



HAL
open science

Modélisation d'expériences permettant la manipulation de molécules biologiques uniques

Nicolas Douarche

► **To cite this version:**

Nicolas Douarche. Modélisation d'expériences permettant la manipulation de molécules biologiques uniques. Biophysique [physics.bio-ph]. Université Pierre et Marie Curie - Paris VI, 2005. Français. NNT: . tel-00011190

HAL Id: tel-00011190

<https://theses.hal.science/tel-00011190>

Submitted on 12 Dec 2005

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

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



Thèse de doctorat
de l'université Paris VI
« Pierre et Marie CURIE »

présentée par
Nicolas DOUARCHE

pour obtenir le grade de
docteur de l'université Paris VI

spécialité « physique ».

Modélisation d'expériences
permettant la manipulation
de molécules biologiques uniques

Soutenue le jeudi 8 décembre 2005
devant le jury composé de

Ralph BLOSSEY (Lille)	rapporteur
Martin CASTELNOVO (Lyon)	examinateur
Didier CHATENAY (Paris)	directeur
Simona COCCO (Paris)	co-directrice
François HESLOT (Paris)	président
Jean-François JOANNY (Paris)	rapporteur

Un jour que la maîtresse, à bout de patience parceque le gamin ne savait toujours pas lire, lui dit :

– Mais enfin, Elmo, tu le vois ce tableau ou tu es aveugle ?

– Je suis aveugle, madame la maîtresse.

– Et pourquoi ne l’as-tu pas dit avant ?

Stefano Benni, *Bar 2000*.

Merci !

Avant tout, merci à toi ma thèse : « grâce à ma thèse, j’ai perdu du poids ». Cela ressemble à un slogan publicitaire. J’ai beaucoup apprécié ce petit jeu de poupées russes auquel nous nous sommes livrés toi et moi.

Plus sérieusement je veux remercier les CROUS d’Orléans-Tours, de Versailles et de Toulouse qui m’ont accompagné durant ma formation pré-doctorale.

Ce travail de doctorat a commencé en octobre 2002 à Strasbourg, pour une période de trois années financée par le ministère de l’éducation nationale et de la recherche. La première moitié de mon contrat s’est effectuée au Laboratoire de Dynamique des Fluides Complexes (LDFC) de l’université Louis Pasteur. Il s’est finalisé au Laboratoire de Physique Statistique (LPS) de l’École Normale Supérieure (ENS), à Paris. Merci aux administrations concernées d’avoir réalisé cette transition sans encombres majeures.

Merci à mon directeur de thèse, Didier Chatenay, pour son soutien pratique¹ à cette époque qui ne s’est pas démenti ensuite. Merci pour ta franchise et ton impartialité Didier. Merci à Simona Cocco pour son aide précieuse sur la « partie ADN » et la relecture du manuscrit correspondant.

Je suis reconnaissant aux membres de mon jury d’avoir accepté cette charge de travail, et la responsabilité qui va avec, en un temps extrêmement court.

Concrètement les premiers résultats obtenus ont été numériques. Je suis reconnaissant à Florent Calvo de m’avoir donné goût aux méthodes numériques durant mon stage au Laboratoire de Physique Quantique (LPQ) de l’Institut de Recherche sur les Systèmes Atomiques et Moléculaires Complexes (IRSAMC) de Toulouse. Merci à Fabrice Thalmann, maintenant à l’Institut Charles Sadron (ICS) de Strasbourg, d’avoir partagé ses astuces numériques. L’utilisation de l’algorithme de Lado (chapitre 6) lui est due.

Merci à Philippe Thomen d’avoir clarifié avec moi la « thématique plasmide ».

Merci à Jérôme Robert, parceque... Parcequ’il faut remercier cet homme là. Merci à Alexis Prévost pour son bavardage incessant, à Georges – allô ? – Debrégeas d’avoir simplement souligné une mauvaise direction dans laquelle je m’étais engagé (ah bah non, c’était pas stochastique !), me permettant ainsi de rapidement « corriger le tir ».

J’espère que les gens que j’aime et qui m’aiment (c’est beau !) n’ont pas attendu de voir leur nom apparaître sur ce document pour le savoir $\overleftarrow{\overrightarrow{\cup}}$. Si si, vous êtes sûrement quelque part là-dedans !

¹Bureau, ordinateur, lecture pertinente, encouragements, etc.

Table des matières

Avant propos	1
I Une régulation à l'étape de transcription facilitée par la cyclisation de l'ADN : étude statique	4
1 Initiation de la transcription et cyclisation de l'ADN	5
1.1 Une action à distance fréquemment observée	5
1.2 Facteur de cyclisation : $J(L)$	10
1.3 Une modélisation pertinente selon la longueur de boucle	13
1.4 Pourquoi considérer élasticité <i>et</i> entropie?	14
1.5 Modification de la rigidité élastique.	14
1.6 Récapitulatif : plan de travail	16
2 Le modèle du ver : un polymère semi-flexible	18
2.1 Formalisme : quelques résultats exacts	18
2.2 Limite élastique : approximation de point col	23
2.3 Chaîne entropique : polymère gaussien	28
2.4 Fluctuations au point col : résultat de Shimada et Yamakawa	31
3 Une histoire d'inextensibilité	37
3.1 Un polymère gaussien plus rigide?	38
3.2 Champ moyen : relaxation de la contrainte	39
3.3 Faibles perturbations à l'étirement maximal	41
3.4 Bien d'autres pistes	43
4 Fermeture de la boucle : effets dus à la taille finie des protéines	47
4.1 Relation de « force/extension » : une digression?	47
4.2 Un calcul indirect et numérique du facteur de cyclisation J	50
4.2.1 Principe de l'algorithme	50
4.2.2 Exponentielle de matrice	51
4.2.3 Distributions P et S : discussion et utilisation	54
4.3 Pertinence des effets de tailles selon la longueur de boucle	58
4.4 Taille du répresseur LacR de l'opéron Lac	62

5	Perte de rigidité de l'ADN : cassure effective à mi-longueur	67
5.1	Modélisation : faire le tri	68
5.2	La limite élastique du modèle du ver change	70
5.3	Comment modifier l'algorithme numérique?	76
5.4	Importance de mécanismes entraînant une perte de rigidité	81
6	Contrôle du brin d'ADN par une force extérieure	88
6.1	Apparition d'une direction privilégiée	88
6.2	Étirement de l'ADN par la force	92
7	Vers des modèles de cyclisation de plus en plus complets?	100
7.1	Fermeture de la boucle	100
7.2	Rigidité de torsion	102
7.3	Écrantage et volume exclu	103
7.4	Défauts dans la double hélice et effets de séquence	105
7.5	Étirement par une force	106
7.6	Mais aussi...	107
A	Analogie entre WLC et rotateur rigide quantique	108
A.1	Formulation lagrangienne de la mécanique quantique	108
A.2	Équation de Schrödinger du WLC	110
B	Opérateurs $\widehat{\cos \theta}$ et $\widehat{\cos \varphi \sin \theta}$ développés sur $Y_\ell^m(\Omega)$	113
B.1	Composantes des opérateurs	113
B.2	Rappels sur les harmoniques sphériques	115
C	Configuration de point col	117
D	Relation entre distributions $S(r; L)$ et $P(z; L)$	119
D.1	Dérivation de la relation	119
D.2	Quelques distributions	120
E	Modèle du lignes & cercles : une alternative au point col	123
F	Transformée de Fourier discrète 1D ou 2D	126
F.1	Dimension un : repliement, FFT	126
F.2	Dimension deux : fonction radiale	130
G	Publications	133
II	Régulation à l'étape de réplication : bases d'une étude dynamique	169
8	Expression d'un gène	170

8.1	Retour sur le dogme central	170
8.2	Sensibilité au bruit	172
8.3	Constantes de réaction. Cycle cellulaire	173
8.4	Dynamique stochastique : algorithmes de Gillespie	176
8.5	Résultat : l'« étalon » chromosome	179
9	Contrôle du nombre de copies d'un plasmide	183
9.1	Généralités [SC02, SR01, Weg05]	183
9.2	ColE1 est un plasmide relâché	184
9.3	Modélisation du CNC	186
9.3.1	Schémas hyperbolique/exponentiel	186
9.3.2	Réponse à la concentration en plasmide	189
9.3.3	Constantes de réaction	193
9.4	Distribution conditionnelle	195
10	Perspectives : signal plasmide/chromosome	200
	Bibliographie	217

Table des figures

1	La double hélice d'ADN	2
2	Application d'une torsion additionnelle	3
1.1	Initiation de la réplication : recrutement par CAP, allostérie utilisée par NtrC et facilitée par courbure additionnelle par IHF . . .	6
1.2	Régulation de l'opéron Lac	7
1.3	Boucle $O_1O_3 = 76$ bp impliquée dans la répression de l'opéron Lac	8
1.4	Boucle $O_EO_I = 113$ bp impliquée dans la répression de l'opéron Gal, facilitée par courbure additionnelle par HU	9
1.5	Extraction d'une synapse par le recombinaise Cre. Emmagasiner d'une boucle par l'enzyme de restriction de type I	10
1.6	Facteur de cyclisation $J(L)$	11
2.1	Modèle du ver ou WLC, de l'anglais « Worm Like Chain »	19
2.2	Description de la configuration de point col	24
2.3	Configuration de point col : limite élastique du WLC	26
2.4	Polymère gaussien : limite entropique du WLC	28
2.5	Boucle factice approchant le point col de $r > 0$	33
2.6	Énergie de la boucle factice, et du modèle du lignes & cercles . . .	35
3.1	Flambage d'une fine barre élastique	43
4.1	Relation de force/extension	49
4.2	Distribution $P(z; L)$	55
4.3	Distribution isotrope $S(r; L)$	57
4.4	Distribution $Q(\vec{r}; L)$, rapport $K(r; L)$ et facteur $J(L)$	60
4.5	Une énergie libre de cyclisation	63
5.1	Cassure à mi-longueur du point col	72
5.2	Régression linéaire pour la boucle du point col cassé	73
5.3	Boucle factice cassée à mi-longueur, approchant le point col . . .	75
5.4	Effet de la cassure sur le facteur de cyclisation	82
5.5	Réduction de la longueur la plus probable, et de l'énergie libre de cyclisation des 113 bp impliquées dans la répression de l'opéron Gal	85
5.6	Effet de la cassure sur l'énergie libre de cyclisation	86

6.1	Distribution $P(z; 113 \text{ bp}, f)$ et relation de force/extension	93
6.2	Distribution $S(\rho, z; 113 \text{ bp}, f)$ pour $f = 0$ et 0.2 pN	95
6.3	Distribution $S(\rho, z; 113 \text{ bp}, f)$ pour $f = 0.4$ et 0.6 pN	96
6.4	Distribution $S(\rho, z; 113 \text{ bp}, f)$ pour $f = 0.8$ et 1 pN	97
6.5	Distributions $S(r; 113 \text{ bp}, f)$ contre $P(z; 113 \text{ bp}, f)$	98
6.6	Étirement d'une molécule contenant une boucle	99
7.1	Orientations parallèles et anti-parallèles	101
E.1	Modèle du lignes & cercles	124
F.1	Repliement dans une transformée de Fourier discrète	129
8.1	Dynamique de l'expression au cours de cycles cellulaires : schéma détaillé	179
8.2	Dynamique de l'expression au cours de cycles cellulaires : schéma simplifié	180
8.3	État stationnaire de l'expression génétique	181
8.4	Histogrammes du nombre de protéines	182
9.1	Réplicon du plasmide ColE1	185
9.2	Convergence du cas hyperbolique vers le cas exponentiel	190
9.3	Distribution $R(\beta \alpha)$	197
9.4	Convergence de l'histogramme de R	197
9.5	Convergence du schéma hyperbolique vers le schéma exponentiel	198
9.6	Distribution $P_n(\gamma)$ pour diverses $D(\gamma \beta)$	199

Avant propos

Le travail présenté dans ce manuscrit se situe à l'*interface* entre la physique et la biologie. Cette pluridisciplinarité n'est pas sans imposer quelques contraintes dans sa présentation. J'espère :

- que les physiciens ne trouveront pas rébarbatif le détail des calculs ou de certains rappels. Que cette lecture sera pour eux l'occasion de mettre la main sur quelques références bibliographiques leur permettant d'affronter la vaste littérature de biologie ;
- que les biologistes pardonneront les nombreux raccourcis ou abus de langage. Qu'ils ne tourneront pas les talons devant « tant de calculs » et ne considèreront pas ce travail comme une simple préoccupation de physicien.

Le lecteur peut aborder ce manuscrit avec un point de vue relativement *naïf*. La problématique principale reste celle de relier un *gène* (ce que nous voulons comprendre) à son *expression* (ce que nous mesurons), c'est-à-dire aux protéines dont il gouverne la synthèse. Cette relation peut s'imager à l'aide du schéma simple



connu sous le nom de *dogme central*. Il est constitué de deux étapes :

la transcription est la lecture de l'un des deux brins de la molécule d'ADN² par l'ARN polymérase et la synthèse de l'ARN³, simple brin, complémentaire selon la règle d'*appariement* (liaison hydrogène) « A → U, T → A et C ↔ G » des quatre *bases* Adénine, Thymines (ADN) ou Uracile (ARN), Cytosine et Guanine ;

la traduction est la lecture de la molécule d'ARN (messenger) par le ribosome et la synthèse de la protéine selon la correspondance « codon → acide aminé » *non univoque*, entre chaque triplet de bases de l'ARN et les unités composant le polypeptide formant la protéine.

Des précisions sont apportées au cours du manuscrit, lorsqu'elles s'avèrent nécessaires. L'organisme procaryote étudié est la bactérie *Escherichia Coli*, abrégée en *E Coli*. Dans la première partie nous nous focalisons sur la *transcription*. Nous étudions un processus particulier permettant à cette étape de *réguler* l'expression

²Pour Acide DésoxyriboNucléique.

³Pour Acide RiboNucléique.

génétique : la *cyclisation* de l'ADN, c'est-à-dire la formation d'une *boucle* dans la molécule. L'ARN polymérase, et la synthèse de l'ARN ne jouent dans cette étude qu'un rôle secondaire par rapport à l'ADN, ou d'autres protéines susceptibles de s'y accrocher. Dans la deuxième partie nous incluons la *traduction*, afin d'as-

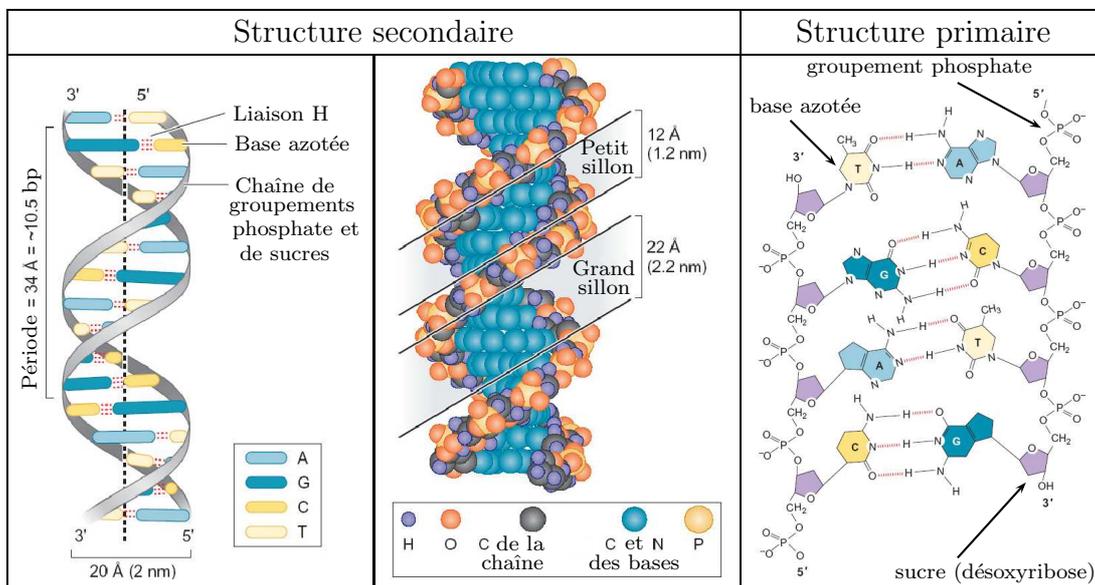


FIG. 1 – Droite : chaque base azotée est liée à un sucre (désoxyribose) formant un nucléoside, qui associé à un groupement phosphate constitue un nucléotide. La liaison phosphodiester assurée par un groupement phosphate entre deux sucres, associe les nucléotides en un brin (structure primaire). Les quatre types de bases s'apparient par liaison hydrogène, selon : $A \leftrightarrow T$ et $C \leftrightarrow G$. Elles constituent deux familles : les purines (A, G) et les pyrimidines (C, T). Gauche : l'appariement et l'empilement des bases donne à la molécule d'ADN double brin une structure stable, en forme de double hélice (structure secondaire). Son enroulement, ou torsion intrinsèque est d'environ 10 paires de bases (bp). Elle est orientée vers la droite (convention : sens positif). La géométrie des paires de bases se traduit par l'existence d'un petit et d'un grand sillon. Les figures sont extraites de [WBB⁺04].

sur une comparaison future avec les expériences de *fluorescence* réalisée dans l'équipe de Didier Chatenay.⁴ C'est l'occasion de discuter brièvement le dogme central. L'ADN conserve toutefois le rôle principal : nous étudions un mécanisme permettant le contrôle du *nombre de copies* du matériel génétique, au cours du *cycle cellulaire*. Cette régulation joue sur :

la réplication de la molécule d'ADN. La lecture de chacun de ses deux brins par l'ADN polymérase permet la synthèse du brin complémentaire, auquel est progressivement associé un deuxième brin. Les appariements entre bases se font selon la règle énoncée ci-dessus ($A \leftrightarrow T$, $C \leftrightarrow G$).

⁴avec Rick, Poirier, Robert et Wong Ng.

Quelques détails sont donc donnés sur la structure de la molécule d'ADN *double brin*, en figure 1. La *double hélice* montrée est la forme B de l'ADN. Elle est idéale dans la mesure où en pratique, l'appariement et l'empilement varient *localement* en fonction du contenu en bases, c'est-à-dire de la séquence. Elle est en fait relativement *flexible* : la majorité de ce travail porte sur une modélisation pertinente de cette propriété. La molécule d'ADN interagit également avec son *milieu* (solvant, protéines, etc.). Elle peut donc être courbée, surenroulée (torsion additionnelle), étirée, etc., produisant ainsi de multiples structures tertiaires. Nous négligeons systématiquement les effets liés à la torsion de l'ADN, ce qui peut se justifier dans un premier temps si les extrémités de la molécule sont libres, et que la torsion n'est donc pas contrainte. Cela est *schématisé* en figure 2 où un ADN linéaire est étiré au moyen d'une force extérieure.

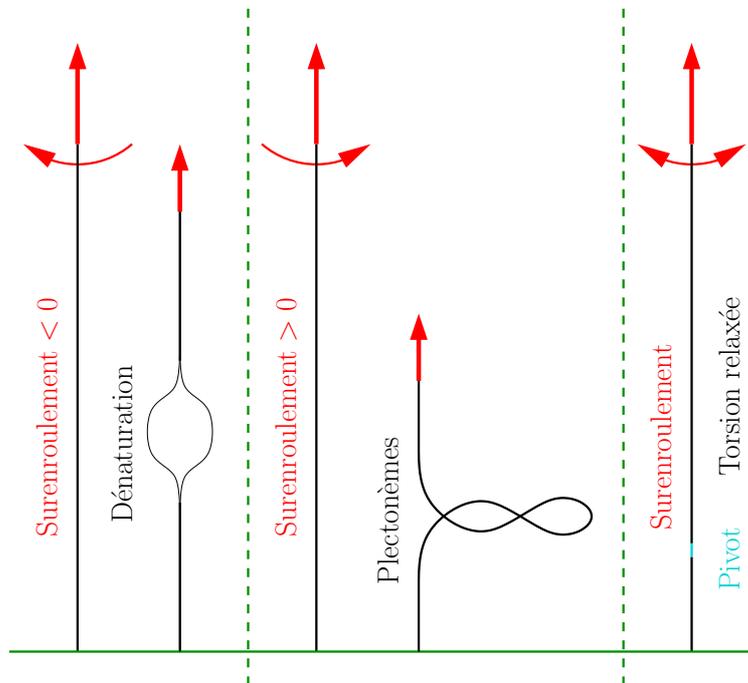


FIG. 2 – Pour un (double) brin d'ADN dont les extrémités sont contraintes, l'application d'une torsion additionnelle (superenroulement ou surenroulement) modifie sa structure. Gauche : surenroulement négatif menant à l'ouverture de la molécule (dénaturation). Milieu : surenroulement positif transformé en pléctonèmes. Droite : l'introduction d'un joint permet de relaxer la torsion additionnelle (ne reste que la torsion intrinsèque).

