 pN et à surenroulement négatif, aucune transition de flambage n'est visible dans les courbes force-extension [67]. À la place, l'appariement des bases se brise en certains endroits, donnant lieu à des bulles de dénaturation [54] composées d'ADN simple-brin dont les propriétés physiques sont différentes de celles de l'ADN double-brin (longueur de persistance beaucoup plus courte, notamment). Une telle transition structurale de la molécule n'est pas prise en compte dans le modèle actuel. Elle pourrait cependant présenter un intérêt biologique, étant donné que l'ADN de la plupart des bactéries est négativement surenroulé et donc potentiellement dénaturé par endroits, sous certaines conditions de fortes contraintes mécaniques. Ici aussi, une approche locale des contraintes est susceptible de résoudre ce problème.

La prise en compte locale du twist et l'approche consistant à préserver la continuité du brin pour respecter la contrainte de conservation du linking number ont également permis de réaliser un système relativement simple constitué de deux domaines indépendants. Cette approche semble prometteuse et pourrait s'avérer fructueuse pour modéliser d'autres types de barrières topologiques, telles que la machine de transcription, l'ancrage à une membrane ou la fourche de réplication, par exemple, ou encore l'effet de topoisomérases [117] ou de cassures simple-brin.

La nécessité d'un algorithme performant s'est avérée d'autant plus cruciale que certaines questions requièrent un niveau de discrétisation particulièrement détaillé. En particulier, l'étude du nombre de plectonèmes formés le long d'une grande molécule impose une description fine de la courbure à l'échelle des boucles constituant ces structures. Mes simulations nous ont permis de constater l'écart entre les prédictions du modèle Worm-Like Chain surenroulé et les mesures expérimentales [69] à ce sujet, et suggèrent que la prise en compte de la séquence pourrait permettre un meilleur accord avec les observations. Par exemple, l'utilisation d'une longueur de persistance non plus homogène, mais modulée par la séquence (notamment la densité de guanine et cytosine par rapport à celle d'adénine et de thymine), est susceptible de jouer sur le nombre, la taille et la localisation des structures formées.

Enfin, la comparaison avec les expériences de Van Loenhout *et. al.* [69] souligne par ailleurs une des limites de ces simulations : l'impossibilité d'étudier la dynamique du repliement de l'ADN, qui est le prix à payer pour le gain en performances d'une approche Monte-Carlo par rapport à des techniques de dynamique brownienne et de dynamique de Langevin, par exemple. Néanmoins, cet algorithme constitue un outil multi-échelles puissant pour étudier la colocalisation de gènes distants (notamment régulés par un même facteur de transcription), leur fréquence de contact et leur éventuelle aggrégation sous forme d'usine de transcription, ce qui n'était possible jusqu'ici qu'en ignorant l'effet du surenroulement [9].

Ces simulations constituent en particulier un outil intéressant pour la biologie de synthèse, en offrant la possibilité de prédire le degré de colocalisation d'un ensemble donné de gènes régulés par un même facteur de transcription

(ou formant un réseau de régulation plus complexe), et d'étudier l'influence de la position de ces gènes sur leur co-régulation. Les résultats pourraient alors être comparés à des mesures expérimentales de la transcription de ces gènes en fonction de leur position génomique pour mieux comprendre la coopération entre facteurs de transcription, protéines structurantes, surenroulement et micro-domaines notamment, permettant la colocalisation de ces gènes. À plus long terme, ces travaux pourraient servir de base à un outil de « conception assistée par ordinateur » intégrant les nombreuses contraintes imposées à l'organisation du génome [118] et permettant de déterminer la meilleure façon de placer un ensemble de gènes le long du chromosome de manière à optimiser leur transcription, dans le but de créer des voies métaboliques synthétiques rendant une bactérie capable de produire efficacement une molécule d'intérêt (médicament, biocarburant, etc.).