Première partie

Une régulation à l'étape de
transcription facilitée par la
cyclisation de l'ADN : étude
statique

Chapitre 1

Initiation de la transcription et cyclisation de l'ADN

Dans l'avant propos, le traitement de l'information génétique codée dans la séquence du double brin d'ADN semble être le travail d'*une seule enzyme*, l'ARN polymérase, s'accrochant en *un site* de la molécule pour ensuite la lire. De même pour la traduction de l'ARN messager en protéines, par le ribosome. Cette image est de loin très *naïve*. La bactérie *E Coli* montre effectivement une grande *diversité* de mécanismes pour *réguler* son expression génétique. Les processus élémentaires qui composent ces mécanismes assurent aussi bien une régulation positive (*activation*) que négative (*répression*) de l'expression. Ils sont observés à chaque étape de l'expression en question : de la traduction à la transcription. Au niveau moléculaire ces mécanismes se comprennent comme des interactions entre molécules biologiques. De fait une grande variété de molécules est impliquée. Ce peuvent être des brins d'ADN ou d'ARNs, ou bien encore diverses protéines.¹ La nature de leurs interactions est également très variée. Il peut s'agir de complexations (dimérisation, tétramérisation, etc.) plus ou moins stables, d'hybridation. Ces interactions sont souvent *coopératives* et produisent des *structures complexes* observables expérimentalement, comme des boucles parmi le double brin d'ADN, différents motifs de repliements ou d'appariements de simples brins d'ARNs comme des boucles en épingle à cheveux (« hairpin » en anglais).

1.1 Une action à distance fréquemment observée

En quoi la cyclisation du double brin d'ADN, c'est-à-dire la formation d'une boucle dans sa séquence est-elle impliquée dans la régulation de l'expression génétique ? Il existe beaucoup d'exemples pour lesquels la transcription d'un gène

¹L'idée de gènes dont l'expression a pour but la régulation de celle d'autres gènes a été en autres avancée par Jacob et Monod, au cours de travaux sur l'opéron Lac réalisés durant les années 50 et 60.

dépend de l'interaction entre des protéines qui ont la capacité de reconnaître des séquences bien spécifiques (spécificité en site) le long de l'ADN, et de s'y accrocher. En rapprochant ces protéines la formation d'une boucle dans l'ADN *facilite* leur interaction nécessaire. Nous donnons quelques exemples dans cette section. Pour initier la transcription d'un *opéron* (groupe de gènes), l'ARN polymérase doit avant tout s'accrocher à son *promoteur*. Selon l'opéron, le promoteur réalise un complexe plus ou moins stable avec la polymérase. Un exemple de promoteur *faible* est celui de l'opéron Lac (voir plus bas). Le *recrutement* de la polymérase par la protéine CAP – pour « Catabolite Activator Protein » – suivit de la cyclisation de la séquence entre le promoteur et le site amplificateur auquel CAP est accrochée, facilite l'initiation de la transcription [WBB⁺04]. Cet exemple d'*accrochage coopératif* est schématisé en figure 1.1 (gauche).

Le métabolisme de nitrogène dépend d'un gène dont le promoteur forme un

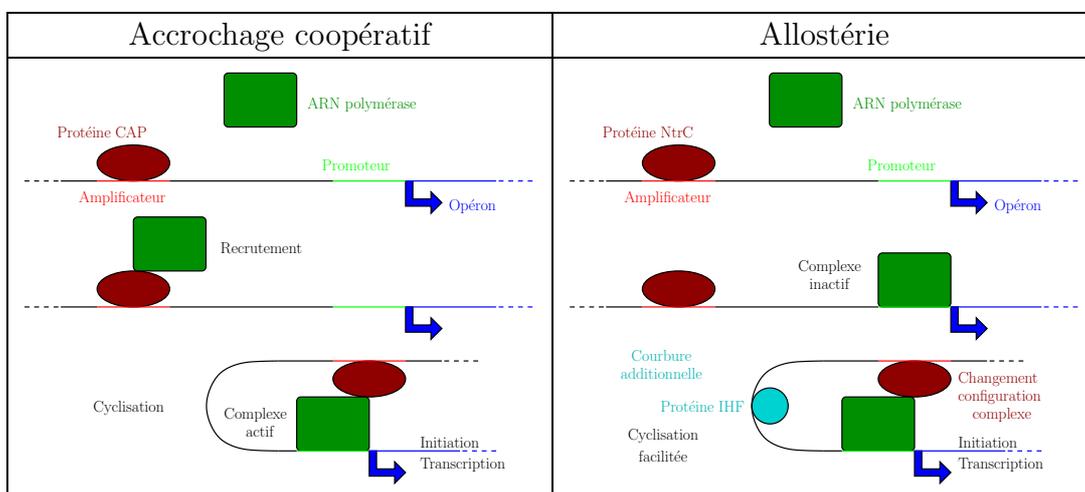


FIG. 1.1 – Activation de l'initiation de la transcription par cyclisation. Gauche : la protéine CAP reconnaît une séquence spécifique, amplificatrice du promoteur de l'opéron. L'ARN polymérase est recrutée par accrochage coopératif avec CAP. La formation d'une boucle permet à la polymérase de complexer avec le promoteur et d'initier la transcription. Droite : la polymérase arrive à complexer, mais la transcription n'est pas initiée. La protéine NtrC permet de l'activer, en changeant la configuration du complexe. Cette régulation allostérique est possible par cyclisation, qui peut être facilitée par la courbure additionnelle locale, provoquée par l'accrochage de la protéine IHF.

complexe stable avec la polymérase, mais incapable d'initier la transcription. Elle peut être déclenchée par la protéine NtrC – pour « Nitrogen regulatory protein C » – capable de *modifier* favorablement la configuration du complexe promoteur/polymérase. Comme NtrC reconnaît également un site amplificateur du promoteur, la cyclisation de la séquence entre ces derniers rend plus probable la modification du complexe [WBB⁺04, SLR00, RGvHB97, WRN⁺97, SPKE90].

Cet exemple de régulation par *allostérie*² est schématisé en figure 1.1 (droite). La séquence est de 150 bp, mais elle peut être artificiellement augmentée jusqu'à un facteur dix, c'est-à-dire un ordre de grandeur. Dans cette séquence se trouve le site d'accrochage de la protéine IHF – pour « Integrating Host Factor ». Cet accrochage courbe *localement* la séquence, facilitant ainsi la cyclisation [JAAB⁺01, RYMN96].

L'opéron Lac (figure 1.2) permet d'exprimer les protéines réalisant le métabolisme

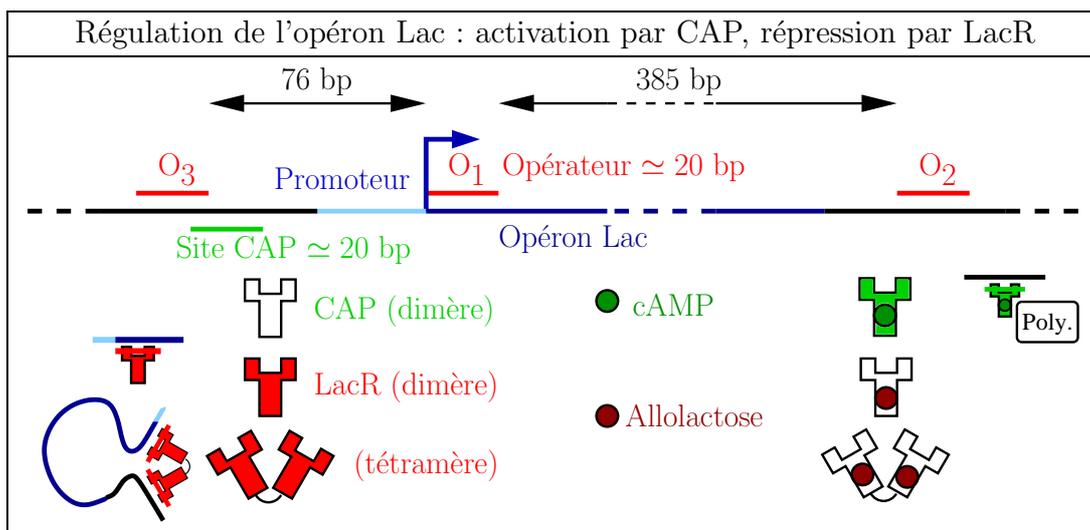


FIG. 1.2 – L'opéron Lac est composé des gènes exprimant les protéines nécessaires au métabolisme du sucre lactose. L'activation de sa transcription est assurée par recrutement de l'ARN polymérase, par la protéine CAP. C'est complexée avec la molécule cAMP que ce dimère s'accroche à une séquence spécifique, en amont du promoteur de l'opéron. La concentration en cAMP du milieu est diminuée en présence du sucre glucose. D'autre part, la répression dépend de la protéine LacR empêchant l'accrochage de la polymérase, ou sa lecture de l'opéron. En l'absence de lactose, ses dimères s'accrochent aux opérateurs $O_{1,2,3}$. Sous forme de tétramère, le LacR peut alors cycliser les séquences entre O_1O_3 (figure 1.3) et O_1O_2 . C'est un produit du métabolisme du lactose (l'allolactose), qui empêche ces accrochages par ajustement dans chaque dimère.

du lactose, au cours duquel ce sucre est coupé en galactose et en glucose. Ces deux sucres sont utilisés par *E Coli* comme source d'énergie, avec une préférence pour le glucose. L'opéron Lac est réprimé en l'absence de lactose, qu'il y ait ou non du glucose. Il n'est au contraire plus réprimé en présence de lactose, et se trouve même activé en l'absence de glucose [WBB⁺04, Mat92].³ L'activation se fait par recrutement de la polymérase par CAP.⁴ Le site amplificateur étant situé 60 bp

²La fonction de la molécule est rendue possible, ou au contraire impossible, par changement de configuration, provoqué par une autre molécule.

³L'absence d'activation rend peu probable l'expression de l'opéron (promoteur faible).

⁴Cette protéine intervient en fait dans la régulation de plus d'une centaine de gènes.

en amont de l'opéron, il n'y a pas de cyclisation⁵ au contraire du précédent exemple. Le signal contrôlant l'accrochage de CAP est la concentration en glucose du milieu : c'est l'ajustement de la molécule cAMP – pour « Cyclic Adénosine MonoPhosphate » – dans le dimère qui permet l'accrochage⁶, et la concentration de cette dernière diminue en présence de glucose. La répression consiste à empêcher le complexe entre la polymérase et le promoteur, ou encore la lecture de l'opéron par cette dernière. Les dimères du répresseur LacR reconnaissent trois courtes séquences spécifiques $O_{1,2,3}$. Il peut s'accrocher individuellement en chacun des ces trois opérateurs, mais c'est sous forme de *tétramère* que le LacR est le plus efficace : il est alors capable de former une *boucle* entre O_1O_3 (76 bp)⁷ ou O_1O_2 (385 bp) [BMS99, BMS03]. C'est un produit du métabolisme du lactose, l'allolactose, qui rend impossible ces accrochages en changeant la configuration des dimères du LacR [LCH⁺96].⁸ Dans la suite nous nous focalisons sur la *répression*, facilitée par *cyclisation* (figure 1.3).

Le métabolisme du sucre galactose dépend de l'expression de l'opéron Gal, dont

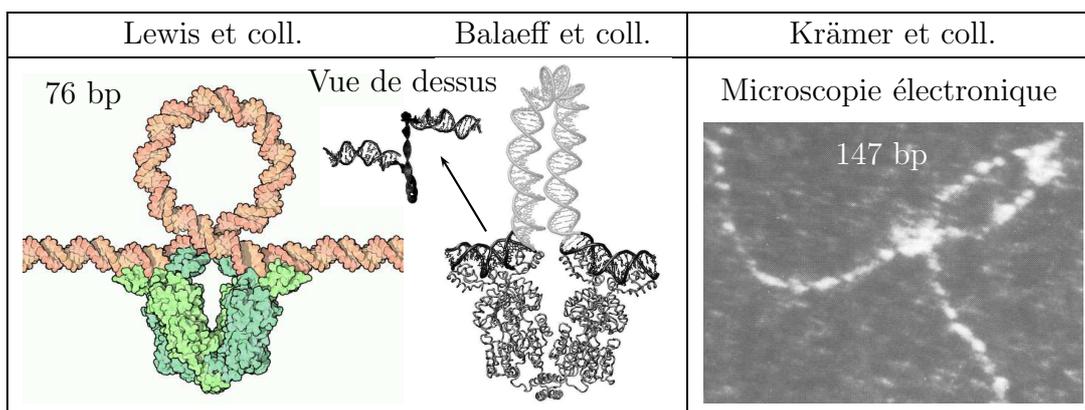


FIG. 1.3 – Petite boucle intervenant dans la répression de l'opéron Lac. Gauche : illustration basée sur la structure cristallographique du LacR obtenue en 1996 par Lewis, Chang, Horton, Kercher, Pace, Schumacher, Brennan et Lu [LCH⁺96] comparée à une structure proposée en 2003 par Balaëff, Mahadevan et Schulten [BMS99, BMS03] en résolvant numériquement le problème d'élasticité correspondant. Droite : visualisation par microscopie électronique d'une boucle de 147 bp fixée par un LacR, réalisée en 1987 par Krämer, Niemöller, Amouyal, Revet, von Wilcken-Bergmann et Müller-Hill [KNA⁺87].

la régulation est très semblable à celle du l'opéron Lac. Il est réprimé en l'ab-

⁵L'accrochage de CAP courbe toutefois localement la séquence [KEL⁺01, KC98, BMS04].

⁶CAP est aussi appelée CRP – pour « CAMP Receptor Protein ».

⁷En présence de glucose, cette petite boucle peut être gênée par CAP dont le site jouxte O_3 (figure 1.2). L'accrochage coopératif dont font en fait preuve CAP et le LacR peut exiger une réinterprétation du rôle de CAP comme activateur de la transcription seule [BMS04].

⁸En l'absence de lactose l'opéron est rarement mais tout de même exprimé. Cela permet l'initiation du métabolisme, autrement paradoxale.

sence de galactose. En sa présence, il est en revanche exprimé voir même activé si le milieu contient du glucose [WBB⁺04, Mat92]. L'activateur est encore CAP. Les deux opérateurs O_E et O_I , distants d'une séquence de 113 bp comprenant les deux promoteurs de l'opéron, permettent l'accrochage du répresseur GalR en l'absence de galactose (contrôle allostérique). Le GalR est différent du LacR : les dimères interagissent trop *faiblement* pour que le tétramère forme une boucle entre $O_E O_I$ [MH98, GA02]. Il a été démontré *in vitro* [LBC⁺03] que l'ajout de la protéine HU – pour « Heat Unstable » – permettait de retrouver une répression *efficace*. Cette protéine, présente dans *E Coli* en un grand nombre de copies⁹, est connue pour s'accrocher à la séquence d'ADN de manière non spécifique [vNVG⁺04]. Parmi les scénarii imaginés d'interactions entre le GalR, HU et le brin d'ADN, [SVA05, SGLA02, GVZA01, VLK⁺03, ACA96, LSAA97, KA01, AA97, LGA99] il est maintenant admis le suivant : HU s'accroche sur la séquence¹⁰ *entre* les deux opérateurs, et cette courbure locale *additionnelle* rend l'interaction des dimères *suffisante* pour la cyclisation. Une illustration est disponible en figure 1.4

En fait, la formation de boucle dans l'ADN est un mécanisme fréquemment

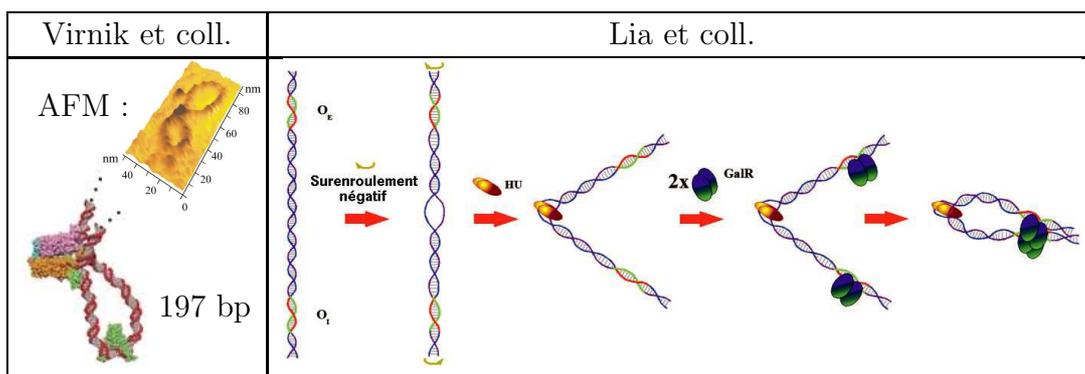


FIG. 1.4 – Boucle intervenant dans la répression de l'opéron Gal. Gauche : image AFM (Microscopie à Force Atomique) obtenue en 2003 par Virnik, Lyubchenko, Karymov, Dahlgren, Tolstorukov, Semsey, Zhurkin et Adhya et son illustration basée sur les structures cristallographiques du GalR et de HU [VLK⁺03]. La séquence cyclisée de 197 bp fait partie d'un ADN circulaire ($\simeq 600$ bp). Droite : schéma expliquant comment HU, en courbant localement la séquence de 113 bp séparant les opérateurs O_E et O_I , peut permettre au tétramère GalR instable de fixer la boucle. Tiré d'un travail [LBC⁺03] publié en 2003 par Lia, Bensimon, Croquette, Allemand, Dunlap, Lewis, Adhya et Finzi.

observé pour contrôler l'expression génétique, en empêchant l'initiation de la transcription [WBB⁺04, Mat92]. Nous pourrions encore citer les opérons Ara-

⁹entre 15000 et 60000 selon l'instant observé dans le cycle cellulaire.

¹⁰HU s'ajuste dans une bulle de dénaturation dont l'ouverture est facilitée par un surenroulement négatif de la séquence.

BAD [HMS88] et Deo ou encore la *très* élaborée commutation du bactériophage¹¹ λ entre les états lysogène (prophage, phage dormant) et lytique (phage actif). Mentionnons également la recombinaise Cre – pour « CAMP Responsive Element » – qui assure la présence d’une unique copie du génome du prophage P1, dans le chromosome de la bactérie infectée, par reconnaissance du site LoxP – pour « Locus of crossover P1 ». L’ADN du bactériophage ne contenant qu’une copie de ce site, si cette protéine accroche deux LoxP elle cyclise (la boucle est appelée synapse) puis *extrait* du chromosome la séquence comprise entre les deux sites¹² (figure 1.5, gauche). Le processus peut être inversé faisant de ce système un *outil* remarquable pour la modification de séquences d’ADN [Nag00]. Finissons avec les enzymes de restriction de type I. Ces moteurs moléculaires s’accrochent en un site spécifique et peuvent, en utilisant de l’ATP, graduellement *emmagasiner* une boucle à partir de la molécule d’ADN (figure 1.5, droite). Les facteurs qui déclenchent l’initiation comme la terminaison (coupure de la boucle) du processus restent incompris [DMR01, Mur00]. Notons que les changements de configurations des répresseurs des opérons Lac et Gal ne requiert au contraire *aucun* apport énergétique.

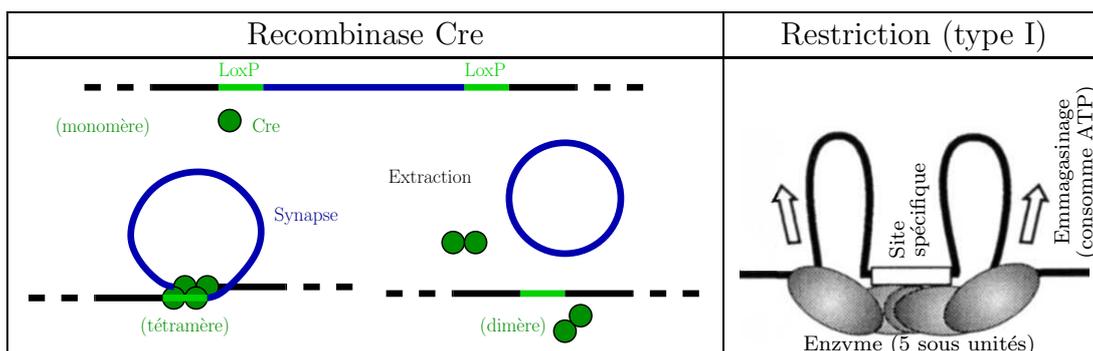


FIG. 1.5 – Gauche : extraction de la séquence entre deux LoxP, en une boucle (synapse), réalisée par la recombinaise Cre sous forme de tétramère. Droite : l’enzyme de restriction de type I peut emmagasiner une boucle à partir d’une molécule d’ADN. Le processus nécessite un apport énergétique. L’illustration est tirée de [Mur00]. Dans les deux cas, la longueur de la séquence est très variable.