Bibliographie

- [1] F. Jacob and J. Monod, “Genetic regulatory mechanisms in the synthesis of proteins.”, *J Mol Biol*, vol. 3, pp. 318–356, 1961.
- [2] B. Müller-Hill, “Some repressors of bacterial transcription.”, *Curr Opin Microbiol*, vol. 1, pp. 145–151, 1998.
- [3] J. M. G. Vilar and S. Leibler, “DNA looping and physical constraints on transcription regulation.”, *J Mol Biol*, vol. 331, pp. 981–989, 2003.
- [4] L. Saiz, J. M. Rubi, and J. M. G. Vilar, “Inferring the in vivo looping properties of DNA”, *Proc Natl Acad Sci USA*, vol. 102, no. 49, pp. 17642–17645, 2005.
- [5] I. B. Dodd, K. E. Shearwin, A. J. Perkins, T. Burr, A. Hochschild, and J. B. Egan, “Cooperativity in long-range gene regulation by the lambda CI repressor.”, *Genes Dev*, vol. 18, pp. 344–354, 2004.
- [6] F. Képès, “Periodic transcriptional organization of the E.coli genome.”, *J Mol Biol*, vol. 340, pp. 957–964, 2004.
- [7] G. Kolesov, Z. Wunderlich, O. N. Laikova, M. S. Gelfand, and L. A. Mirny, “How gene order is influenced by the biophysics of transcription regulation.”, *Proc Natl Acad Sci USA*, vol. 104, pp. 13948–13953, 2007.
- [8] I. Junier, J. Hérisson, and F. Képès, “Periodic pattern detection in sparse boolean sequences.”, *Algorithms Mol Biol*, vol. 5, p. 31, 2010.
- [9] I. Junier, O. Martin, and F. Képès, “Spatial and Topological Organization of DNA Chains Induced by Gene Co-localization”, *PLoS Comput Biol*, vol. 6, p. e1000678, 2010.
- [10] M. Elati, R. Nicolle, I. Junier, D. Fernández, R. Fekih, J. Font, and F. Képès, “PreCisIon : PREdiction of CIS-regulatory elements improved by gene’s positION”, *Nucleic Acids Research*, 2012.
- [11] D. J. Jin and J. E. Cabrera, “Coupling the distribution of RNA polymerase to global gene regulation and the dynamic structure of the bacterial nucleoid in Escherichia coli.”, *J Struct Biol*, vol. 156, pp. 284–291, 2006.
- [12] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, “Quantifying E. coli proteome and transcriptome with

- single-molecule sensitivity in single cells.”, *Science*, vol. 329, pp. 533–538, 2010.
- [13] T. E. Kuhlman and E. C. Cox, “Gene location and DNA density determine transcription factor distributions in *Escherichia coli*”, *Mol Syst Biol*, vol. 8, 2012.
- [14] W. Wang, G. W. Li, C. Chen, X. S. Xie, and X. Zhuang, “Chromosome Organization by a Nucleoid-Associated Protein in Live Bacteria”, *Science*, vol. 333, no. 6048, p. 1445–1449, 2011.
- [15] X. Weng and J. Xiao, “Spatial organization of transcription in bacterial cells”, *Trends in Genetics*, vol. 30, no. 7, pp. 287–297, 2014.
- [16] B. Müller-Hill, “The function of auxiliary operators”, *Molecular Microbiology*, vol. 29, no. 1, pp. 13–18, 1998.
- [17] O. G. Berg, R. B. Winter, and P. H. von Hippel, “Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory.”, *Biochemistry*, vol. 20, pp. 6929–6948, 1981.
- [18] R. B. Winter and P. H. von Hippel, “Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The *Escherichia coli* repressor–operator interaction : equilibrium measurements.”, *Biochemistry*, vol. 20, pp. 6948–6960, 1981.
- [19] R. B. Winter, O. G. Berg, and P. H. von Hippel, “Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor–operator interaction : kinetic measurements and conclusions.”, *Biochemistry*, vol. 20, pp. 6961–6977, 1981.
- [20] T. Ruusala and D. M. Crothers, “Sliding and intermolecular transfer of the lac repressor : kinetic perturbation of a reaction intermediate by a distant DNA sequence.”, *Proc Natl Acad Sci USA*, vol. 89, pp. 4903–4907, 1992.
- [21] F. Képès and C. Vaillant, “Transcription-Based Solenoidal Model of Chromosomes”, *ComplexUs*, vol. 1, no. 4, pp. 171–180, 2003.
- [22] S. Schoenfelder, T. Sexton, L. Chakalova, N. F. Cope, A. Horton, S. Andrews, S. Kurukuti, J. A. Mitchell, D. Umlauf, D. S. Dimitrova, C. H. Eskiw, Y. Luo, C.-L. Wei, Y. Ruan, J. J. Bieker, and P. Fraser, “Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells.”, *Nat Genet*, vol. 42, pp. 53–61, 2010.
- [23] M. Valens, S. Penaud, M. Rossignol, F. Cornet, and F. Boccard, “Macrodomain organization of the *Escherichia coli* chromosome”, *The EMBO Journal*, vol. 23, no. 21, pp. 4330–4341, 2004.
- [24] O. Espéli, R. Mercier, and F. Boccard, “DNA dynamics vary according to macrodomain topography in the *E. coli* chromosome.”, *Mol Microbiol*, vol. 68, pp. 1418–1427, 2008.
- [25] I. Junier, F. Boccard, and O. Espéli, “Polymer modeling of the *E. coli* genome reveals the involvement of locus positioning and macrodomain structuring for the control of chromosome conformation and segregation.”, *Nucleic Acids Res*, vol. 42, pp. 1461–1473, 2014.