1.2 Facteur de cyclisation : $J(L)$

La cyclisation de l’ADN permet notamment de rapprocher deux sites spécifiques *a priori* séparés par une séquence ne permettant pas leur interaction directe. Pour fixer les idées nous raisonnons sur la répression de l’opéron Lac, au moyen de la boucle $O_1O_3 = 76$ bp. Un indicateur pertinent de l’efficacité

¹¹Virus n’infectant que les bactéries.

¹²Ainsi qu’une moitié de chaque LoxP afin de n’en laisser qu’une copie.

de ce phénomène consiste par exemple à mesurer la *concentration* moyenne de l'opérateur O_3 à *proximité* du O_1 . Par « proximité » nous entendons :

- au mieux le volume permettant à la boucle de se former, prenant ainsi en compte les *propriétés* de la protéine qui la fixe : le LacR. On parle parfois de *volume de réaction* ;
- au pire un *petit* volume, choisit arbitrairement.

Supposons maintenant que nous puissions faire varier la longueur de la séquence séparant les deux opérateurs. Sans changer le volume (les espèces en jeu n'ont pas changées), nous obtenons alors une collection de concentrations en fonction de la longueur. C'est une caractéristique de la capacité du LacR (et des opérateurs) à former une boucle d'une longueur quelconque dans la molécule d'ADN [RvHL95, Rip01]. Si nous considérons la situation « idéale » où la boucle se forme quand les extrémités de sa séquence *coïncident* exactement, Le volume pertinent est alors *infinitésimal* : les concentrations deviennent ce que l'on nomme le *facteur de cyclisation*¹³, écrit par convention $J(L)$ comme fonction de la longueur L de la boucle [HS62, Yam97]. Un graphe *typique* de $J(L)$ est proposé en figure 1.6. Nous y situons certaines des boucles évoquées en section précédente. Nous remettons à plus tard une description et une discussion du profil de $J(L)$. Dans la fin de cette section nous soulignons deux aspects importants.

Tout d'abord, nous avons conclu l'avant propos de ce manuscrit sur le caractère

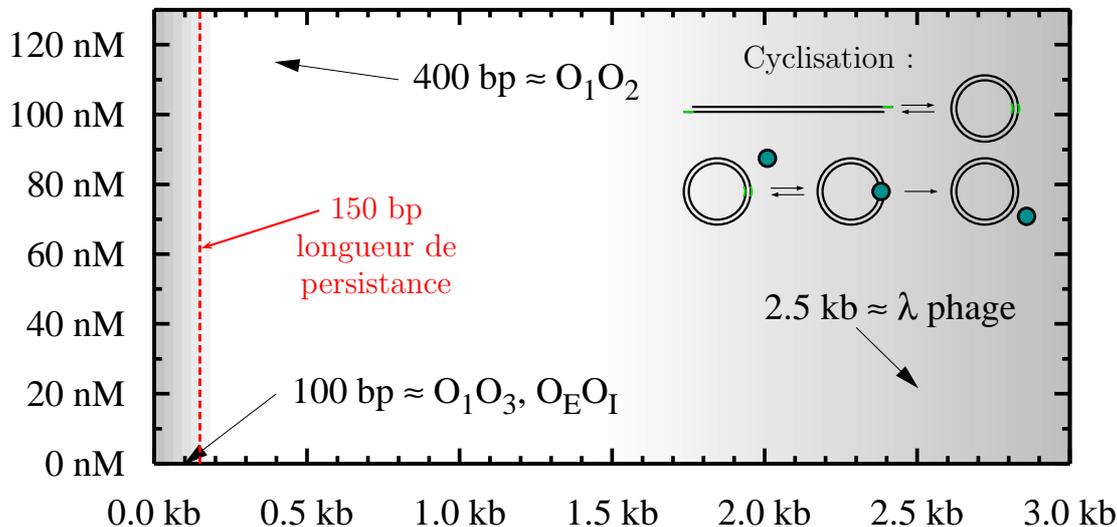


FIG. 1.6 – La concentration d'une extrémité de la boucle au voisinage direct de la seconde s'appelle le facteur de cyclisation, noté $J(L)$. C'est la distribution statistique en longueurs de boucle. Médaillon : expérience type de cyclisation d'un fragment d'ADN, par hybridation irréversible des extrémités par une enzyme.

dynamique de l'ADN au cours du traitement de son information. La formation d'une boucle est une belle illustration de la capacité de la molécule à être *courbée*. La figure 1.6 montre que ce phénomène est peu probable pour des séquences

¹³Où plus brièvement le facteur- J .

contenant moins de 150 bp (pointillés rouges). Cette longueur *caractéristique* de sa rigidité élastique s'appelle la *longueur de persistance*. Exprimée dans des unités plus conventionnelles en physique, elle mesure environ 50 nm [LWS02]. Nous proposerons une définition rigoureuse plus tard. Quand au facteur de cyclisation, c'est en fait la *distribution* en longueur de boucles de la molécule d'ADN. Pour revenir à des unités là aussi plus habituelles :

$$\begin{aligned} (1 \text{ mole de sites}) / (1 \text{ L}) &= (6.02252 \times 10^{23} \text{ sites}) / \left[(10^8)^3 \text{ nm}^3 \right] \\ &= \frac{6.02252}{10} \times (1 \text{ site}) / (1 \text{ nm}^3), \end{aligned}$$

autrement dit $1 \text{ nm}^{-3} \simeq 1.66 \text{ mol} \cdot \text{L}^{-1} \equiv 1.66 \text{ M}$, où M signifie « molaire ». Nous remarquons que c'est une quantité très *faible* : nous l'avons exprimée en nanomolaires dans la figure 1.6.

Enfin nous avons implicitement adopté depuis le début une *convention* peu usuelle : nous ne faisons pas de différence entre les termes de *cyclisation* et de *formation d'une boucle*. Il est parfaitement possible de reproduire *in vitro* les boucles fixées par des protéines, comme le LacR [FG95, LBC⁺03, BPJ91]. Une autre expérience typique consiste à provoquer l'hybridation des extrémités de fragments d'ADN de diverses longueurs. Des enzymes, comme les *ligases* permettent une telle hybridation de manière irréversible [DVK⁺05, DSS⁺05]. Bien qu'il s'agisse toujours de la formation d'une boucle, on parle dans ce cas de cyclisation. Ce sont Shore, Langowski et Baldwin qui ont les premiers (1981) réalisés de telles expériences [SLB81, SB83].¹⁴ Il s'agit en effet d'un type *particulier* de formation de boucle : l'enzyme requiert une coïncidence exacte des extrémités (volume *nul*) et l'hybridation laisse un ADN fermé, c'est-à-dire cyclique (conditions limites *périodiques*). Ce scénario, schématisé en figure 1.6 (médaillon) n'est *a priori* pas pertinent pour la boucle O_1O_3 fixée par le LacR (figure 1.3).¹⁵ La formation d'une boucle apparaît *sensible* à ces deux propriétés. Nous n'en disons pas plus pour le moment : leur étude constitue une grande partie de ce travail. La convention que nous adoptons se retrouve par exemple en 2001 dans un article de Klenin et Langowski [KL04]. La circularité, que la cyclisation de courts fragments impose, a également été remise en cause en 2005 suite aux (certes très controversées) expériences de Cloutier et Widom [CW04, CW05].

¹⁴Depuis répétées un grand nombre de fois.

¹⁵Le $J(L)$ de la figure 1.6 correspond en fait à des conditions limites *quelconques*. Ce n'est pas un résultat expérimental, mais semi-analytique : il s'agit de la formule (3.1) de $p = 0$.

1.3 Une modélisation pertinente selon la longueur de boucle

Comme l'illustrent les deux sections précédentes, ces boucles d'ADN s'observent *in vivo* pour une *large échelle* de longueurs. Cette dernière couvre au moins deux ordres de grandeur : d'une centaine à quelques milliers de paires de bases. Pour des séquences significativement inférieures à la longueur de persistance de l'ADN (disons jusqu'à une demi, soit 75 bp), modéliser la boucle est un problème classique de courbure d'une *barre élastique*. Des résultats analytiques exacts tels que son énergie élastique de courbure ou l'équation décrivant sa configuration peuvent alors être obtenus [Yam97, SY84, YS72, IB79, TW86] à condition toutefois d'omettre des propriétés de l'ADN potentiellement pertinentes, comme sa nature électrostatique par exemple. De tels effets peuvent néanmoins être étudiés par résolution numérique du problème [MRKL98, KML98, BMS99, BMS03, ZRG02, ZRG03a, ZRG03b]. En revanche dès que la boucle inclut plus de dix longueurs de persistance (1500 bp, ou 1.5 kb), l'ADN se comporte comme un polymère gaussien [Yam97] : la *perte entropique* associée à sa cyclisation est la contribution principale à son coût énergétique [HM03].

Pour des longueurs de boucles *intermédiaires* – région non grisée de la figure 1.6 – les effets élastique et entropique sont *tous deux* importants et une étude basée sur des méthodes de *physique statistique* est alors judicieuse. C'est notamment le cas des boucles dont la longueur est la plus probable (environ 500 bp, figure 1.6). Disposer de résultats quantitatifs ne serait-ce que pour ce cas particulier permettrait une meilleure compréhension du mécanisme de régulation initial, et ce malgré sa complexité globale.

Un modèle simple de polymère continu incluant *à la fois* des effets élastique et entropique est le *modèle du ver* [KP49, Yam97, Kle04], décrit au chapitre suivant. Une quantité d'intérêt tant théorique qu'expérimental est le *facteur de cyclisation* (§ 1.2) c'est-à-dire la *distribution statistique* en longueurs de boucle $J(L)$, traduisant donc la capacité de former une boucle selon sa longueur. Malgré une étude théorique assez intense depuis la formulation initiale du modèle du ver une cinquantaine d'années auparavant, on ne dispose encore d'aucune expression *analytique* exacte du facteur de cyclisation. Beaucoup de méthodes d'*approximations* ont en revanche été proposées. Ce sont des développements qui partant d'un des deux régimes limites (élastique [WF96, SY84] ou entropique [Dan52, GYSM72]) essayent d'inclure au mieux les effets ainsi initialement négligés (entropiques ou élastiques respectivement). Un rapide bilan de ces approximations est proposé au chapitre 3. Toutefois ajoutées à un nombre non négligeable de résultats *numériques* [PV00, RvHL95, MRKL98, YKM05] et *expérimentaux* [DSS⁺05, BPJ91, SLB81, SB83], les principales propriétés du facteur de cyclisation sont alors *globalement* bien comprises.

1.4 Pourquoi considérer élasticité *et* entropie ?

Si les comportements « purement » élastique et entropique de l'ADN sont individuellement bien caractérisés (résultats analytiques exacts) et qu'une série d'approximations permet une description qualitative de la cyclisation des longueurs auxquelles ces deux effets sont importants – description pouvant par ailleurs être quantitativement vérifiée grâce aux données numériques et expérimentales – alors *pourquoi* tenter une analyse théorique exacte du modèle du ver ? Sans rapport direct avec la modélisation de systèmes biologiques il existe tout d'abord un intérêt physique *fondamental* : le modèle du ver est en fait un cas particulier de *modèle sigma non linéaire* décrivant un système de spins unitaires, à trois dimensions (*modèle d'Heisenberg classique*). Malgré son étude également intensive ces dernières décennies, une solution exacte n'a été proposée qu'en dimension infinie [Sir02]. Il s'agit effectivement d'un problème théorique complexe, toutefois nos exigences sont moindre en ce qui concerne la cyclisation de l'ADN. En effet une approximation incluant une description quantitative du régime pour lequel effets élastique et entropique sont importants, simplifierait déjà grandement l'étude théorique, par exemple de la dynamique de cyclisation. Un tel résultat assurerait par conséquent des progrès certains dans la modélisation globale du mécanisme de régulation à l'étape de transcription.¹⁶ Même en tenant compte de ce qui vient d'être dit, la cyclisation *seule* du double brin d'ADN « nu » montre en effet rapidement ses *limites* face aux situations biologiquement observables [Mat92]. Par exemple, nous avons déjà mentionné l'observation significative *in vivo* de boucles d'ADN de longueurs très diverses. Ce résultat est au premier abord *contre intuitif* : pourquoi une telle *dispersion* plutôt qu'un choix quasi-systématique de la longueur la plus probable ? Quel intérêt pour « la nature » d'utiliser apparemment sans distinction des processus de coûts énergétiques aussi divers, de natures aussi bien élastique (courbure de la séquence cyclisée) qu'entropique (coïncidence des extrémités de la boucle) ? En plus du besoin déjà évoqué d'une modélisation unitaire et pertinente quelque soit la longueur de boucle (formalisée au chapitre 2) ces quelques faits justifient de *préciser* le processus de cyclisation de l'ADN lui-même.

1.5 Modification de la rigidité élastique.

Tout d'abord il est utile de garder à l'esprit qu'une molécule d'ADN n'est jamais « nue » *in vivo*. Nous entendons par là qu'un grand nombre de molécules diverses sont *accrochées* le long d'un brin d'ADN, et ce à tout moment du cycle

¹⁶La distribution en longueur de boucles serait de plus utilisables pour quantité d'autres matériaux [TC02, Zho01], non nécessairement biologiques [SKOS01], dont on connaît ou dont on veut connaître la rigidité élastique.

cellulaire d'*E. Coli*. Ce peuvent être des brins d'ARNs, des enzymes comme les polymérasés ou bien encore les protéines régulant l'initiation de la transcription que nous étudions.

À leur propos nous avons pour le moment été discret. D'avantage d'attention a été accordée à la cyclisation de l'ADN, phénomène dont le but est pourtant bien de faciliter l'interaction de telles protéines. Une fois la boucle formée ce sont elles qui en assurent alors la fermeture, dont la stabilité dépend de leur interaction. Considérant la taille *finie* de ces protéines la « boucle » n'est donc pas à proprement parler fermée. Intuitivement cette simplification est sans incidence majeure pour des séquences grandes devant la taille des protéines. C'est par exemple certainement le cas des séquences suffisamment longues pour que l'ADN ait un comportement entropique.¹⁷ En revanche dans le cas limite opposé où longueur des séquences et taille des protéines sont du même ordre de grandeur on ne peut décemment plus parler de boucle et on doit s'attendre à des changements importants du processus de cyclisation, concernant son énergétique par exemple. Qu'en est-il pour des longueurs de séquences de plus en plus grandes devant la taille des protéines ? Ces considérations motivent l'étude quantitative de ces *effets de tailles*, présentée au chapitre 4.

Plus généralement l'ensemble des molécules accrochées courbent *localement* le brin d'ADN. mêmes faibles individuellement la somme de ces courbures peut être suspectée de modifier la rigidité élastique de l'ADN, et donc sa cyclisation. Plus spécifiquement rappelons que la régulation de l'opéron Gal fait intervenir une protéine type histone (HU) pour plier la séquence d'ADN, trop rigide sinon pour que le répresseur maintienne stablement la boucle (figure 1.4). Toujours en section 1.1 nous avons mentionné une fonction semblable de la protéine IHF (figure 1.1, droite). Une rigidification peut aussi être observée : c'est le cas de la protéine RecA – pour « Recombinase A » – qui entoure la double hélice [LRB⁺98, HSB99], ou encore d'une saturation de protéine HU, comme démontré récemment [vNVG⁺04]. De manière plus générale l'existence de mécanismes coopératifs entraînant une *modification* de la rigidité élastique de l'ADN apparaît comme primordiale pour la régulation de l'expression génétique, reposant sur une cyclisation du brin d'ADN. Le chapitre 5 précise brièvement ces mécanismes puis explore la possibilité de les reproduire de manière *effective*, c'est-à-dire sans s'arrêter sur leurs *détails* respectifs.

Sans prétendre être exhaustif, nous verrons que ces deux types d'effets permettent une amélioration sensible des modèles de boucles d'ADN, dans le cadre d'une régulation de l'expression génétique à l'étape de transcription.

¹⁷La taille typique des protéines que nous étudions est d'une dizaine de nanomètres.

1.6 Récapitulatif : plan de travail

Le prochain chapitre présente le modèle du ver (§ 2.1), incluant les effets élastique et entropique nécessaires à une description pertinente et unitaire de la cyclisation de l'ADN, selon la longueur de boucle. Les limites élastique et entropique sont ensuite individuellement étudiées (§§ 2.2 et 2.3). Nous proposons pour finir (§ 2.4) une adaptation aux boucles non rigoureusement fermées, du résultat empirique obtenu en 1984 par Shimada et Yamakawa [SY84, Yam97] étendant la validité de la limite élastique.

Le chapitre 3 montre pourquoi un calcul exact des quantités associées au facteur de cyclisation est compliqué. Suit un rapide bilan des approximations proposées au cours des cinquante dernières années (§§ 3.1–3.4).

Le chapitre 4 traite des effets de taille des protéines fermant la boucle d'ADN. Bien qu'il ne soit pas encore question de soumettre la molécule d'ADN à une force extérieure, une étude numérique de la relation de « force/extension » du modèle du ver est reprise (§ 4.1). Il existe en effet une forte analogie entre ce problème et l'algorithme que nous proposons pour calculer le facteur de cyclisation (§ 4.2). Ses résultats sont dans un premier temps comparés à ceux des régimes purement élastique et entropique ainsi qu'aux approximations les plus fréquemment utilisées et pertinentes (§ 4.3). Les effets sur la cyclisation de l'ADN, dus à la taille des protéines sont enfin estimés et illustrés dans le cas de la répression de l'opéron Lac (§ 4.4).

Le chapitre 5 s'intéresse aux mécanismes provoquant une perte de rigidité de l'ADN (§ 5.1). Une estimation effective de leur rôle est tentée en cassant arbitrairement le modèle du ver à mi-longueur. Après reconsidération de sa limite élastique (§ 5.2) nous montrons également comment adapter l'algorithme numérique (§ 5.3). Ces résultats, comparés aux données expérimentales (répression de l'opéron Gal) et numériques disponibles, mettent en évidence l'importance de tels mécanismes et permettent de discuter la pertinence d'une description effective (§ 5.4).

Le chapitre 6 considère le cas d'une molécule d'ADN étirée par une force uniforme. Cette étude est motivée par les dispositifs expérimentaux développés ces dernières années et permettant la manipulation de molécules uniques [BMW00, SDC⁺03]. Pour contrôler le brin d'ADN une telle force extérieure est en effet fréquemment utilisée. L'apparition d'une direction privilégiée (brisure de symétrie) modifie peu le principe du calcul du facteur de cyclisation développé précédemment mais elle le rend en revanche moins direct, nécessitant quelques développements numériques supplémentaires (§ 6.1). Nous ne produisons pas une étude complète de la force sur la cyclisation de l'ADN, mais nous vérifions en revanche que la méthode proposée reproduit convenablement l'étirement d'un fragment de longueur fixé (§ 6.2).

Le chapitre 7 esquisse les moyens d'inclure des effets initialement négligés mais qui peuvent jouer un rôle significatif dans la cyclisation de l'ADN comme le contenu (en bases) de sa séquence, ses conditions limites, sa torsion ou encore sa nature électrostatique.

Chapitre 2

Le modèle du ver : un polymère semi-flexible

Les ouvrages de référence en physique des polymères introduisent le modèle du ver comme limite continue d'une chaîne discrète de monomères présentant un couplage entre plus proches voisins, type modèle de spins classiques d'Heisenberg. Apparaît alors de manière naturelle une corrélation à plus grande échelle, sur une longueur de persistance¹. Cette approche est motivée par son caractère pédagogique : on voit comment l'introduction d'une corrélation dans une chaîne initialement *flexible* (une marche aléatoire) peut se traduire par une *rigidité* élastique à la limite continue. Nous préférons considérer d'emblée cette limite *semi-flexible*, le cas discret apparaissant de manière assez intuitive lors des calculs d'intégrales de chemin (annexe A).

2.1 Formalisme : quelques résultats exacts

Dans le modèle du ver le polymère est une courbe continue et différentiable de longueur L , couverte par l'abscisse curviligne $0 \leq s \leq L$. Il possède une énergie élastique de courbure, et peut subir le travail d'une force uniforme $f\vec{e}_z$ constante appliquée par un opérateur extérieur dans le but de l'étirer (figure 2.1). En utilisant le vecteur tangent \vec{t} à sa courbe, le ver a alors une *énergie libre*

$$E[\vec{t}; L, f\vec{e}_z] = \frac{1}{2} (k_B T \times A) \int_0^L \left[\frac{d\vec{t}(s)}{ds} \right]^2 ds - f \int_0^L \vec{e}_z \cdot \vec{t}(s) ds.$$

Nous écrivons son module d'élasticité ($k_B T \times A$) en fonction de la constante de Boltzmann k_B , de la température T du milieu ainsi que d'une longueur A . Nous reparlerons en détails de A plus tard, mais notons que cette longueur est une mesure de la rigidité élastique du polymère. Dans ce sens elle est assimilable à sa longueur de persistance. Nous insistons sur le fait que E est bien une énergie

¹En toute rigueur il est préférable à ce niveau de parler de longueur de corrélation.

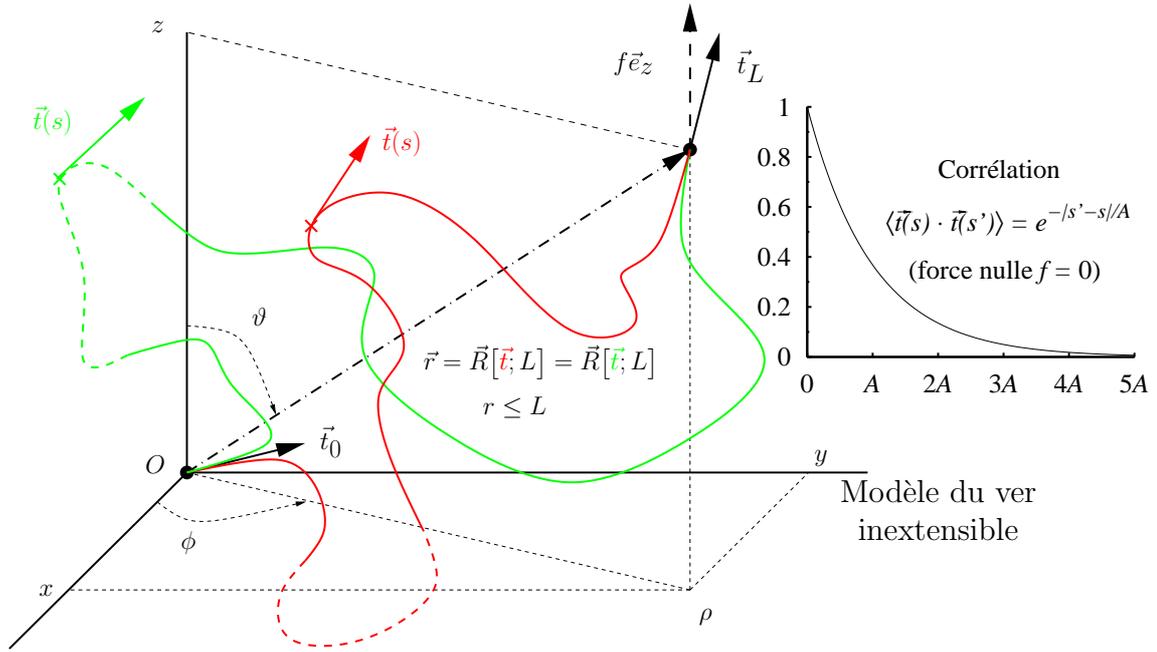


FIG. 2.1 – Le modèle du ver ou WLC, de l’anglais « Worm Like Chain ». Le polymère est une courbe continue et différentiable en chacun de ses points, repérés par l’abscisse curviligne s . Il est décrit par son vecteur tangent $\vec{t}(s)$ unitaire, reflet de son inextensibilité : sa distance bout à bout $r = |\vec{r}|$ ne peut excéder sa longueur L . La courbure de fragments dont la taille avoisine sa longueur de persistance A est particulièrement coûteuse.

libre, même si nous parlons le plus souvent d’« énergie élastique » au cours de ce manuscrit : nous ne faisons pour le moment aucune hypothèse sur les configurations \vec{t} étudiées. Finalement nous utiliserons l’acronyme WLC de l’anglais « Worm-Like Chain » pour désigner le modèle du ver.

L’ensemble de l’information statistique (*i.e.* configurationnelle) du WLC est contenu dans sa fonction de partition. Elle s’exprime comme l’intégrale du poids de Boltzmann d’une configuration quelconque $\vec{t}(s)$, sur l’ensemble des chemins possibles

$$Z(L, \vec{t}_L | 0, \vec{t}_0 ; f) = \int \mathcal{D}\vec{t} \exp \left\{ -\beta E[\vec{t}; L, f\vec{e}_z] \right\} \quad \text{avec} \quad \beta = \frac{1}{k_B T}.$$

Par « chemins possibles » nous entendons les configurations dont la courbe, conservant sa propriété de différentiabilité vérifie les orientations limites (initiale et finale) $\vec{t}(0) = \vec{t}_0$ et $\vec{t}(L) = \vec{t}_L$ en « début » et « fin » de chaîne respectivement (figure 2.1). Nous restreignons notre étude aux polymères inextensibles, y compris dans le cas d’une importante force d’étirement extérieure. Cela signifie

que la *distance bout à bout*, ou l'*extension*

$$\vec{R}[\vec{t}; L] = \int_0^L \vec{t}(s) ds = \left[\int_0^L \cos \varphi(s) \sin \theta(s) ds \right] \vec{e}_x + \left[\int_0^L \sin \varphi(s) \sin \theta(s) ds \right] \vec{e}_y + \left[\int_0^L \cos \theta(s) ds \right] \vec{e}_z$$

ne peut jamais réaliser de valeur \vec{r} dont la norme $r = |\vec{r}|$ est plus grande que la longueur L . Cette contrainte supplémentaire se traduit localement par l'*unitarité* de $\vec{t}(s)$. C'est d'ailleurs pourquoi nous en avons omis la norme dans les composantes de la distance bout à bout, qui sont exprimées ci-dessus en coordonnées sphériques (θ, φ) de \vec{t} que le lecteur prendra garde à ne pas confondre avec celles (ϑ, ϕ) de $\vec{R}[\vec{t}; L]$. Sans plus de justifications pour l'instant (voir § 3.4), signalons que c'est essentiellement cette unitarité qui rend difficile l'obtention de résultats analytiques exacts.

Le calcul de l'intégrale de chemins est simplifié par l'analogie du WLC avec le problème du rotateur rigide quantique soumis à un champ extérieur uniforme, d'origine électrique par exemple [Mes95, Mer97]. L'équivalence est démontrée dans l'annexe A. À ce niveau il suffit de saisir qu'elle réduit l'étude des propriétés du WLC à un problème aux valeurs propres. La fonction de partition, ou *propagateur* devient

$$Z(\Omega_L | \Omega_0; L, f) = \sum_{\ell, m} \sum_{\ell', m'} Y_{\ell}^m(\Omega_L) \tilde{Y}_{\ell'}^{m'}(\Omega_0) \langle \ell, m | \exp \left[-\frac{L}{A} \widehat{H}(f) \right] | \ell', m' \rangle.$$

Nous conservons le repérage du vecteur tangent par ses coordonnées sphériques, que nous symbolisons par l'angle solide $\Omega = (\varphi, \theta)$. L'opérateur $\widehat{H}(f)$ est le hamiltonien du rotateur sous champ (nous avons factorisé son moment d'inertie A)

$$\widehat{H}(f) = \frac{\widehat{L}^2}{2} - \beta A f \widehat{\cos \theta},$$

qui se réduit en champ nul ($f = 0$) à l'opérateur moment cinétique \widehat{L}^2 [Mes95, Mer97]. Les harmoniques sphériques $\langle \Omega | \ell, m \rangle = Y_{\ell}^m(\Omega)$ sont donc la base indiquée pour exprimer ses éléments de matrice (voir rappels en § B.2). L'application d'un champ ($f \neq 0$) laisse le système invariant par rotation autour de la direction d'application. Cette symétrie cylindrique se traduit par des éléments de matrice nuls suivant les composantes $m' \neq m$. Les sous matrices restantes sont tridiagonales et symétriques. Plus précisément $\langle \ell, m | \widehat{H}(f) | \ell', m' \rangle = H_{\ell, \ell'}^{(m)}(f) \delta_{m, m'}$ avec (§ B.1)

$$H_{\ell, \ell'}^{(m)}(f) = \frac{\ell(\ell+1)}{2} \delta_{\ell, \ell'} - \beta A f \left[\sqrt{\frac{(\ell'-m)(\ell'+m)}{(2\ell'+1)(2\ell+1)}} \delta_{\ell, \ell'-1} + \sqrt{\frac{(\ell-m)(\ell+m)}{(2\ell+1)(2\ell'+1)}} \delta_{\ell', \ell-1} \right]$$

Nous serons dans un premier temps (§§ 4.1 et 4.2, voir plus bas) intéressés par la seule sous matrice de $m = 0$ dont les composantes sont

$$H_{\ell,\ell'}^{(0)}(f) = \frac{\ell(\ell+1)}{2} \delta_{\ell,\ell'} - \beta A f \frac{\ell' \delta_{\ell,\ell'-1} + \ell \delta_{\ell',\ell-1}}{\sqrt{(2\ell'+1)(2\ell+1)}}.$$

Nous retenons que le calcul du propagateur se réduit à celui de l'exponentielle de cette matrice, c'est-à-dire à sa diagonalisation. À part en champ nul où le hamiltonien est déjà diagonal, une telle diagonalisation exacte n'est pas possible analytiquement². Elle est toutefois réalisable numériquement comme nous le verrons au chapitre 4 (§ 4.1). Nous y proposons une évaluation du facteur de cyclisation du WLC à partir de ce propagateur, ou plus exactement de sa version indépendante des orientations limites Ω_0 et Ω_L . En sommant sur ces dernières, puis en reconnaissant (§ B.2) l'harmonique sphérique $Y_0^0(\Omega) = 1/\sqrt{4\pi}$ dans la mesure d'intégration³ symétrisée $\int d\Omega_0 d\Omega_L/(4\pi)$, l'orthonormalisation des $Y_\ell^m(\Omega)$ donne alors une expression du type *amplitude du vide*

$$Z(L, f) = \int \frac{d\Omega_0 d\Omega_L}{4\pi} Z(\Omega_L|\Omega_0; L, f) = \langle 0, 0 | \exp \left[-\frac{L}{A} \widehat{H}(f) \right] | 0, 0 \rangle.$$

Nous la notons également Z , car le contexte ne permet aucune confusion.

Le calcul analytique exact de la fonction de partition est impossible en champ non nul. En fait nous n'avons encore présenté aucun résultat analytique exact pour le WLC, nous permettant même de laisser en suspens une interprétation convenable du paramètre A . Dans ces conditions il est légitime que le lecteur se demande pourquoi nous avons pris la peine d'insister sur l'analogie avec un problème quantique. Rappelons tout d'abord que les résultats analytiques exacts sont peu nombreux pour le WLC. Enfin, les plus utilisés d'entre eux sont obtenus grâce à l'analogie en question. Nous terminons cette section par la présentation de certaines de ces propriétés, en champ nul ($f = 0$).

Par isotropie (symétrie sphérique) la variance en distance bout à bout du WLC coïncide avec son moment d'ordre 2

$$\sigma_r^2 = \langle \vec{R}^2[\vec{t}; L] \rangle = \int_0^L \int_0^L \langle \vec{t}(s) \cdot \vec{t}(s') \rangle ds ds'.$$

Sa fonction de corrélation ($s' > s$) s'exprime au moyen du propagateur par

$$\langle \vec{t}(s) \cdot \vec{t}(s') \rangle = \int \frac{d\Omega d\Omega'}{4\pi} \vec{t}(s) \cdot \vec{t}(s') \times Z(\Omega'|\Omega; s' - s, 0) = \exp\left(-\frac{s' - s}{A}\right).$$

²On pourrait tenter une approximation perturbative [Mes95, Sch68, Sak94] en $\beta A f$ (§ 3.4).

³Notations : $\int d\Omega = \int_0^\pi r d\theta \int_0^{2\pi} r \sin\theta d\varphi = r^2 \int_0^{2\pi} d\varphi \int_{-1}^1 d(\cos\theta)$.

L'égalité finale s'obtient par exemple en reconnaissant dans le produit scalaire des vecteurs tangents $\vec{t}(s) \cdot \vec{t}(s')$, les harmoniques sphériques de $\ell = 1$ (§ B.2)

$$\begin{aligned}\cos \varphi \sin \theta &= \sqrt{\frac{2\pi}{3}} [Y_1^{-1}(\Omega) - Y_1^1(\Omega)] \\ \sin \varphi \sin \theta &= i\sqrt{\frac{2\pi}{3}} [Y_1^{-1}(\Omega) + Y_1^1(\Omega)] \quad \text{idem } \Omega' = (\theta', \varphi') \\ \cos \theta &= 2\sqrt{\frac{\pi}{3}} Y_1^0(\Omega).\end{aligned}$$

Quand nous injectons le développement de Z sur les harmoniques sphériques, n'en subsiste alors que les termes $\ell = 1$ grâce, là encore, à l'orthonormalisation des harmoniques sphériques. La longueur de corrélation du WLC est en fait A .⁴ C'est dans ce sens qu'elle s'interprète comme sa longueur de persistance.⁵ Enfin, la fonction de corrélation partage avec le propagateur l'*invariance par translation*, le long du polymère : ils ne dépendent que de $(s' - s)$. Revenons à la variance du WLC

$$\sigma_r^2 = 2A^2 (L/A + e^{-L/A} - 1).$$

Le comportement obtenu est cohérent. Aux petites longueurs $L \ll A$ les effets élastiques dominant et le polymère est rigide $\sigma_r \lesssim L$ (élongation quasi-maximale). Au contraire les grandes longueurs $L \gg A$ favorisent les effets entropiques et le polymère est alors flexible $\sigma_r \simeq \sqrt{2A \times L}$ (marche quasi-aléatoire). Ces deux limites sont détaillées aux prochaines sections (§§ 2.2 et 2.3). Bien que similaire le calcul des moments d'ordres pairs supérieurs devient rapidement pénible [STY67, Kle04]. En guise d'illustration, à l'ordre 4

$$\frac{\langle \vec{R}^4[\vec{t}; L] \rangle}{A^4} = \frac{4}{27} \left[45 \left(\frac{L}{A} \right)^2 - 156 \frac{L}{A} + 214 - \left(54 \frac{L}{A} + 216 \right) e^{-L/A} + 2e^{-3L/A} \right].$$

Notons que nous avons obtenu ces moments *sans avoir besoin* de la distribution en distance bout à bout elle-même. Nous verrons au chapitre 3 qu'elle n'est pour le moment pas calculable analytiquement. La prochaine section caractérise la limite élastique du WLC, pour laquelle il n'est alors pas question de distribution. La section suivante (§ 2.3) propose le polymère gaussien comme limite entropique du WLC. Ce modèle est en revanche exactement résoluble : c'est l'occasion pour nous d'introduire des quantités *fondamentales*. Nous nous référerons en effet fréquemment à cet *exemple pédagogique* dans la première partie de ce travail (chapitres 3 et 4).

⁴Si l'on souhaite conserver la définition $D = 3$ de A , alors la longueur de corrélation/persistance en dimension D quelconque est $A \times 2/(D - 1)$ [Kle04].

⁵La définition usuellement retenue [Yam97] est basée sur la corrélation de la distance bout à bout à l'orientation initiale, d'un long polymère : $A = \lim_{L \rightarrow \infty} \langle \vec{t}_0 \cdot \vec{R}[\vec{t}; L] \rangle$.

2.2 Limite élastique : approximation de point col

Dans la limite élastique ($L \ll A$), les contributions entropiques à l'énergie du WLC sont négligeables et les configurations sont alors essentiellement rigides. Contrairement à la section précédente, nous faisons donc maintenant une hypothèse forte sur les configurations \vec{t} étudiées : à distance bout à bout et orientations limites fixées, nous ne retenons que le chemin minimisant l'énergie du WLC.⁶ Nous sommes en particulier intéressés par les orientations limites réalisant, à distance bout à bout fixée, la configuration d'énergie élastique minimale.

Avant de faire une recherche systématique de cet extremum, est-il possible de l'intuiter ? Nous nous plaçons toujours en l'absence de force extérieure : $f = 0$. Nous voulons modéliser à terme des boucles dans la molécule d'ADN. Il est donc naturel de considérer en premier lieu le cas d'un polymère « fermé », c'est-à-dire de distance bout à bout nulle. Le cercle est alors la configuration la plus simple à laquelle on puisse penser. Dans le plan (\vec{e}_1, \vec{e}_2) du cercle le vecteur tangent du WLC

$$\vec{t}_\odot(s) = \cos\left(\frac{2\pi}{L}s\right)\vec{e}_1 + \sin\left(\frac{2\pi}{L}s\right)\vec{e}_2, \quad \text{produit} \quad \beta E[\vec{t}_\odot; L] = \frac{2\pi^2}{L/A}.$$

Pour de grandes boucles cette énergie est en effet faible : dix longueurs de persistance coûtent moins de $2 k_B T$. C'est alors l'entropie de l'ensemble des configurations omises (et qui peuvent beaucoup différer du cercle) qui est significative. En revanche la boucle d'une longueur de persistance possède une énergie d'environ $20 k_B T$, une estimation somme toute convenable comme nous le verrons avec l'ADN (§ 4.4).

Le cas du cercle est simple à traiter, mais en jouant sur ses orientations limites $\vec{t}_0 = \vec{t}_L = \vec{e}_1$ nous pouvons obtenir une énergie plus petite. C'est la configuration ainsi obtenue qui réalise l'énergie minimale et que nous appelons *point col*. Nous traitons le cas d'un polymère *a priori* « ouvert » car il demande peu d'efforts supplémentaires. Les résultats font néanmoins intervenir des fonctions spéciales *a priori* peu familières. C'est pourquoi nous proposons également en annexe E un modèle géométrique simple, inspiré de l'exemple du cercle ci-dessus. Nous n'utilisons pas ce modèle du *lignes & cercles* dans le reste de notre travail mais il constitue une bonne première approximation du point col, que nous allons maintenant chercher.