- [26] S. Meyer and G. Beslon, “Torsion-Mediated Interaction between Adjacent Genes”, *PLoS Comput Biol*, vol. 10, p. e1003785, 2014.
- [27] J. Müller, S. Oehler, and B. Müller-Hill, “Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator.”, *J Mol Biol*, vol. 257, pp. 21–29, 1996.
- [28] J. Huang, T. Schlick, and A. Vologodskii, “Dynamics of site juxtaposition in supercoiled DNA”, *Proc Natl Acad Sci USA*, vol. 98, no. 3, pp. 968–973, 2001.
- [29] T. B. K. Le, M. V. Imakaev, L. A. Mirny, and M. T. Laub, “High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome”, *Science*, vol. 342, no. 6159, pp. 731–734, 2013.
- [30] A. Travers and G. Muskhelishvili, “DNA supercoiling - a global transcriptional regulator for enterobacterial growth?”, *Nat Rev Micro*, vol. 3, pp. 157–169, 2005.
- [31] B. J. Peter, J. Arsuaga, A. M. Breier, A. B. Khodursky, P. O. Brown, and N. R. Cozzarelli, “Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*.”, *Genome Biol*, vol. 5, no. 11, p. R87, 2004.
- [32] C. J. Dorman, “Regulation of transcription by DNA supercoiling in *Mycoplasma genitalium* : global control in the smallest known self-replicating genome”, *Molecular Microbiology*, vol. 81, no. 2, pp. 302–304, 2011.
- [33] E. P. C. Rocha, “The organization of the bacterial genome.”, *Annu Rev Genet*, vol. 42, pp. 211–233, 2008.
- [34] B. Audit and C. A. Ouzounis, “From genes to genomes : universal scale-invariant properties of microbial chromosome organisation.”, *J Mol Biol*, vol. 332, pp. 617–633, 2003.
- [35] I. Junier, “Conserved patterns in bacterial genomes : A conundrum physically tailored by evolutionary tinkering.”, *Comput Biol Chem*, 2014.
- [36] T. Schlick, “Modeling superhelical DNA : recent analytical and dynamic approaches”, *Current Opinion in Structural Biology*, vol. 5, pp. 245–262, 1995.
- [37] K. Klenin, H. Merlitz, and J. Langowski, “A Brownian Dynamics Program for the Simulation of Linear and Circular DNA and Other Wormlike Chain Polyelectrolytes”, *Biophys J*, vol. 74, pp. 780–788, 1998.
- [38] A. Vologodskii, “Determining protein-induced DNA bending in force-extension experiments : theoretical analysis.”, *Biophys J*, vol. 96, pp. 3591–3599, 2009.
- [39] P. Carrivain, M. Barbi, and J.-M. Victor, “*In Silico* Single-Molecule Manipulation of DNA with Rigid Body Dynamics”, *PLoS Comput Biol*, vol. 10, p. e1003456, 2014.
- [40] A. V. Vologodskii and N. R. Cozzarelli, “Conformational and Thermodynamic Properties of Supercoiled DNA”, *Annual Review of Biophysics and Biomolecular Structure*, vol. 23, no. 1, pp. 609–643, 1994.