Nous nous efforçons de rester fidèle dans la présentation du calcul à l'article de Yamakawa et Stockmayer [Yam97, YS72] pour des raisons qui deviendront apparente en section 2.4. Nous faisons les hypothèses suivantes concernant la configuration de point col. Tout d'abord le polymère est dans un plan $(\vec{e}_r, \vec{e}_\perp)$

⁶En d'autres termes, nous nous intéressons au chemin *classique* minimisant l'action du système dont nous venons de définir l'intégrale des chemins (annexe A, [FH65]).

porté par sa distance bout à bout $\vec{r} = r \vec{e}_r$ et un vecteur \vec{e}_\perp qui lui est perpendiculaire. Les configurations sont alors symétriques par rapport à l'axe porté par \vec{e}_\perp . Enfin, celles présentant des structures du type plectonémique (croisements) sont écartées d'emblée (figures 2.2 et 2.3). La minimisation se fait sur l'ensemble des

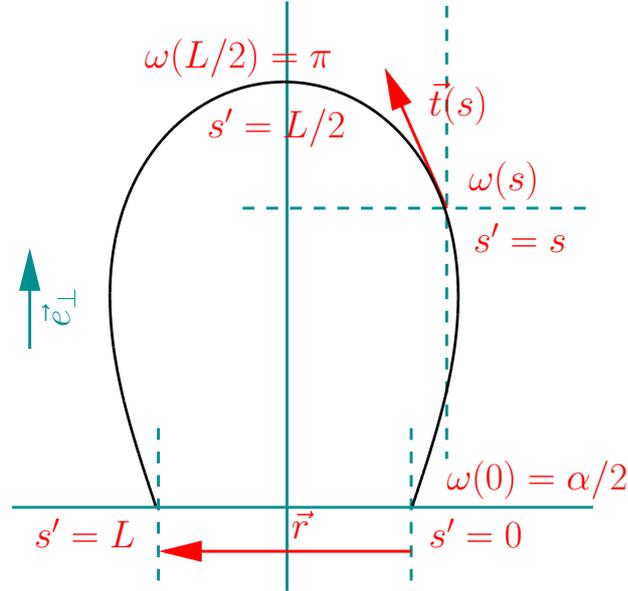


FIG. 2.2 – Nous cherchons la configuration planaire et symétrique d'un WLC, dont les orientations initiales α minimisent l'énergie élastique.

configurations rigides de distance bout à bout r fixée. La symétrie axiale nous permet de ne considérer qu'une moitié de polymère. Si $\omega(s)$ est l'angle entre $\vec{t}(s)$ et \vec{r} dans le plan $(\vec{e}_r, \vec{e}_\perp)$, alors le polymère a une énergie

$$\beta E[\omega; L, f = 0] = 2 \times \frac{A}{2} \int_0^{L/2} \dot{\omega}^2(s) ds \quad \dot{\omega}(s) = \frac{d\omega(s)}{ds}.$$

et sa configuration doit vérifier la contrainte

$$\int_0^{L/2} \cos \omega(s) ds = -r/2 \quad \text{soit} \quad \int_0^{L/2} [\cos \omega(s) + r/L] ds = 0.$$

Nous prenons en compte cette dernière en introduisant un multiplicateur de Lagrange [Lan86] (adimensionné) γ . L'équation décrivant la configuration du polymère d'extension r fixée, est par analogie l'équation du mouvement du rotateur rigide sous champ (annexe A). Nous cherchons donc à résoudre l'équation d'Euler/Lagrange associée à l'action (pour le polymère : l'énergie contrainte) [LL, Gol80]

$$\mathcal{S}[\omega; L, \gamma, r] = \int_0^{L/2} \underbrace{\left\{ \dot{\omega}^2(s) - \frac{\gamma}{L^2} [\cos \omega(s) + r/L] \right\}}_{\text{lagrangien } \mathcal{L}(\omega, \dot{\omega}) = T - V} ds. \quad (2.1)$$

La différentiabilité du WLC exige $\omega(L/2) = \pi$ (pas de cassure en $L/2$), mais nous pouvons résoudre l'équation pour une orientation initiale $\omega(0) = \alpha/2$ quelconque. L'équation du mouvement est

$$\frac{d}{ds} \left[\frac{\partial \mathcal{L}(\omega, \dot{\omega})}{\partial \dot{\omega}} \right] - \frac{\partial \mathcal{L}(\omega, \dot{\omega})}{\partial \omega} = 0 \quad \text{soit} \quad 2\ddot{\omega}(s) - \frac{\gamma}{L^2} \sin \omega(s) = 0,$$

et le système est alors complètement intégrable. Il n'a en effet qu'un degré de liberté ω , et il possède une intégrale première du mouvement (intégrer ci-dessus)

$$\dot{\omega}^2(s) + \frac{\gamma}{L^2} \cos \omega(s) = \text{cte} \quad \text{notée} \quad \dot{\omega}^2(s) + \frac{\gamma}{L^2} [\cos \omega(s) + r/L] = \frac{2\varepsilon}{L^2}.$$

Nous reconnaissons le hamiltonien $\mathcal{H} = T + V$ du système, qui ne dépend pas explicitement du temps s et s'identifie donc à son énergie. Plus précisément, l'énergie du polymère s'exprime en $k_B T$ comme $\frac{\varepsilon}{L/A}$. Intégrons cette constante du mouvement, c'est-à-dire

$$L \dot{\omega}(s) = \pm \sqrt{2\varepsilon - \gamma [\cos \omega(s) + r/L]}.$$

Les configurations possibles sont de deux types selon leur courbure $\dot{\omega}(s)$:

- (i) les premières ont une courbure positive tout du long $0 \leq s \leq L/2$;
- (ii) celle des secondes passe de négative à positive pour un s intermédiaire.

La généralisation à (ii) étant directe, la suite du calcul n'est ici présentée que pour le cas (i). En passant des variables ω , γ et ε à

$$\Omega(s) = \frac{\pi - \omega(s)}{2}, \quad 16\xi^2 = 2\varepsilon + \gamma(1 - r/L) \quad \text{et} \quad k^2 = \frac{2\gamma}{2\varepsilon + \gamma(1 - r/L)}$$

nous intégrons alors (le « - » vient du changement de variables)

$$-\frac{L}{2} \dot{\Omega}(s) = \xi \sqrt{1 - k^2 \sin^2 \Omega(s)}$$

sur $0 \leq s \leq L/2$, ce qui fournit une première équation intégrale

$$\xi = \int_0^\beta \frac{d\Omega}{\sqrt{1 - k^2 \sin^2 \Omega}} = F(\beta, k), \quad (2.2)$$

avec l'« orientation » initiale $\beta = \Omega(0)$ et F l'intégrale elliptique de Legendre du premier type [AS70]. Notons que l'orientation à mi-longueur n'apparaît plus explicitement car $0 = \Omega(L/2)$. Comme $\sin^2 \Omega(s) = \frac{1}{2} [1 + \cos \omega(s)]$, la constante de mouvement peut se réécrire

$$-\frac{L}{2} \dot{\Omega}(s) \sin^2 \Omega(s) = \xi \frac{1}{2} [1 + \cos \omega(s)] \sqrt{1 - k^2 \sin^2 \Omega(s)}$$

qui intégrée sur $0 \leq s \leq L/2$ fournit une deuxième équation intégrale

$$(1 - r/L) \xi = 2 \int_0^\beta \frac{\sin^2 \Omega}{\sqrt{1 - k^2 \sin^2 \Omega}} d\Omega,$$

où l'on a utilisé la contrainte fixant l'extension r . Nous pouvons l'exprimer en fonction de l'intégrale elliptique de Legendre du deuxième type E [AS70], en réutilisant la première équation intégrale (2.2)

$$[2 - k^2(1 - r/L)] \xi = 2 \int_0^\beta \sqrt{1 - k^2 \sin^2 \Omega} \, d\Omega = 2E(\beta, k). \quad (2.3)$$

La solution générale à l'équation du mouvement est caractérisée par le paramètre k (annexe C). Fixant les valeurs de l'extension relative r/L et de l'orien-

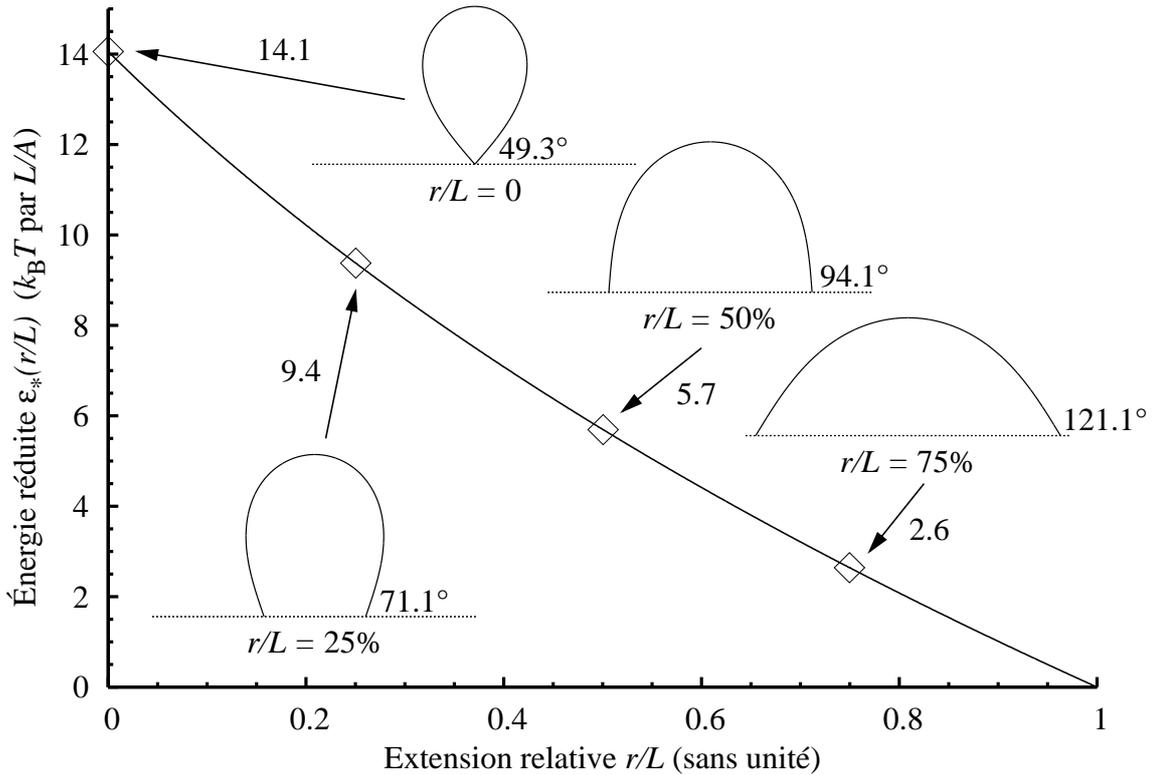


FIG. 2.3 – Pour un WLC de distance bout à bout relative r/L fixée, la configuration de point col possède les orientations initiale et finale (voir angles) correspondant au minimum de l'énergie élastique. Elle constitue sa configuration *la plus stable*, dominant aux faibles longueurs $L \ll A$ (limite élastique). Ainsi, les boucles ($r/L = 0$) les plus stables ne sont pas des cercles. L'énergie du point col s'exprime en $k_B T$ à partir de l'énergie réduite $\frac{\varepsilon_*(r/L)}{L/A}$.

tation initiale β , la satisfaction commune des deux équations intégrales (2.2) et (2.3) produit la valeur de k cherchée

$$[2 - k^2(1 - r/L)] F(\beta, k) = 2E(\beta, k).$$

Avant de pousser plus loin rappelons que nous nous intéressons à la configuration d'orientation initiale donnant l'énergie élastique minimale (le point col). En

fonction de ξ et de k que nous venons de déterminer, ε s'écrit alors

$$\varepsilon = 8\xi^2 - \frac{\gamma}{2}(1 - r/L) = 4\xi^2 [2 - k^2(1 - r/L)].$$

Sa dérivée par rapport à α (en fait β) s'annule pour

$$8\xi \left\{ [2 - k^2(1 - r/L)] \frac{d\xi}{d\beta} - \xi k(1 - r/L) \frac{dk}{d\beta} \right\} = 0.$$

Comme $\xi \neq 0$ (car $\beta \geq 0$) et qu'en dérivant également (2.3) on s'aperçoit que⁷

$$\begin{aligned} [2 - k^2(1 - r/L)] \frac{d\xi}{d\beta} - 2\xi k(1 - r/L) \frac{dk}{d\beta} &= \\ &= 2\sqrt{1 - k^2 \sin^2 \Omega} - k \frac{dk}{d\beta} \times 2 \underbrace{\int_0^\beta \sqrt{1 - k^2 \sin^2 \beta} d\Omega}_{=(1-r/L)\xi}, \end{aligned}$$

nous voyons que la condition initiale α_* réalisant le minimum de l'énergie réduite ε est reliée au k_* de la configuration correspondante (le point col), par

$$k_* \sin \beta_* = 1 \quad \text{et donc} \quad k_* > 1.$$

Or les intégrales elliptiques de (2.2) et (2.3) ne sont valables que pour $k < 1$. Nous réalisons donc la transformation de Jacobi réelle : $x = \frac{1}{k}$ et $\sin y = k \sin \Omega$. Le paramètre x de la configuration recherchée est alors solution de

$$(1 + r/L) F[B_0(x), x] = 2 E[B_0(x), x]$$

avec la fonction $B_0(x) = B(x, 0)$ issue de $B(x, s) = \arcsin[\sin \Omega(s)/x]$. De même l'énergie élastique réduite devient

$$\varepsilon(r/L; x, \alpha) = 4 F^2[B_0(x), x] [2x^2 - (1 - r/L)].$$

Le point col, réalisant l'énergie élastique minimale

$$\varepsilon_*(r/L) \equiv \varepsilon(r/L; x_*, \alpha_*)$$

vérifie $\sin \beta_*/x_* = 1$, c'est-à-dire $\omega_*(0) = \alpha_*/2 = 2 \arccos x_*$. Alors $B_0(x_*) = \frac{\pi}{2}$ et F, E deviennent les intégrales elliptiques de Legendre *complètes* du premier et deuxième type respectivement [AS70]. Une configuration type (i) correspond à une orientation initiale $\alpha < \alpha_*$, tandis que $\alpha > \alpha_*$ produit un cas (ii).

Nous vérifions pour finir que la configuration de point col totalement étirée $r = L$ est obtenue pour l'orientation initiale $\omega_*(0) = \pi$ (ou $x_* = 0$) et qu'elle est d'énergie nulle $\varepsilon_*(1) = 0$. Ainsi, parmi les distances bout à bout r possibles (figure 2.3), elle réalise bien la moins coûteuse des configurations de point col. Nous confirmons également que la boucle $r = 0$ d'énergie élastique minimale n'est pas un cercle : elle est obtenue pour $\omega_*(0) \simeq 49.3^\circ$ (ou $x_* \simeq 0.90891$). Si son énergie $\frac{\varepsilon_*(0)}{L/A} \simeq \frac{14.055}{L/A} k_B T$ est aussi inversement proportionnelle à la longueur réduite de la boucle, elle est toutefois inférieure d'environ 28.8 % à celle du cercle (voir aussi la figure 2.6).

⁷Utiliser $\frac{d}{dt} \left[\int_{a(t)}^{b(t)} f(x, t) dx \right] = \frac{db}{dt}(t) f[b(t), t] - \frac{da}{dt}(t) f[a(t), t] + \int_{a(t)}^{b(t)} \frac{\partial f}{\partial t}(x, t) dx$ (Leibniz).

2.3 Chaîne entropique : polymère gaussien

Dans la chaîne gaussienne [Yam97], le polymère est un ensemble de $(N + 1)$ points couplés par N ressorts de constantes de raideur $\frac{3}{2} \frac{k_B T}{\sigma_r^2/N}$. C'est une chaîne de N monomères *extensibles*, dont les longueurs sont individuellement distribuées selon une gaussienne de moyenne nulle et de variance σ_r^2/N (figure 2.4).⁸ Ce

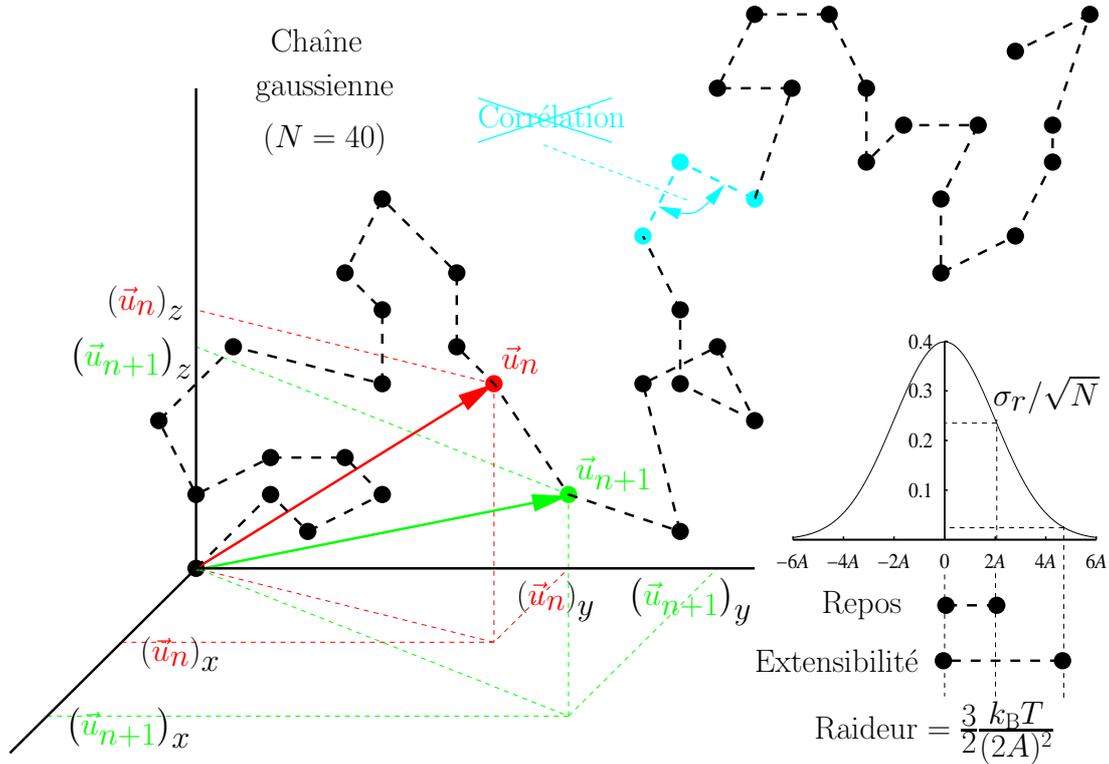


FIG. 2.4 – Les N monomères d'une chaîne gaussienne sont *indépendants* et *extensibles*. Leur longueur est distribuée selon une *gaussienne* dont l'écart-type est en général assimilé au double de la longueur de persistance (élément de Kuhn) du polymère étudié. Autrement dit chaque monomère est un ressort de raideur *entropique*, de longueur au repos $2A$.

modèle ne présente *aucune corrélation* dans l'orientation des monomères, c'est pourquoi on parle de polymère flexible ou encore de chaîne entropique. À cause de son extensibilité ce modèle est peu pertinent pour décrire l'étirement par une force extérieure. Nous nous plaçons donc d'emblée à force nulle $f = 0$. Nous fixons l'origine \vec{u}_0 de la chaîne. Sa fonction de partition s'obtient en sommant le poids de Boltzmann associé à la chaîne $(\vec{u}_1, \dots, \vec{u}_N)$, sur l'ensemble des configurations possibles

$$Z = \prod_{n=0}^{N-1} \int d\vec{u}_{n+1} \exp\left[-\frac{3}{2} \frac{(\vec{u}_{n+1} - \vec{u}_n)^2}{\sigma_r^2/N}\right] = \zeta^3$$

⁸Moyenne nulle et écart-type $\sigma_r \sim \sqrt{N}$ sont des caractéristiques d'une marche aléatoire.

L'isotropie permet la dernière factorisation. Soit v_n la composante de \vec{u}_n selon l'une quelconque des directions \vec{e}_x , \vec{e}_y ou \vec{e}_z . Nous gardons délibérément intacte l'intégrale selon v_N et calculons la fonction de partition réduite ζ par *changement de jauge* $w_{n+1} = v_{n+1} - v_n$. Le produit des $(N-1)$ intégrales couplant deux à deux les N variables v_n devient un produit de N intégrales sur N variables vérifiant la contrainte globale $\sum_{n=0}^{N-1} w_{n+1} = v_N - v_0$. Nous fixons en fait la distance bout à bout de la chaîne selon la direction considérée ($v_N - v_0$)

$$\zeta = \int_{-\infty}^{\infty} dv_N \left[\prod_{n=0}^{N-1} \int_{-\infty}^{\infty} dw_{n+1} \exp\left(-\frac{w_{n+1}^2}{2\sigma_r^2/3}\right) \right] \delta\left[\sum_{m=0}^{N-1} w_{m+1} - (v_N - v_0)\right].$$

La contrainte est imposée par une distribution de Dirac (§ D.2), dont la représentation spectrale

$$\delta\left[\sum_{m=0}^{N-1} w_{m+1} - (v_N - v_0)\right] = \int_{-\infty}^{\infty} \frac{dk}{2\pi} e^{-ik(v_N - v_0)} \prod_{m=0}^{N-1} \exp(ik w_{m+1})$$

permet de réécrire ζ

$$\zeta = \int_{-\infty}^{\infty} dv_N \int_{-\infty}^{\infty} \frac{dk}{2\pi} e^{-ik(v_N - v_0)} \left[e^{-\frac{\sigma_r^2}{6} k^2} \underbrace{\int_{-\infty}^{\infty} e^{-\frac{3}{2\sigma_r^2} (w_{n+1} - i\frac{\sigma_r^2}{3} k)^2} dw_{n+1}}_{=\sqrt{2\pi\sigma_r^2/3}} \right]^N$$

où nous avons réalisé l'intégrale gaussienne (modulo un changement de variable dans le plan complexe). En procédant de même pour l'intégrale sur k ne reste alors que celle sur v_N , qui apparaît également gaussienne

$$\zeta = \left(\frac{\pi\sigma_r^2}{3N/2}\right)^{N/2} \times \frac{1}{\sqrt{2\pi\sigma_r^2/3}} \int_{-\infty}^{\infty} dv_N \exp\left[-\frac{(v_N - v_0)^2}{2\sigma_r^2/3}\right] = \left(\frac{\pi\sigma_r^2}{3N/2}\right)^{N/2}.$$

Nous avons en fait obtenu deux résultats. La fonction de partition est

$$Z = \left(\frac{\pi\sigma_r^2}{3N/2}\right)^{3N/2}.$$

Elle permet de normaliser le poids de Boltzmann. Dans le calcul précédent, si nous décidons finalement de ne pas intégrer sur v_N nous obtenons la distribution en distance bout à bout selon la direction considérée

$$P(v_N - v_0; \sigma_r^2) = \frac{1}{\sqrt{2\pi\sigma_r^2/3}} \exp\left[-\frac{(v_N - v_0)^2}{2\sigma_r^2/3}\right].$$

Ici $(v_N - v_0)$ représente indifféremment l'une des trois composantes de la distance bout à bout $\vec{r} = (x, y, z)$. Elles ont toutes une distribution gaussienne de moyenne nulle et de variance $\sigma_r^2/3$.

Comment utiliser le modèle gaussien pour décrire un polymère dont nous connaissons les longueurs totale L et de persistance A ? Oublions pour un moment

l'extensibilité des monomères de la chaîne gaussienne. Pour cela nous assimilons l'écart-type de leur longueur σ_r/\sqrt{N} , ou encore leur longueur au repos, au double de la longueur de persistance $2A$ appelé *élément de Kuhn* [Yam97]. Par rapport au choix naturel d'une longueur de persistance, l'élément de Kuhn permet en quelque sorte de prendre en compte l'extensibilité de la chaîne de manière effective. L'identification de la longueur du polymère $L = N \times 2A$ permet alors d'estimer la variance en distance bout à bout $\sigma_r^2 = 2AL$ du modèle gaussien adéquat.

Pour récapituler, dans le modèle gaussien un polymère de longueur totale L et de persistance A voit les composantes de sa distance bout à bout $\vec{r} = (x, y, z)$ distribuées selon une gaussienne de moyenne nulle et d'écart-type $\sqrt{2AL/3}$

$$P(x; L) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{x^2}{2\sigma_x^2}\right), \quad \sigma_x^2 = \frac{2AL}{3} \quad (\text{idem } y \text{ et } z).$$

La distance bout à bout \vec{r} est alors distribuée selon une gaussienne de moyenne nulle et d'écart-type $\sqrt{2AL}$

$$Q(\vec{r}; L) = P(x; L) P(y; L) P(z; L) = \left(\frac{3}{2\pi\sigma_r^2}\right)^{3/2} \exp\left(-\frac{3\vec{r}^2}{2\sigma_r^2}\right), \quad \sigma_r^2 = 2AL.$$

Nous insistons sur le caractère *marginal* des distributions P des composantes de la distance bout à bout. Ainsi $P(x; L)$ ne contient aucune information sur les composantes y et z , etc. Elle s'obtient par intégration⁹ selon y et z de la distribution $Q(\vec{r}; L)$.

Comme le modèle gaussien permet un calcul exact de la distribution en distance bout à bout nous pouvons également obtenir une expression pour le *facteur de cyclisation*

$$J(L) = Q(\vec{0}; L) = \left(\frac{3}{4\pi AL}\right)^{3/2} = \frac{1}{A^3} \left(\frac{3}{4\pi}\right)^{3/2} (L/A)^{-3/2}.$$

C'est une caractéristique des configurations de boucle $\vec{r} = \vec{0}$. Notons qu'il a bien la dimension d'une *concentration* (inverse d'un volume). Il dépend de la taille de la boucle en $(L/A)^{-3/2}$ et diverge par conséquent pour de faibles tailles. Comme le modèle gaussien n'est pertinent qu'aux grandes longueurs $L \gg A$ cela ne doit pas nous inquiéter. Nous pouvons associer aux boucles de taille L une énergie libre par le logarithme de $V \times Q(\vec{0}; L)$, avec V *a priori* petit

$$\beta G(L) = -\ln[A^3 J(L)] - \ln(V/A^3) \simeq \frac{3}{2} \left[\ln(L/A) - \ln\left(\frac{3}{4\pi}\right) \right].$$

En pratique $L \gg A$, et A ne représente que la moitié d'un monomère de la chaîne gaussienne. C'est pourquoi nous choisissons le volume typique $V \simeq A^3$. Pour dix

⁹Sur $]-\infty; \infty[$ et non $[-L; L]$ du fait de l'extensibilité.

longueurs de persistance on obtient alors une énergie d'environ $5.6 k_B T$. Notons toutefois que le choix $V \simeq (2A)^3$ donne $2.5 k_B T$, tandis qu'un cube de 0.5 nm d'arête produit $26.3 k_B T$.¹⁰ En fait, la symétrie sphérique suggère l'utilisation d'une boule pour V . Une estimation plus quantitative est possible à partir de la distribution *isotrope* en distance bout à bout, c'est-à-dire la distribution en extension r

$$S(r; L) = r^2 \int_0^{2\pi} d\phi \int_0^\pi \sin \vartheta d\vartheta Q(r\vec{e}_r; L) = 4\pi r^2 Q(r\vec{e}_r; L),$$

présentant un unique maximum pour $r = \sigma_r \sqrt{2/3} = \sqrt{4AL/3} \neq 0$. La probabilité pour la distance bout à bout $r\vec{e}_r$ de se trouver dans la boule de rayon r_B

$$\text{Prob}(r_B; L) = \int_0^{r_B} S(r; L) dr = \text{erf}\left(\sqrt{\frac{3r_B^2}{2\sigma_r^2}}\right) - \frac{1}{2} \frac{2\sigma_r^2}{3r_B^2} r_B S(r_B; L)$$

permet pour un petit rayon r_B d'estimer une énergie libre de cyclisation [PV00]

$$\beta G(r_B; L) = -\ln[\text{Prob}(r_B; L)]$$

Son expression analytique fait intervenir la fonction erreur $\text{erf}(x)$, calculable numériquement. Pour dix longueurs de persistance nous retrouvons une énergie d'environ $5.5 k_B T$, pour une boule de rayon 1 nm .

Remarquons pour finir que cette énergie ne permet pas de considérer les configurations de « vraie boucle », dans le sens de *fermée* : elle diverge quand $r_B \rightarrow 0^+$. Cette limite n'est toutefois pas pathologique. On peut en effet montrer que

$$\lim_{r_B \rightarrow 0^+} \left[\text{Prob}(r_B; L) / \left(\frac{4}{3} \pi r_B^3 \right) \right] = J(L)$$

tout comme $\lim_{r \rightarrow 0^+} [S(r; L)/(4\pi r^2)] = J(L)$, par définition. Plus loin dans ce manuscrit (§§ 4.2 et 5.4), nous généralisons ce raisonnement. Nous notons alors r_B comme l'extension r , le contexte évitant en général toute confusion.

2.4 Fluctuations au point col : résultat de Shimada et Yamakawa

Le point col déterminé en section 2.2 est une approche *classique* du WLC dans la mesure où, à extension relative r/L et orientations limites fixées ne correspond qu'une configuration¹¹. Nous avons en effet négligé les contributions entropiques

¹⁰C'est ce volume de 1 nm^{-3} , rapporté à $(2A)^3$, qu'utilisent Sankararaman et Marko dans leur travail sur $J(L)$ [SM05b].

¹¹Modulo une rotation du plan de la configuration autour de la distance bout à bout \vec{r} .

à l'énergie du WLC, modèle pour lequel il n'existe pas de solution exacte (approche *quantique*) comme expliqué au prochain chapitre 3. Nous pouvons toutefois espérer prendre en compte approximativement les effets entropiques omis en fluctuant la configuration de point col (approche *semi-classique*). Cette dernière constitue la configuration la *plus stable* du WLC, dans le sens où elle domine l'ensemble des configurations possibles aux faibles longueurs $L \ll A$ (grande rigidité). Des fluctuations, mêmes faibles, doivent permettre d'étendre ce domaine de validité.

Le point col est planaire et les fluctuations que nous considérons sont, dans un premier temps, également planaires. Par définition, le point col annule les fluctuations linéaire de l'action/énergie contrainte (2.1) en $\omega(s)$. Nous poussons donc jusqu'à l'ordre quadratique, et

$$\int \mathcal{D}\omega e^{-A\mathcal{S}[\omega;L,\gamma,r]} \simeq e^{-\frac{\varepsilon_*(r/L)}{L/A}} \int \mathcal{D}(\delta\omega) e^{-\frac{A}{2} \int_0^L \delta\omega(s) \widehat{M}(s;\gamma_*,r) \delta\omega(s) ds}$$

où l'intégrale sur l'opérateur $\widehat{M}(s;\gamma_*,r)$ correspond à la dérivée fonctionnelle seconde de \mathcal{S} , par rapport au chemin $\omega(s)$, évaluée en le point col. Notons qu'en pratique nous discrétisons, comme en annexe A, l'intégrale de chemin : $\delta\omega(n \frac{L}{N}) = \delta\omega_n$, les $(N+1)$ composantes d'un vecteur $|\delta\omega\rangle$. L'intégrale apparaît alors comme le produit scalaire $\langle \delta\omega | \widehat{M}(\gamma_*,r) | \delta\omega \rangle$, ou bien encore une forme quadratique en les composantes de $|\delta\omega\rangle$. En diagonalisant¹² la matrice \widehat{M} l'intégrale de chemin se ramène à un produit, d'intégrales gaussiennes sur ses vecteurs propres. Une hypothèse supplémentaire se cache derrière cette méthode : il se peut qu'une des valeurs propres soit petite, voir nulle auquel cas nous ne pouvons pas étendre l'intégration sur le vecteur propre correspondant, à $]-\infty; \infty[$ comme nous l'avons implicitement supposé. La résolution de ce problème est quelque peu technique, mais faisable [Kle04, Kul04, KML⁺05]. Nous n'explicitons pas la suite du calcul : en réalité nous sommes intéressé par le cas tridimensionnel, où les fluctuations peuvent sortir du plan du point col.

Nous choisissons arbitrairement le plan $\varphi(s) = 0$ et nous identifions $\theta(s) = \omega(s)$. Les fluctuations quadratiques en $\varphi(s)$ se séparent alors de celles en $\theta(s)$, qui restent identiques au cas 2D (matrice \widehat{M}). La nouvelle matrice possède notamment une valeur propre nulle et une autre négative. La première correspond à l'arbitraire dans le choix du plan du point col (invariance par rotation autour de la distance bout à bout \vec{r}) mais la seconde indique une instabilité du point col sous l'effet des perturbations en $\varphi(s)$. Quel sens physique accorder à ce résultat ? Si nous perturbons faiblement la boucle ($r = 0$) d'un polymère contraint à rester dans un plan, le système va revenir dans son état initial et la boucle est donc stable. En revanche, dès que le polymère n'est plus contraint au plan la moindre perturbation fait disparaître la boucle. Autrement dit les configurations de boucles sont *stables à 2D* mais *instables à 3D*. En pratique, les boucles que

¹²Cela est beaucoup plus simple pour le cercle que pour le point col.

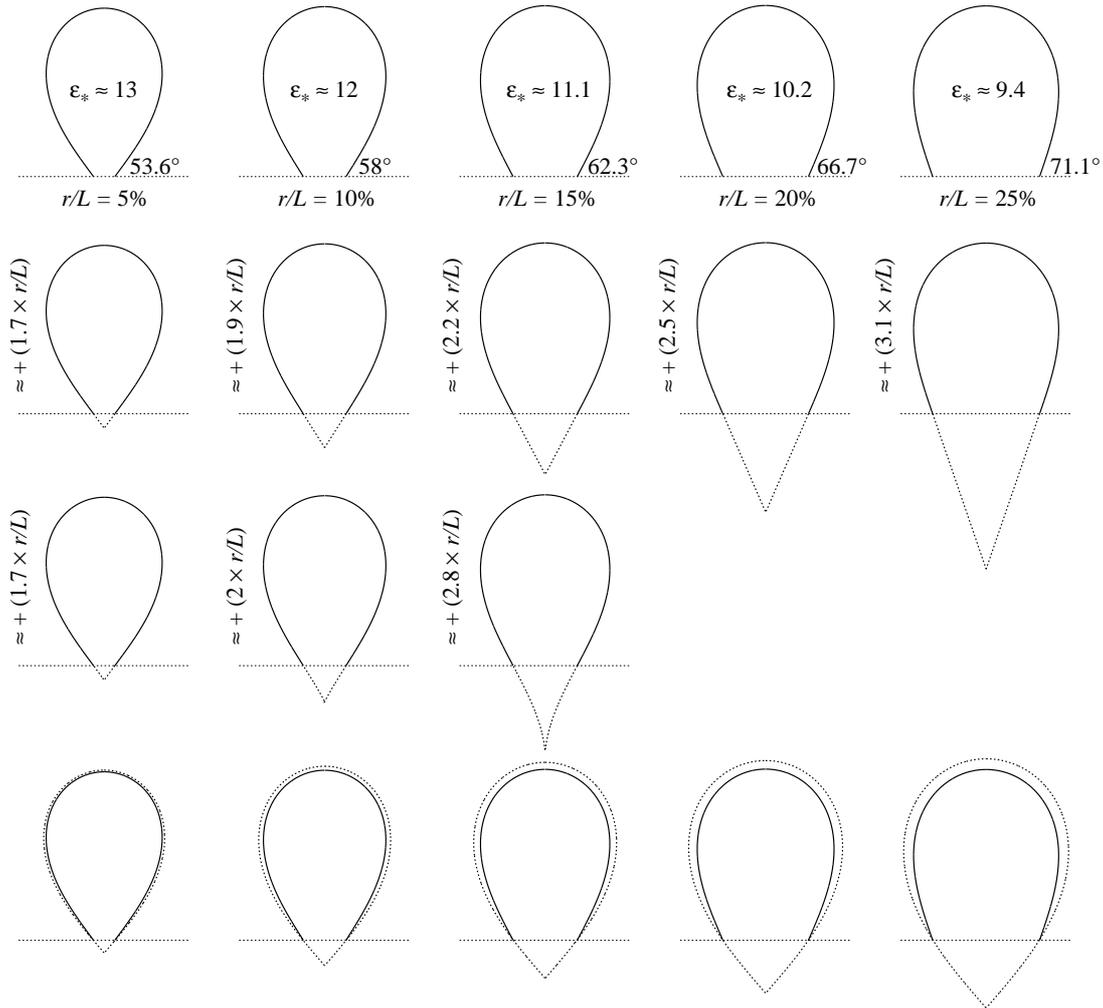


FIG. 2.5 – Estimation des fluctuations $C(r; L)$ au point col d'extension relative r/L (première ligne, repris en trait plein sur les autres lignes), comme celles $C(0; \bar{L})$ à une boucle *factice* dont la longueur $\bar{L}(r) \geq L$ prend en compte l'extension $r \geq 0$ étudiée. Deuxième ligne : fermeture du point col par des segments *linéaires*. Troisième ligne : le point col comme partie d'une *configuration* de boucle du WLC. Quatrième ligne : utilisation *du point col* pour la boucle également, avec une longueur additionnelle $\bar{L} - L = 2 \times r$. Ce choix simple est motivé par les $(\bar{L} - L)$ obtenues grâce aux deux autres méthodes (exprimées comme fonctions linéaires de r/L sur la figure). Il permet une description quantitative jusqu'à $r/L \lesssim 25\%$ (moins de 3% d'écart relatif à l'énergie, figure 2.6).

nous étudions son stabilisées par des protéines capables de s'accrocher à l'ADN en deux sites distants (le LacR p. ex.), ou des ligases liant les extrémités de la molécule de manière irréversible. L'application d'une torsion additionnelle (sur-enroulement) à l'ADN permet également une telle stabilisation, bien illustrée par la formation de plectonèmes. Récemment (2004), Kulić a également souligné

l'auto-attraction de la molécule d'ADN par exemple observable en jouant sur le solvant, ou encore l'application possible d'un champ magnétique dans le plan de la boucle [Kul04]. L'instabilité doit donc être vue comme un artefact dû à l'approximation considérée.

Les difficultés inhérentes à cette approximation ont été rigoureusement mais compréhensivement exposées, dans le cas du cercle ($r = 0$ et $\alpha = 0$), par Kulić dans sa thèse [Kul04].¹³ Une partie de ce travail est repris (2005) dans un article avec Mohrbach, Lobaskin, Thaokar et Schiessel, mais sous une forme significativement plus condensée [KML⁺05]. Dans la littérature scientifique, il existe en fait peu de travaux originaux sur cette approximation dont le type est pourtant très courant en physique. En 2003 Zhang et Crothers [ZC03] ont mené une étude similaire sur un modèle discret, résolu numériquement au final. Nous devons ensuite remonter une vingtaine d'années en arrière pour trouver l'article de Shimada et Yamakawa (S&Y). Les auteurs notent l'existence des valeurs propres nulle ou négative, mais leur traitement est quelque peu déroutant [Yam97, SY84].¹⁴ Il apparaît finalement que les fluctuations sont calculées *numériquement* en s'aidant d'ajustements *empiriques*. Ils proposent l'expression suivante du facteur de cyclisation

$$J(L) \simeq \frac{1}{A^3} \frac{112.04}{(L/A)^5} \exp\left(-\frac{14.055}{L/A} + 0.246 \times L/A\right),$$

qui montre un accord quantitatif avec les données expérimentales et numériques jusqu'à dix longueurs de persistance, complétant ainsi remarquablement la description gaussienne (§ 2.3). Leur travail est difficile à reproduire et du même coup à généraliser aux autres géométries de boucle possibles, c'est-à-dire pour le calcul des fluctuations car nous identifions tout de même l'énergie réduite du point col $14.055 = \varepsilon_*(0)$. Ce constat a encouragé de nombreux travaux appliquant telle quelle cette formule dans des domaines d'ailleurs relativement variés [SKOS01, JBH03, JHBB04, SM05a, BTM05, CW04, CW05, DSS⁺05], ou bien en proposant des adaptations toutes autant empiriques [RCA⁺99, SLR00, Rip01, DRDK02, SM05a, RSKM05, PT05b].

Comme nous avons généralisé la détermination du point col aux configurations *a priori* ouvertes, nous allons proposer une telle adaptation dans le but d'évaluer la distribution en distance bout à bout \vec{r} . Nous verrons au chapitre 3 qu'elle s'obtient, modulo la normalisation par $Z(L, f = 0)$, par transformée de Fourier (TF)¹⁵ sur les fluctuations $\delta\gamma$ (autour de γ_*) de l'intégrale de chemin ci-

¹³Les outils mathématiques se trouvent par exemple dans le livre de Kleinert [Kle04], mais les chapitres utiles sont plutôt techniques pour un novice en la matière.

¹⁴Ils déclarent que la valeur propre nulle peut être supprimée, sans en expliciter la raison. Pour la négative, ils obtiennent une situation stable en jouant sur l'ordre d'intégration sur les vecteurs propres.