- [41] A. V. Vologodskii and J. F. Marko, "Extension of torsionally stressed DNA by external force.", *Biophys J*, vol. 73, pp. 123–132, 1997.
- [42] L. Postow, C. D. Hardy, J. Arsuaga, and N. R. Cozzarelli, "Topological domain structure of the Escherichia coli chromosome.", *Genes Dev*, vol. 18, pp. 1766–1779, 2004.
- [43] F. Leng, B. Chen, and D. D. Dunlap, "Dividing a supercoiled DNA molecule into two independent topological domains", *Proc Natl Acad Sci USA*, vol. 108, no. 50, pp. 19973–19978, 2011.
- [44] D. G. Priest, L. Cui, S. Kumar, D. D. Dunlap, I. B. Dodd, and K. E. Shearwin, "Quantitation of the DNA tethering effect in long-range DNA looping in vivo and in vitro using the Lac and λ repressors", *Proc Natl Acad Sci USA*, vol. 111, no. 1, pp. 349–354, 2014.
- [45] L. F. Liu and J. C. Wang, "Supercoiling of the DNA template during transcription.", *Proc Natl Acad Sci USA*, vol. 84, pp. 7024–7027, 1987.
- [46] S. Deng, R. A. Stein, and N. P. Higgins, "Organization of supercoil domains and their reorganization by transcription.", *Mol Microbiol*, vol. 57, pp. 1511–1521, 2005.
- [47] B. M. Booker, S. Deng, and N. P. Higgins, "DNA topology of highly transcribed operons in Salmonella enterica serovar Typhimurium", *Molecular Microbiology*, vol. 78, no. 6, pp. 1348–1364, 2010.
- [48] N. P. Higgins, "RNA polymerase : chromosome domain boundary maker and regulator of supercoil density", *Current Opinion in Microbiology*, vol. 22, no. 0, pp. 138 – 143, 2014.
- [49] D. G. Priest, S. Kumar, Y. Yan, D. D. Dunlap, I. B. Dodd, and K. E. Shearwin, "Quantitation of interactions between two DNA loops demonstrates loop domain insulation in E. coli cells.", *Proc Natl Acad Sci USA*, vol. 111, pp. E4449–E4457, 2014.
- [50] J. D. Watson and F. H. C. Crick, "Molecular Structure of Nucleic Acids : A Structure for Deoxyribose Nucleic Acid", *Nature*, vol. 171, pp. 737–738, 1953.
- [51] M. H. F. Wilkins, A. R. Stokes, and H. R. Wilson, "Molecular Structure of Nucleic Acids : Molecular Structure of Deoxypentose Nucleic Acids", *Nature*, vol. 171, pp. 738–740, 1953.
- [52] R. E. Franklin and R. G. Gosling, "Molecular Configuration in Sodium Thymonucleate", *Nature*, vol. 171, pp. 740–741, 1953.
- [53] W. Saenger, *Principle of Nucleic Acid Structure*. Springer, 1988.
- [54] J. F. Allemand, D. Bensimon, R. Lavery, and V. Croquette, "Stretched and overwound DNA forms a Pauling-like structure with exposed bases", *Proc Natl Acad Sci USA*, vol. 95, no. 24, pp. 14152–14157, 1998.
- [55] J. Vinograd, J. Lebowitz, R. Radloff, R. Watson, and P. Laipis, "The twisted circular form of polyoma viral DNA", *Proc Natl Acad Sci USA*, vol. 53, no. 5, pp. 1104–1111, 1965.

- [56] A. Kornberg and T. A. Baker, *DNA Replication*. W. H. Freeman & Company, 2 ed., 1992.
- [57] Y. L. Lyubchenko and L. S. Shlyakhtenko, “Visualization of supercoiled DNA with atomic force microscopy in situ.”, *Proc Natl Acad Sci USA*, vol. 94, pp. 496–501, 1997.
- [58] W. Pohl, “The Self-Linking Number of a Closed Space Curve”, *Indiana Univ. Math. J.*, vol. 17, pp. 975–985, 1968.
- [59] C. C. Adams, *The Knot Book : An Elementary Introduction to the Mathematical Theory of Knots*. W.H. Freeman & Company, 1994.
- [60] J. C. Wang, “DNA topoisomerases.”, *Annu Rev Biochem*, vol. 65, pp. 635–692, 1996.
- [61] J. J. Champoux, “DNA topoisomerases : structure, function, and mechanism.”, *Annu Rev Biochem*, vol. 70, pp. 369–413, 2001.
- [62] R. Strick, M.-N. Dessinges, G. Charvin, N. H. Dekker, J.-F. Allemand, D. Bensimon, and V. Croquette, “Stretching of macromolecules and proteins”, *Reports on Progress in Physics*, vol. 66, pp. 1–45, 2003.
- [63] K. C. Neuman and A. Nagy, “Single-molecule force spectroscopy : optical tweezers, magnetic tweezers and atomic force microscopy”, *Nat Meth*, vol. 5, pp. 491–505, 2008.
- [64] P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J. L. Viovy, D. Chatenay, and F. Caron, “DNA : an extensible molecule.”, *Science*, vol. 271, pp. 792–794, 1996.
- [65] S. B. Smith, Y. Cui, and C. Bustamante, “Overstretching B-DNA : the elastic response of individual double-stranded and single-stranded DNA molecules.”, *Science*, vol. 271, pp. 795–799, 1996.
- [66] R. M. Simmons, J. T. Finer, S. Chu, and J. A. Spudich, “Quantitative measurements of force and displacement using an optical trap.”, *Biophys J*, vol. 70, pp. 1813–1822, 1996.
- [67] T. R. Strick, J.-F. Allemand, D. Bensimon, and V. Croquette, “Behavior of Supercoiled DNA”, *Biophys J*, vol. 74, pp. 2016–2028, 1996.
- [68] C. Gosse and V. Croquette, “Magnetic tweezers : micromanipulation and force measurement at the molecular level.”, *Biophys J*, vol. 82, pp. 3314–3329, 2002.
- [69] M. T. J. van Loenhout, M. V. de Grunt, and C. Dekker, “Dynamics of DNA Supercoils”, *Science*, 2012.
- [70] S. B. Smith, L. Finzi, and C. Bustamante, “Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads.”, *Science*, vol. 258, pp. 1122–1126, 1992.
- [71] S. Forth, C. Deufel, M. Y. Sheinin, B. Daniels, J. P. Sethna, and M. D. Wang, “Abrupt Buckling Transition Observed during the Plectoneme Formation of Individual DNA Molecules”, *Phys. Rev. Lett.*, vol. 100, p. 148301, 2008.