¹⁵En fait par transformée de Laplace, car γ est réel.

dessus¹⁶. Nous l'écrivons donc dans le cas général

$$Q(r \vec{e}_r; L) \simeq \frac{1}{A^3} C(r; L) \exp \left[-\frac{\varepsilon_*(r/L)}{L/A} \right], \quad C(0; L) = \frac{112.04}{(L/A)^5} e^{0.246 \times L/A}.$$

Nous proposons d'estimer $C(r; L)$ à partir de $C(0; \bar{L})$ en considérant une boucle *factice* de longueur $\bar{L}(r) \geq L$, prenant en compte l'extension $r \geq 0$ de la configuration de point col *réellement* considérée (première ligne de la figure 2.5, reprise en trait plein sur les autres lignes). Pour être pertinente, cette boucle ne doit pas réaliser une énergie significativement différente de celle du point col $\varepsilon_*(r/L)$. Le plus simple consiste à prolonger artificiellement ce dernier par deux segments

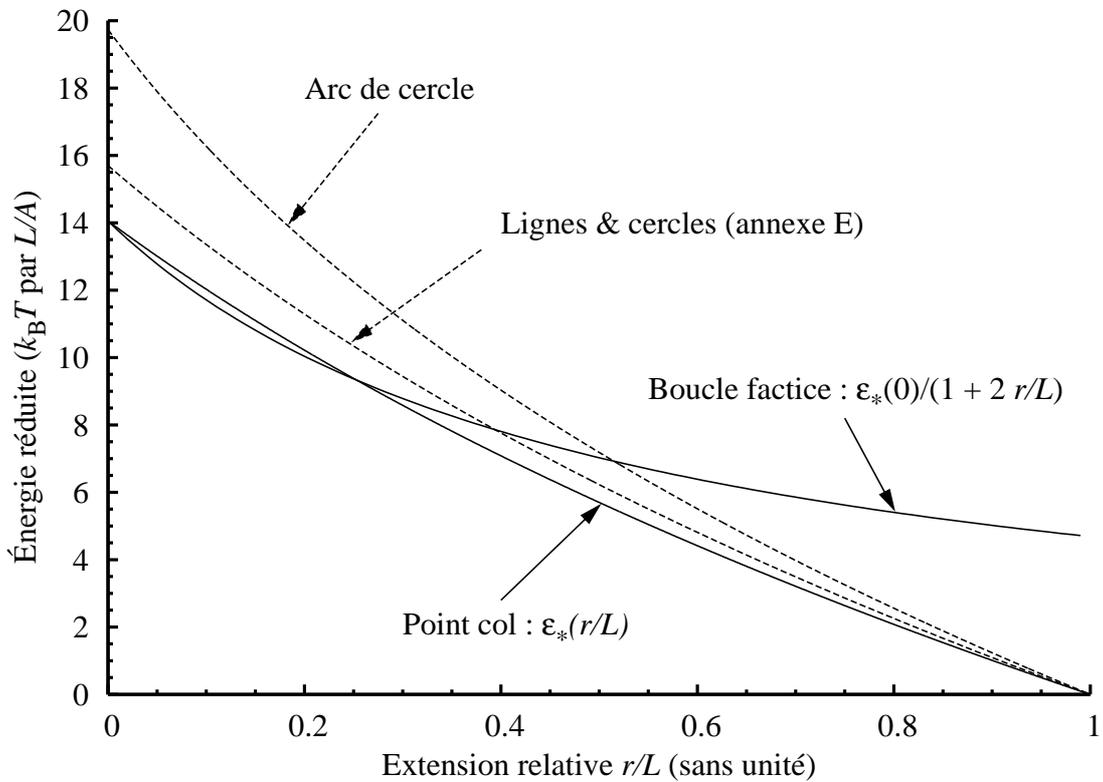


FIG. 2.6 – Le point col d'un WLC de longueur L et d'extension $r \geq 0$ est qualitativement approché par une boucle factice (celle de point col également) de longueur $(L + 2r)$. Leurs énergies élastiques (traits pleins) diffèrent de moins de 3% pour des extensions relatives allant jusqu'à 25%. Cet écart augmente significativement aux plus grandes extensions ($r/L \gtrsim 50\%$). En guise d'illustration, sont également montrées (en pointillés) l'énergie du modèle du lignes & cercles (annexe E) ainsi que celle d'un simple arc de cercle.

linéaires (donc sans énergie élastique) de longueur totale $\bar{L} - L = r / \cos(\alpha_*/2)$. La boucle obtenue¹⁷ (deuxième ligne de la figure 2.5) n'est toutefois pas « phy-

¹⁶Nous avons omis les fluctuations en γ par soucis de clarté. Cela ne change en rien nos précédentes conclusions.

¹⁷Attention, ce n'est pas le modèle du lignes & cercles présenté en annexe E.

sique », dans le sens de solution de l'équation du mouvement du WLC (§ 2.2). Nous pouvons néanmoins considérer la *boucle* de paramètre $\bar{x} = x_*$ (troisième ligne de la figure 2.5). Ce *n'est pas* le point col des configurations de boucle mais une configuration de type (ii) ($\bar{x} \leq \bar{x}_*$, voir § 2.2), qui vérifie

$$\int_0^{\frac{\bar{L}-L}{2}} \cos \bar{\omega}(s) ds = r/2 \quad \text{avec} \quad \begin{cases} \bar{\alpha} \geq \bar{\alpha}_* \simeq 2 \times 49.3^\circ & (\text{cf. figure 2.3 p. ex.}); \\ \text{et } \bar{\omega}\left(\frac{\bar{L}-L}{2}\right) = \alpha_*/2. \end{cases}$$

Pour de faibles extensions du point col originel ($r \geq 0$), ces boucles diffèrent peu de celles obtenues par l'adjonction des deux segments linéaires. En revanche, pour de plus grandes extensions l'écart se marque quelque peu, mais surtout à partir de $r/L \gtrsim 15\%$ (en fait dès que $\bar{\alpha} > \pi$) ces boucles présentent des croisements et ne sont donc plus pertinentes. Les longueurs additionnelles ($\bar{L} - L$) de ces deux types de boucles factices sont exprimées en figure 2.5 comme fonction *linéaire* de l'extension r : elles varient *grosso modo* de $1.5 \times r$ à $3 \times r$. Plutôt que de considérer la boucle de type (ii) nous proposons de choisir *systématiquement* celle du point col ($\bar{x} = \bar{x}_*$) avec une longueur additionnelle $2 \times r$ (dernière ligne de la figure 2.5). Nous n'avons ainsi plus de paramètre à ajuster

$$Q(r \vec{e}_r; L) \simeq \frac{1}{A^3} C(0; L + 2r) \exp\left[-\frac{\varepsilon_*(r/L)}{L/A}\right], \quad (2.4)$$

Bien que *simpliste* et *arbitraire*, ce choix produit des boucles factices d'énergies réduites¹⁸ $\frac{\varepsilon_*(0)}{1+2r/L}$ peu différentes de celles des points col originels $\varepsilon_*(r/L)$, pour des extensions relatives allant jusqu'à $r/L \lesssim 25\%$ (moins de 3% d'écart relatif). Cet écart devient toutefois significatif aux plus grandes extensions (environ 25% pour $r/L = 50\%$, voir la figure 2.6). Il n'est donc pas pertinent de normaliser (2.4).

Mis à part les quelques remarques que nous venons de faire, nous ne pouvons pas encore à ce niveau juger quantitativement de la pertinence de notre adaptation. Pour cela nous devons la confronter à des résultats exacts, analytiques comme numériques (§§ 4.2 et 4.3), ou encore à des approximations dont nous maîtrisons les domaines de validité (chapitre 3).

¹⁸Par rapport à la vraie longueur L .

Chapitre 3

Une histoire d'inextensibilité

Nous n'avons pas encore parlé du facteur de cyclisation pour le WLC. Au chapitre 2 l'exemple pédagogique du modèle gaussien (§ 2.3) nous a permis de vérifier qu'il est calculable à partir des diverses distributions en extension. Nous avons ensuite proposé une extension de l'approximation de point col (§ 2.2), pour laquelle le calcul de la distribution en distance bout à bout $Q(\vec{r}; L)$ est apparu problématique (§ 2.4). Bien que nous ayons obtenu une expression analytique pour les premiers moments de la distance bout à bout du WLC (§ 2.1), la méthode utilisée permettait de s'affranchir du calcul explicite des distributions $Q(\vec{r}; L)$ ou $S(r; L)$. En présence d'une force d'étirement *a priori* non nulle, l'expression de départ est le propagateur contraint

$$Q(\vec{r}; L, f) = \frac{1}{Z(L, f)} \int \mathcal{D}\vec{t} \delta\left\{\vec{r} - \vec{R}[\vec{t}; L]\right\} \exp\left\{-\beta E[\vec{t}; L, f\vec{e}_z]\right\}$$

sélectionnant parmi les configurations possibles, de distance bout à bout $\vec{R}[\vec{t}; L]$, celles réalisant la valeur particulière \vec{r} . En utilisant la représentation spectrale de la distribution de Dirac (D.2), le calcul se réduit à la transformée de Fourier (TF) du propagateur prenant en compte une *force complexe* supplémentaire, appliquée selon une direction quelconque

$$Q(\vec{r}; L, f) = \frac{1}{Z(L, f)} \int \frac{d\vec{k}}{(2\pi)^3} e^{i\vec{k}\cdot\vec{r}} \int \mathcal{D}\vec{t} \exp\left\{-\beta E[\vec{t}; L, f\vec{e}_z - i\vec{k}/\beta]\right\}.$$

Nous pouvons généraliser pour une telle force l'analogie quantique, dérivée en annexe A pour une force selon \vec{e}_z . Le calcul exact¹ de l'intégrale de chemin discrétisée requiert l'évaluation d'intégrales non gaussiennes en φ et θ , les coordonnées sphériques de \vec{t} . Une autre manière de mener le calcul consiste à écrire l'intégrale de chemins sur les vecteurs tangents non normés $\vec{\tau}(s)$ en assurant la

¹C'est-à-dire sans développements limités, au contraire du calcul ayant permis l'établissement de l'équation de Schrödinger (§ A.2).

contrainte $|\vec{\tau}(s)|^2 = 1$ par une distribution de Dirac²

$$\int \mathcal{D}\vec{t} e^{-\beta E[\vec{t}; L, f\vec{e}_z - i\vec{k}/\beta]} = \int \mathcal{D}\vec{\tau} \delta[\vec{\tau}^2 - 1] e^{-\beta E[\vec{\tau}; L, f\vec{e}_z - i\vec{k}/\beta]},$$

où la contrainte s'applique en tout s .³ En pratique nous discrétisons, comme en annexe A, l'intégrale des chemins : $\vec{\tau}(n \frac{L}{N}) = \vec{\tau}_n$. En utilisant également la représentation spectrale du Dirac (multiplicateur de Lagrange associé κ), l'intégrale de chemin s'interprète comme la TF de la fonction de partition du système de spins classiques non unitaires $\vec{\tau}_n$, auto-interagissant entre eux et couplés entre plus proches voisins avec les intégrales de recouvrement :

- pour l'auto-interaction $J_{n,n} = \frac{N}{L/A} + i\kappa$ (modulo les conditions aux bords⁴);
- entre plus proches voisins $J_{n,n+1} = -\frac{N}{L/A}$ (symétrisée);

de plus soumis à un champ $-\beta A \frac{L/A}{N} (f\vec{e}_z - i\vec{k}/\beta)$. Nous ne savons pas plus calculer cette fonction de partition, c'est-à-dire diagonaliser la matrice de recouvrement \hat{J} afin de se ramener à des intégrales gaussiennes. Ce constat est indépendant de la présence ou non d'une force extérieure ($f = 0$), ou bien encore de la contrainte fixant la distance bout à bout ($\vec{k} = \vec{0}$). Ce problème *technique* identifié, quelle en est l'interprétation *physique*? Les termes en cause permettent de fixer la contrainte d'unitarité du vecteur tangent $\vec{\tau}(s)$. Cette dernière est l'expression de l'*inextensibilité* du polymère qui est donc responsable, en premier lieu, de notre incapacité à calculer la distribution de la distance bout à bout du WLC.

3.1 Un polymère gaussien plus rigide ?

Si nous supprimons cette contrainte d'inextensibilité, nous retrouvons en fait une dépendance gaussienne. Partant de la chaîne gaussienne, pouvons nous toutefois trouver un moyen d'introduire de manière perturbative une rigidité élastique? Un développement en puissances inverses de L/A semble indiqué : dans la limite entropique $L \gg A$, ces effets additionnels s'annulent. En 1952 Daniels proposa une procédure systématique de construction d'un tel développement, pour la distribution en distance bout à bout [Dan52]. Se plaçant à force nulle $f = 0$, la correction d'ordre n doit tout d'abord conserver la normalisation de la distribution $Q(\vec{r}; L) = Q_{\text{gauss}}(\vec{r}; L) \times D_n(r; L)$. Au premier ordre $n = 1$ elle doit produire des moments d'ordre 2 et 4 s'accordant, au premier ordre en $(L/A)^{-1}$, avec les développements aux grandes longueurs $L \gg A$ des mo-

²Nous évitons d'écrire la contrainte $|\vec{\tau}(s)| = 1$, car cela produit un modèle non analytique.

³D'où la notation « fonctionnelle » (avec des crochets) du Dirac. La contrainte est souvent écrite $\prod_{0 \leq s \leq L} \delta[|\vec{\tau}(s)|^2 - 1]$, anticipant sur la discrétisation de l'intégrale de chemin.

⁴Pour lesquelles $J_{0,0} = J_{N,N} = \frac{N}{2L/A} + i\kappa$.

ments exacts du WLC (§ 2.1), à savoir :

$$\begin{aligned}\langle r^2 \rangle &\simeq 2A^2 (L/A - 1) = 2AL \times \left[1 - (L/A)^{-1} \right] \\ \langle r^4 \rangle &\simeq \frac{4}{27} A^4 \left[45 (L/A)^2 - 156 L/A + 214 \right] = (2AL)^2 \times \frac{1}{9} \left[15 - 52 (L/A)^{-1} \right].\end{aligned}$$

En vérifiant ces trois conditions, nous trouvons

$$D_1(r; L) = 1 - \frac{5/4}{L/A} + 2(r/L)^2 - \frac{33}{80} (r/L)^4 L/A.$$

En itérant la procédure pour $n > 1$ [YS72], cela produit une expression du facteur de cyclisation

$$J(L) = J_{\text{gauss}}(L) \left[1 - \frac{5/4}{L/A} + \dots \right],$$

qui diverge toujours aux faibles longueurs. Bien qu'améliorant l'évaluation gaussienne, ce développement ne permet pas de prendre en compte des effets de rigidité élastique.

3.2 Champ moyen : relaxation de la contrainte

En fait nous ne voulons pas supprimer la contrainte d'inextensibilité : nous avons déjà un modèle de polymère extensible, à savoir la chaîne gaussienne⁵. Est-il possible de trouver une contrainte rendant le calcul faisable tout en assurant une *inextensibilité effective* du polymère ? Cette recherche d'une *relaxation* pertinente de la contrainte d'unitarité a fait l'objet de nombreuses études (et de débats), s'étendant sur près de trente années [HH66, Fre71, BF85, LNN91, WRH94, LE95, ZSZ97, HT95, TH97, BTB97, Win03]. La solution retenue à l'heure actuelle consiste à contraindre $\vec{\tau}(s)$ suivant

1. $\int_0^L |\vec{\tau}(s)|^2 ds = L$ assurant une inextensibilité globale et non plus locale,
2. sauf aux extrémités du polymère où l'on conserve $|\vec{\tau}_0|^2 = |\vec{\tau}_L|^2 = 1$.

Considérons le propagateur prenant en compte ces deux contraintes, au moyen des multiplicateurs de Lagrange κ_{glob} et κ_{extr}

$$\int \mathcal{D}\vec{\tau} \exp \left\{ -\beta E \left[\vec{\tau}; L, f\vec{e}_z - i\vec{k}/\beta \right] - \kappa_{\text{glob}} \int_0^L |\vec{\tau}(s)|^2 ds - \kappa_{\text{extr}} (|\vec{\tau}_0|^2 + |\vec{\tau}_L|^2) \right\}.$$

Nous devons nous contenter dans un premier temps, des valeurs κ_{glob}^* et κ_{extr}^* qui réalisent le minimum de l'énergie libre définie *via* le logarithme de ce propagateur. Ce sont elles qui contribuent le plus aux deux transformées de Fourier⁶ intervenant dans le problème initial. Ces deux valeurs dépendent de la quantité

⁵Nous venons de voir comment « l'améliorer ».

⁶En fait, transformées de Laplace car κ_{glob} et κ_{extr} sont réels.

que l'on souhaite calculer.

Par exemple l'utilisation de l'énergie contrainte, c'est-à-dire l'argument de l'exponentielle ci-dessus, avec κ_{glob}^* et κ_{extr}^* obtenus pour $f = 0$ et $\vec{k} = \vec{0}$ (cf. expressions ci-dessus), donne une fonction de corrélation similaire à celle du WLC (§ 2.1) : *invariante par translation* le long du polymère ainsi qu'*exponentiellement décroissante*. La longueur de corrélation apparaît en revanche comme $\frac{2}{3}A = A_{\text{eff}}$. La relaxation de la contrainte d'inextensibilité du WLC produit un polymère également semi-flexible, mais *moins rigide* que ce dernier. Cela est dû aux fluctuations locales dans la norme en $\vec{r}(s)$, qu'autorise le long du polymère la contrainte globale. C'est alors à condition d'identifier A_{eff} comme la vraie longueur de persistance que ce modèle approche le mieux le WLC. En fonction de ce paramètre

$$\kappa_{\text{glob}}^* = \frac{3}{4} (A_{\text{eff}})^{-1}, \quad \kappa_{\text{extr}}^* = \frac{3}{4} \quad \text{et} \quad \frac{A}{2} = \frac{3}{4} A_{\text{eff}}.$$

Sur ce principe, considérant $\vec{k} \neq \vec{0}$ (mais toujours $f = 0$) il a été proposé, après réalisation de la TF (en \vec{k}), l'approximation

$$S(r; L) = 4\pi r^2 Q(\vec{r}; L) \propto \frac{1}{L} \frac{(r/L)^2}{[1 - (r/L)^2]^{9/2}} \exp\left[-\frac{3}{4} \frac{L/A_{\text{eff}}}{1 - (r/L)^2}\right],$$

exprimée en fonction de la longueur de persistance effective précédemment introduite. La constante de proportionnalité (non reprise ici) assurant la normalisation des distributions, ne dépend que de la longueur réduite L/A_{eff} . À longueur fixée, la distribution $S(r; L)$ en extension r :

- s'annule de manière continue, pour un polymère atteignant son régime d'étirement maximal ($r \rightarrow L$). Cela reflète l'inextensibilité globale du polymère qui est semi-flexible ;
- se comporte comme r^2 aux faibles extensions, assurant ainsi l'existence du facteur de cyclisation (voir la fin de section 2.3) dans la limite $r \rightarrow 0$.

Elle réalise son maximum pour une extension intermédiaire calculable analytiquement. Aux faibles longueurs $L \ll A_{\text{eff}}$ la distribution est de plus en plus piquée, autour d'un étirement quasi-maximal. Aux grandes longueurs $L \gg A_{\text{eff}}$ le comportement gaussien est également retrouvé. Toutefois, cette expression ne permet pas une évaluation convenable du facteur de cyclisation

$$J(L) = Q(\vec{0}; L) = \left(\frac{3}{4\pi A_{\text{eff}} L}\right)^{3/2} \underbrace{\left[1 + \frac{4}{L/A_{\text{eff}}} + \frac{20/3}{(L/A_{\text{eff}})^2}\right]^{-1}}_{\text{déviation (et non corrections!) au modèle gaussien}}.$$

La méthode esquissée ici est celle que Ha, Thirumalai, Bhattacharjee et Bryngelson ont proposée et développée il y a environ dix ans [HT95, TH97, BTB97]. Elle a permis la synthèse de nombreux résultats issus de raisonnements divers, sur la relaxation de l'inextensibilité, autour du principe de minimisation de l'énergie libre du système étudié : un WLC d'inextensibilité *effective*. C'est dans ce sens

qu'elle constitue pour ce modèle une approximation du type *champ moyen*, devenant exacte en dimension D infinie. Elle est valable quelque soit la longueur du polymère ; « valable » dans la mesure où elle ne diverge pas. L'expression $S(r; L)$ proposée est toutefois obtenue par un *développement supplémentaire* en $1/L$, d'où l'apparition d'un terme gaussien dans l'expression de $J(L)$.⁷

Pour finir, précisons pourquoi nous avons conservé l'inextensibilité locale aux extrémités du polymère. En 1965, Harris et Hearst proposèrent la contrainte globale seule ($\kappa_{\text{extr}} = 0$) comme première relaxation possible [HH66, HHB66, HH67]. En identifiant κ_{glob} à une force d'étirement aux faibles courbures, la définition de la longueur de persistance leur permit d'obtenir κ_{glob}^* . L'énergie contrainte correspondante donne à la fois une variance en distance bout à bout du type marche aléatoire (indépendante de A) et une fonction de corrélation non nulle. Le polymère ainsi défini est bien *semi-flexible* mais son comportement élastique est *pathologique*. Sa fonction de corrélation n'est pas invariante par translation – sauf loin des extrémités $0 \ll s, s' \ll L$ – et l'inhomogénéité correspondante $\lim_{s' \rightarrow s} \langle \vec{\tau}' \cdot \vec{\tau} \rangle = \langle \vec{\tau}^2 \rangle$ est significative. Elle est due aux fluctuations dans la longueur du polymère, qu'autorise la contrainte globale. En 1991 Lagowski, Noolandi et Nickel comprirent que la majeure partie de ces fluctuations pouvait être supprimées en exigeant l'inextensibilité locale des extrémités du polymère [LNN91, ZSZ97]. En interprétant κ_{extr} comme l'énergie de courbure locale en ces extrémités, ils obtinrent κ_{glob}^* et κ_{extr}^* en fonction de la longueur de persistance⁸ en assurant *a posteriori* l'homogénéité $\langle \vec{\tau}^2 \rangle = 1$.