- [72] E. Florin, V. Moy, and H. Gaub, “Adhesion forces between individual ligand-receptor pairs”, *Science*, vol. 264, no. 5157, pp. 415–417, 1994.
- [73] T. Lionnet, J.-F. Allemand, A. Revyakin, T. R. Strick, O. A. Saleh, D. Bensimon, and V. Croquette, “Single-molecule studies using magnetic traps.”, *Cold Spring Harb Protoc*, vol. 2012, pp. 34–49, 2012.
- [74] F. Mosconi, J. F. m. c. Allemand, D. Bensimon, and V. Croquette, “Measurement of the Torque on a Single Stretched and Twisted DNA Using Magnetic Tweezers”, *Phys. Rev. Lett.*, vol. 102, p. 078301, 2009.
- [75] J. Lipfert, J. W. J. Kerssemakers, T. Jager, and N. H. Dekker, “Magnetic torque tweezers : measuring torsional stiffness in DNA and RecA-DNA filaments.”, *Nat Methods*, vol. 7, pp. 977–980, 2010.
- [76] A. La Porta and M. D. Wang, “Optical Torque Wrench : Angular Trapping, Rotation, and Torque Detection of Quartz Microparticles”, *Phys. Rev. Lett.*, vol. 92, p. 190801, 2004.
- [77] P.-G. De Gennes, *Scaling Concepts in Polymer Physics*. Cornell University Press, 1 ed., 1979.
- [78] M. E. Fisher, “Magnetism in One-Dimensional Systems—The Heisenberg Model for Infinite Spin”, *American Journal of Physics*, vol. 32, no. 5, pp. 343–346, 1964.
- [79] T. Odijk, “Stiff Chains and Filaments under Tension”, *Macromolecules*, vol. 28, no. 20, pp. 7016–7018, 1995.
- [80] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block, “Stretching DNA with optical tweezers.”, *Biophys J*, vol. 72, pp. 1335–1346, 1997.
- [81] C. Bouchiat, M. D. Wang, J. Allemand, T. Strick, S. M. Block, and V. Croquette, “Estimating the persistence length of a worm-like chain molecule from force-extension measurements.”, *Biophys J*, vol. 76, pp. 409–413, 1999.
- [82] J. F. Marko and E. D. Siggia, “Statistical mechanics of supercoiled DNA”, *Phys. Rev. E*, vol. 52, pp. 2912–2938, 1995.
- [83] J. D. Moroz and P. Nelson, “Torsional directed walks, entropic elasticity, and DNA twist stiffness”, *Proc Natl Acad Sci USA*, vol. 94, no. 26, pp. 14418–14422, 1997.
- [84] C. Bouchiat and M. Mézard, “Elasticity Model of a Supercoiled DNA Molecule”, *Phys. Rev. Lett.*, vol. 80, pp. 1556–1559, 1998.
- [85] Z. Yang, Z. Haijun, and O.-Y. Zhong-can, “Monte Carlo Implementation of Supercoiled Double-Stranded DNA”, *Biophys J*, vol. 78, pp. 1979–1987, 2000.
- [86] J. A. Gebe, S. A. Allison, J. B. Clendenning, and J. M. Schurr, “Monte Carlo simulations of supercoiling free energies for unknotted and trefoil knotted DNAs.”, *Biophys J*, vol. 68, pp. 619–633, 1995.
- [87] A. V. Vologodskii, S. D. Levene, K. V. Klenin, M. Frank-Kamenetskii, and N. R. Cozzarelli, “Conformational and thermodynamic properties of supercoiled DNA”, *Journal of Molecular Biology*, vol. 227, no. 4, pp. 1224 – 1243, 1992.

- [88] K. V. Klenin, A. V. Vologodskii, V. V. Anshelevich, A. M. Dykhne, and M. D. Frank-Kamenetskii, “Computer simulation of DNA supercoiling”, *Journal of Molecular Biology*, vol. 217, pp. 413–419, 1991.
- [89] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, “Ionic effects on the elasticity of single DNA molecules”, *Proc Natl Acad Sci USA*, vol. 94, no. 12, pp. 6185–6190, 1997.
- [90] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of State Calculations by Fast Computing Machines”, *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.
- [91] K. Klenin and J. Langowski, “Computation of Writhe in Modeling of Supercoiled DNA”, *Inc. Biopoly*, vol. 54, pp. 307–317, 2000.
- [92] F. B. Fuller, “Decomposition of the linking number of a closed ribbon : A problem from molecular biology”, *Proceedings of the National Academy of Sciences*, vol. 75, pp. 3557–3561, 1978.
- [93] S. Neukirch and E. L. Starostin, “Writhe formulas and antipodal points in plectonemic DNA configurations”, *Phys. Rev. E*, vol. 78, p. 041912, 2008.
- [94] L. H. Kauffman, “New Invariants in the Theory of Knots”, *American Mathematical Monthly*, vol. 95, pp. 195–242, 1988.
- [95] Alexander, “Topological invariants of knots and links”, *Trans. Amer. Math. Soc.*, vol. 30, p. 275–306, 1928.
- [96] A. V. Vologodskii, A. V. Lukashin, Kamenetskii, M. D. Frank, and V. V. Anshelevich, “The knot problem in statistical mechanics of polymer chains”, *Soviet Physics JETP*, vol. 39, p. 1059, 1974.
- [97] M. Frank-Kamenetskii and A. Vologodskii, “Topological aspects of the physics of polymers : The theory and its biophysical applications”, *Soviet Physics Uspekhi*, vol. 24, no. 8, p. 679, 1981.
- [98] <http://openmp.org/wp/>.
- [99] http://eigen.tuxfamily.org/index.php?title=Main_Page.
- [100] <http://www.boost.org/>.
- [101] C. Bustamante, J. Marko, E. Siggia, and S. Smith, “Entropic elasticity of lambda-phage DNA”, *Science*, vol. 265, no. 5178, pp. 1599–1600, 1994.
- [102] H. Brutzer, N. Luzzietti, D. Klaue, and R. Seidel, “Energetics at the DNA Supercoiling Transition”, *Biophys J*, vol. 98, p. 1267–1276, 2010.
- [103] C. Maffeo, R. Schöpflin, H. Brutzer, R. Stehr, A. Aksimentiev, G. Wedemann, and R. Seidel, “DNA-DNA interactions in tight supercoils are described by a small effective charge density.”, *Phys Rev Lett*, vol. 105, p. 158101, 2010.
- [104] M. Emanuel, G. Lanzani, and H. Schiessel, “Multiplectoneme phase of double-stranded DNA under torsion”, *Phys. Rev. E*, vol. 88, p. 022706, 2013.

- [105] D. J. Earl and M. W. Deem, “Parallel tempering : Theory, applications, and new perspectives”, *Phys. Chem. Chem. Phys.*, vol. 7, pp. 3910–3916, 2005.
- [106] A. Laio and F. L. Gervasio, “Metadynamics : a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science”, *Reports on Progress in Physics*, vol. 71, no. 12, p. 126601, 2008.
- [107] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragozy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker, “Comprehensive mapping of long-range interactions reveals folding principles of the human genome.”, *Science*, vol. 326, pp. 289–293, 2009.
- [108] E. de Wit and W. de Laat, “A decade of 3C technologies : insights into nuclear organization.”, *Genes Dev*, vol. 26, pp. 11–24, 2012.
- [109] A. Rosa and R. Everaers, “Structure and Dynamics of Interphase Chromosomes”, *PLoS Comput Biol*, vol. 4, p. e1000153, 2008.
- [110] A. Rosa, N. B. Becker, and R. Everaers, “Looping probabilities in model interphase chromosomes.”, *Biophys J*, vol. 98, pp. 2410–2419, 2010.
- [111] L. A. Mirny, “The fractal globule as a model of chromatin architecture in the cell.”, *Chromosome Res*, vol. 19, pp. 37–51, 2011.
- [112] M. Barbieri, M. Chotalia, J. Fraser, L.-M. Lavitas, J. Dostie, A. Pombo, and M. Nicodemi, “Complexity of chromatin folding is captured by the strings and binders switch model.”, *Proc Natl Acad Sci USA*, vol. 109, pp. 16173–16178, 2012.
- [113] F. Benedetti, J. Dorier, Y. Burnier, and A. Stasiak, “Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes.”, *Nucleic Acids Res*, 2013.
- [114] A. Grosberg, Y. Rabin, S. Havlin, and A. Neer, “Crumpled Globule Model of the Three-Dimensional Structure of DNA”, *Europhysics Letters*, vol. 23, no. 5, p. 373, 1993.
- [115] R. Lua, A. L. Borovinskiy, and A. Y. Grosberg, “Fractal and statistical properties of large compact polymers : a computational study”, *Polymer*, vol. 45, no. 2, pp. 717 – 731, 2004.
- [116] V. V. Rybenkov, A. V. Vologodskii, and N. R. Cozzarelli, “The Effect of Ionic Conditions on DNA Helical Repeat, Effective Diameter and Free Energy of Supercoiling”, *Nucleic Acids Research*, vol. 25, no. 7, pp. 1412–1418, 1997.
- [117] G. Charvin, “Études des topoisomérases de type II par micromanipulation d’ADN”. *Thèse de doctorat*, Université Paris 7, 2004.
- [118] F. Képès, B. C. Jester, T. Lepage, N. Rafiei, B. Rosu, and I. Junier, “The layout of a bacterial genome.”, *FEBS Lett*, vol. 586, pp. 2043–2048, 2012.