En 1994 Winkler, Reineker et Harnau présentèrent les mêmes conclusions [WRH94] avec une méthode plus systématique, basée sur le principe d'entropie maximale [WR92]. Winkler a récemment (2003) revisité ces résultats [Win03] obtenant pour S , sous la même hypothèse d'un développement en $1/L$, une expression quasi-identique à celle présentée ci-dessus. Seule l'amplitude change de

$$[1 - (r/L)^2]^{-9/2} \text{ en } [1 - (r/L)^2]^{-3/2} \times [2 - (r/L)^2]^{-3}.$$

3.3 Faibles perturbations à l'étirement maximal

En section 2.2 nous avons pu vérifier que la configuration d'étirement maximum $r = L$ avait une énergie élastique nulle. Pour de faibles longueurs $L \ll A$ les effets d'origine entropique sont négligeables, et cette configuration est alors la plus probable. Dans ce cas le WLC est un bâtonnet rigide, trivialement décrit par un vecteur tangent constant $\vec{t}(s) = \vec{e}_z$ (direction choisie arbitrairement). Nous pouvons penser obtenir une expression de $S(r; L)$ aux grandes extensions $r \lesssim L$, en considérant les faibles perturbations à ce bâtonnet, décrites

⁷On prendra garde à ne pas le confondre avec un développement de Daniels (§ 3.1).

⁸S'ils ne définissent pas A_{eff} dans leur article, elle y apparaît de manière naturelle.

par $\vec{t}(s) = [\epsilon(s)\vec{e}_\rho + \vec{e}_z] / \sqrt{1 + \epsilon^2(s)}$ avec $\vec{e}_\rho = \cos \varphi(s)\vec{e}_x + \sin \varphi(s)\vec{e}_y$. Autrement dit, nous considérons des fluctuations à la configuration la plus probable d'un WLC de faible longueur. Une telle expression pourrait donc se révéler pertinente pour des longueurs allant jusqu'à la longueur de persistance.

En 1996, Wilhelm et Frey obtinrent une telle expression moyennant une approximation harmonique⁹ [WF96, FKWS98, FKW98]. Leur résultat s'accorde quantitativement avec les données numériques pour des longueurs allant jusqu'à deux longueurs de persistance, et une échelle d'extension variable selon la longueur considérée : $1 - \frac{L/A}{2} \lesssim r/L \leq 1$. Il s'agit du développement

$$S(r; L) \propto \frac{(r/L)^2}{L} \mathcal{W}(r; L) \left\{ 1 + \sum_{p=2}^{\infty} (-1)^{p+1} p^2 [\mathcal{W}(r; L)]^{p^2-1} \right\},$$

où la constante de proportionnalité assure la normalisation de la distribution. En fait, des extensions $1 - \frac{L/A}{2} \lesssim r/L \lesssim 1 - \frac{L/A}{5}$ produisent une convergence rapide de la série qui est alors essentiellement dominée par le poids de Boltzmann

$$\mathcal{W}(r; L) = \exp[-\beta f_{\text{Euler}}(L - r)] = \exp\left[-\frac{\pi^2}{L/A}(1 - r/L)\right].$$

Ce dernier correspond au travail approximatif fourni par une force f_{Euler} lors de la courbure du bâtonnet rigide, en une configuration d'extension $r \leq L$. Cette force est bien connue en élasticité. C'est la valeur critique caractérisant l'apparition de l'instabilité d'Euler [LL90], lors de la compression longitudinale d'une fine barre élastique d'extrémités libres. En dessous de cette valeur la configuration rectiligne de la barre est stable si elle est soumise à de faibles perturbations. Au delà, elle est instable : la barre ne redevient pas rectiligne, mais au contraire se courbe¹⁰ atteignant ainsi sa nouvelle configuration stable (figure 3.1).

Pour des extensions $1 - \frac{L/A}{5} \lesssim r/L \leq 1$, ce développement reste valable. Toutefois, plus l'extension est grande et moins la série converge rapidement. Dans ce cas on préférera l'expression

$$S(r; L) \propto \frac{(r/L)^2}{L} \frac{1}{[\eta(r; L)]^{3/2}} \sum_{q=1}^{\infty} \left[\frac{(2q-1)^2}{\eta(r; L)} - 2 \right] \exp\left[-\frac{(2q-1)^2}{4\eta(r; L)}\right],$$

avec $\eta(r; L) = \frac{1-r/L}{L/A}$, dont la contribution principale provient du terme $q = 1$.

Notons finalement que ce développement est moins systématique que celui proposé par Shimada et Yamakawa (§ 2.4), bien qu'ils partagent le même principe. Il ne considère en effet que les fluctuations à la configuration totalement étirée. Son résultat est alors totalement analytique, mais son domaine de validité ne le rend pas utile pour exprimer le facteur de cyclisation.

⁹Toujours dans le but de se ramener à des intégrales gaussiennes.

¹⁰On parle aussi de flambage.

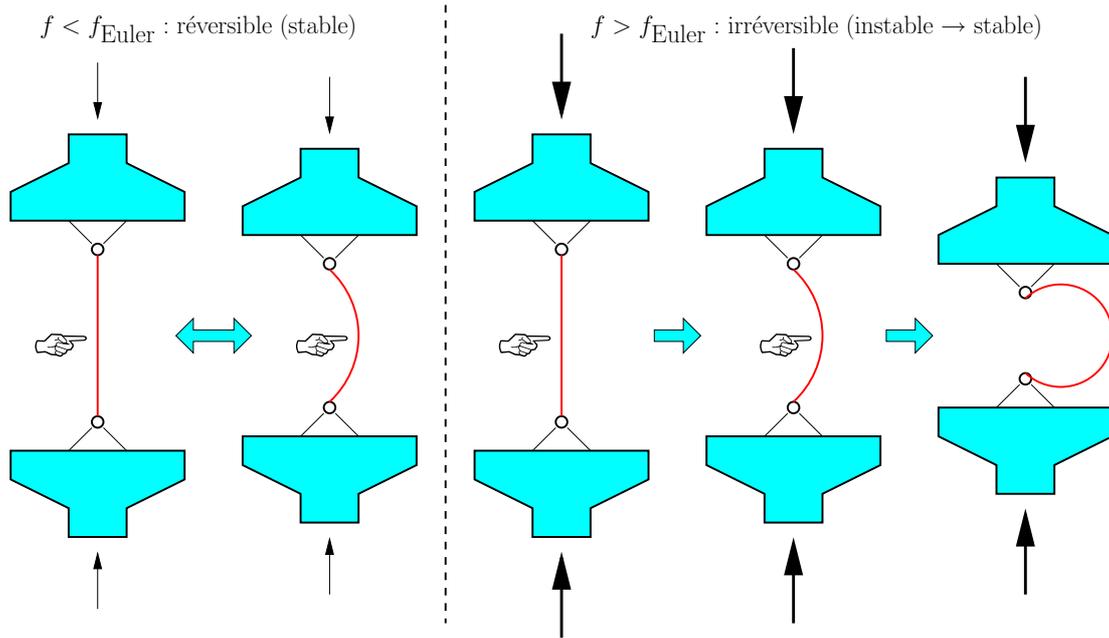


FIG. 3.1 – Instabilité d’Euler, ou flambage d’une fine barre élastique sous l’effet d’une compression longitudinale. Les extrémités de la barre sont montées sur pivots, donc libres. La situation est ainsi *analogue* à la recherche des orientations initiale et finale minimisant l’énergie élastique du WLC (point col, § 2.2). L’effet d’une perturbation de la barre diffère selon que la compression excède ou non la force critique $f_{\text{Euler}} = \frac{A}{\beta} \left(\frac{\pi}{L}\right)^2$.

3.4 Bien d’autres pistes

Nous avons esquissé les principes de quatre méthodes d’approximation au modèle du WLC. Ce sont toutes des développements perturbatifs aux régimes de grande et faible longueurs (§§ 3.1 ,3.2 et §§ 3.3, 2.4 respectivement). La variété d’angle sous lequel le problème est abordé produit toutefois des expressions de forme, et de domaine de validité divers. Malgré leurs imperfections respectives elles ont rencontrées une certaine popularité ces dix dernières années auprès des scientifiques s’intéressant aux propriétés élastiques de l’ADN. Si un physicien décide de travailler sur ce sujet, ce n’est toutefois pas quatre mais une *collection impressionnante* de méthodes qu’il doit s’attendre à trouver dans la littérature. Cette richesse ne rend pas toujours intuitif leurs comparaisons, et c’est pourquoi le choix de « la » méthode la plus pertinente¹¹ n’est pas toujours évident. Toutefois, pour ne pas donner une vue trop partielle du problème nous énumérons ci-dessous quelques unes de ces alternatives, toutes aussi respectables que les quatre précédentes.

Ainsi, nous pouvons penser reconstruire la distribution en extension r du WLC à partir de ses moments $\langle r^{2n} \rangle$. Par définition, la fonction caractéristique s’ex-

¹¹Validité, forme et adaptabilité de l’expression, compétences calculatoires requises. . .

prime alors comme une série sur ces moments, dont la transformée de Fourier (inverse) redonne la distribution initiale. Nous pouvons exprimer exactement les premiers moments (§ 2.1) mais les ordres supérieurs ne sont accessibles que par récurrence, comme développements en puissances de la longueur réduite L/A [HU52, Yam73, YF74, NMF78]. La limite $r \lesssim L \ll A$ montre alors la non pertinence du développement dans ces conditions. Autrement, la forme du hamiltonien du rotateur rigide sous champ (§ 2.1) suggère le traitement du terme de « force » complexe fixant la distance bout à bout, comme perturbation au moment cinétique du rotateur [Mes95, Sch68, Sak94]. En 2002, Stepanow et Schütz ont exploré cette voie [SS02b, Ste03, Ste05] au moyen de la méthode de la résolvante, réutilisée depuis par Spakowitz et Wang (2004) [SW04, SW05, Spa04]. Un an plus tard, Hamprecht et Kleinert ont publié une relation de récurrence exacte entre les ordres successifs de perturbation, pour les vecteurs et énergies propres du hamiltonien [Kle04, HK05, HJK04]. En identifiant les moments ainsi obtenus avec ceux d'une distribution $S(r; L)$ *ad hoc* ils déterminent les paramètres de cette dernière, permettant alors un accord quantitatif sur une large échelle de longueurs.

En 2003, Marucho et Carri [MC03b, MC03a] ont proposé une fonction caractéristique décrivant exactement les deux régimes limites élastique/entropique et vérifiant la contrainte d'inextensibilité exacte. Le comportement aux longueurs intermédiaires est alors ajusté au moyen du moment $\langle r^2 \rangle$ exact, puis jugé *a posteriori* en considérant par exemple la distribution en extension r . L'expression proposée est en fait obtenue par approximations successives de la fonction caractéristique du WLC. Sa transformée de Fourier est réécrite comme une transformée de Borel généralisée, que les auteurs ont employé lors de leurs travaux antérieurs. Bien que peu intuitif, le raisonnement ne présente pas de difficultés majeures. Il en revanche assez calculatoire : il utilise un calcul variationnel mené sur une « ansatz », divers développements limités permettant d'exprimer les intégrales nécessaires comme fonctions spéciales¹² et la transformée de Borel. Les auteurs ont synthétisé leur résultat en termes plus généraux d'une fonction génératrice, mais cela reste très formel [CM04, Car04]. C'est aussi le cas de la généralisation du développement de Wilhelm et Frey, proposée en 2005 par Kleinert et Chervyakov [KC05] au moyen de développements diagrammatiques, n'apportant que peu d'amélioration selon les auteurs eux-mêmes.

Citant Marucho et Carri, le lecteur peut à ce niveau se demander pourquoi il a besoin d'autant de méthodes d'approximation pour résoudre *un seul* modèle. Ne peut-on pas utiliser *la même* méthode pour étudier toutes les propriétés du WLC, complétée par *un* développement perturbatif *systématique* ? La réponse est malheureusement négative pour le moment. Peut-être doit on alors changer

¹²C'est l'occasion pour le lecteur de réviser ses fonctions : de Bessel, gamma, hypergéométriques, de Kummer, de cylindre parabolique, de Whittaker, etc.

de modèle... Si le WLC décrit un comportement élastique de l'ADN en accord avec les nombreuses expériences de molécule unique montées ces dix dernières années, il n'a pas *a priori* de justification microscopique. Depuis 1989, Kholodenko a ainsi fait la promotion de la *chaîne de Dirac* [Kho93, KV94, KBAG98]. Ce modèle d'élasticité de l'ADN n'est pas plus justifiable microscopiquement. Il se résout toutefois de manière exacte et montre un bon accord quantitatif avec la fonction caractéristique du WLC (simulations) et de l'ADN (expériences). Si nous continuons la recherche d'un modèle résoluble, nous devons faire attention à ce que ce dernier s'accorde avec le WLC quelque soit sa longueur. Le groupe de renormalisation nous rappelle en effet que les modèles appartenant à la *classe d'universalité du WLC*, s'ils présentent le même comportement aux grandes longueurs peuvent totalement différer aux faibles longueurs [WN05].¹³

Pour conclure, il existe d'autres travaux proposant des expressions de la distribution en distance bout à bout ou du facteur de cyclisation, valables sur des échelles de longueurs couvrant plusieurs des approximations précédemment citées. Ce sont en fait des combinaisons de ces dernières nécessitant le plus souvent l'ajustement d'un ou plusieurs paramètres. De ce fait nous ne décrivons pas ces dernières dans ce travail.¹⁴ Citons toutefois la formule proposée en 1999 par Ringrose, Chabanis, Angrand, Woodroffe et Stewart [RCA⁺99] pour le facteur de cyclisation. Il s'agit d'une unique expression analytique visant à l'interpolation des résultats du polymère gaussien d'une part (§ 2.3), et de Shimada et Yamakawa¹⁵ d'autre part (§ 2.4). Elle ne dépend d'aucun paramètre ajustable dans sa formulation initiale. En 2001, Rippe [Rip01, DRDK02, SLR00] en proposa sa généralisation non pas à la distribution en distance bout à bout mais à une quantité analogue, importante pour la suite de ce travail (§§ 4.3 et 5.4) et par ailleurs déjà introduite pour le polymère gaussien en fin de section 2.3

$$\text{Prob}(r; L) / \left(\frac{4}{3} \pi r^3 \right) = J_{\text{gauss}}(L) \times \exp \left[- \frac{8 - p}{(L/A)^2 + p} \right]. \quad (3.1)$$

En ajustant le paramètre p par comparaison à des données numériques (voir expérimentales) cette expression permet *notamment* de prendre en compte un rayon de boule $r \geq 0$ (§ 4.3). Si $r = 0$, $p = 0$ et l'on retrouve la formule proposée par Ringrose et coll. pour le facteur de cyclisation. Nous désignons (3.1) dans la suite de ce travail, comme la formule de Rippe et Ringrose (R&R). Nous retenons qu'elle n'est, par principe, pas adaptable de manière *systématique* afin,

¹³Non pertinence des détails microscopiques au-dessus d'une longueur de lissage mésoscopique, ou de « coarse graining » en anglais.

¹⁴Pas plus que nous ne tenterons de faire un bilan exhaustif des diverses simulations numériques réalisées sur ce sujet.

¹⁵Le résultat de S&Y n'est pas à proprement parler inclu : la divergence aux faibles longueurs du $J(L)$ gaussien est supprimée en le modulant par un facteur exponentiel empirique, différent de (2.4) mais reproduisant au mieux le J correspondant.

par exemple, de prendre en compte une extension non nulle ou une cassure à mi-longueur (§ 5.4) : elle nécessite l'introduction d'au moins un paramètre ajustable, ce qui réduit son intérêt théorique.

Pour les lecteurs perdus Rippe a publié deux revues : avec van Hippel et Langowski [RvHL95] en 1995, puis seul [Rip01] en 2001. Leurs contenus sont assez pédagogiques, formulant les problématiques tant au niveau expérimental que théorique (ou numérique). Ces travaux tentent également de prendre en compte la variété de phénomènes biologiques observés *in vivo* lors d'une cyclisation d'ADN.

Chapitre 4

Fermeture de la boucle : effets dus à la taille finie des protéines

Le calcul de la distribution en distance bout à bout d'un WLC de longueur quelconque n'est pas réalisable analytiquement. Pour estimer sa valeur particulière à extension nulle (configurations de boucle), appelée facteur de cyclisation, nous procédons donc numériquement. L'algorithme choisit permet en outre d'évaluer les effets d'une non fermeture rigoureuse de la « boucle », reflet de la taille finie des protéines permettant une cyclisation de l'ADN, impliquée dans une régulation de l'expression génétique à l'étape de transcription. Bien que cette étude soit menée à force extérieure nulle, la relation de force/extension du WLC est tout d'abord traitée. Cette « digression » permet de se faire une meilleure idée du comportement élastique du modèle, et par conséquent d'une molécule d'ADN (pour des forces inférieures à 10 pN). Mais il existe surtout une analogie entre ce problème et le calcul de la distribution en extension, selon la direction d'étirement du polymère, que nous comptons utiliser pour notre évaluation numérique du facteur de cyclisation. Enfin cette relation de force/extension ne se résout exactement que de manière numérique, validant *a priori* notre approche du phénomène de cyclisation.

4.1 Relation de « force/extension » : une digression ?

Dans les dispositifs expérimentaux typiques permettant la manipulation de molécules biologiques uniques (un brin d'ADN par exemple), la molécule est contrôlée au moyen d'une force appliquée par l'opérateur (médaillon de la figure 4.1) [BMW00, SDC⁺03]. Pour étudier une configuration donnée de la molécule, cette force extérieure est maintenue uniforme. Elle correspond au terme $f\vec{e}_z$ que nous avons inclus dans l'énergie du WLC (§ 2.1). La relation de force/extension détermine comment la force modifie l'extension moyenne du polymère, dans la

direction privilégiée \vec{e}_z de l'étirement.¹ La fonction de partition du WLC montre que l'extension moyenne sous force $\langle \vec{R}[\vec{t}; L, f] \cdot \vec{e}_z \rangle = \langle z \rangle$ se calcule sans connaissance explicite de la distribution correspondante, par

$$\frac{\langle z \rangle}{L} = \frac{\partial \ln Z(\Omega_L | \Omega_0; L, f)}{\partial(\beta L f)}.$$

Expérimentalement, l'orientation des extrémités du brin d'ADN est contrainte : l'une est attachée au plan de travail et l'autre à une bille, sensible au champ magnétique responsable de la force. C'est donc bien le propagateur de Ω_0 à Ω_L qu'il faut considérer, et non l'amplitude du vide obtenue par sommation des orientations limites. Nous savons que le problème est résolu si le spectre du hamiltonien du rotateur rigide sous champ est connu, mais sa diagonalisation est impossible analytiquement (§ 2.1). Pour des longueurs $L \gg A$, situation la plus fréquente expérimentalement, seul l'état fondamental (non dégénéré) $H_g(f)$ et $\langle \Omega | \Psi_g(f) \rangle = \Psi_g(\Omega; f)$ est toutefois nécessaire

$$Z(\Omega_L | \Omega_0; L, f) \simeq \Psi_g(\Omega_L; f) \tilde{\Psi}_g(\Omega_0; f) e^{-\frac{L}{A} H_g(f)} \quad \text{donc} \quad \frac{\langle z \rangle}{L} \simeq -\frac{\partial H_g(f)}{\partial(\beta A f)}.$$

L'étude de l'étirement dans la limite élastique $L \ll A$ n'est en effet pas instructive : avec ou sans force, l'ADN est *a priori* quasi-complètement étiré.² La théorie des perturbations stationnaires [Mer97] permet de réécrire

$$\frac{\langle z \rangle}{L} \simeq -\langle \Psi_g(f) | \frac{\partial \hat{H}(f)}{\partial(\beta A f)} | \Psi_g(f) \rangle = \sum_{\ell} \ell \frac{\psi_{\ell-1}(f) \tilde{\psi}_{\ell}(f) + \psi_{\ell}(f) \tilde{\psi}_{\ell-1}(f)}{\sqrt{(2\ell-1)(2\ell+1)}}$$

exprimée en fonction des composantes $\psi_{\ell}(f) = \langle \ell, 0 | \Psi_g(f) \rangle$. Seules les composantes $m = 0$ sont *a priori* non nulles, car le fondamental possède la symétrie cylindrique (§ B.2). Ce dernier n'est pas calculable analytiquement mais estimable numériquement [PTVF02, MS95, BWA⁺99], d'autant plus facilement que le hamiltonien est tridiagonal et symétrique. Ce dernier est tronqué tel que le résultat est stable, condition peu coûteuse en terme de temps de calcul : nous avons utilisé $\ell_{\max} = 100$, mais une valeur bien inférieure fait tout aussi bien l'affaire. En itérant la procédure pour diverses forces, nous obtenons la courbe de force/extension présentée en figure 4.1.

En 1994, Marko et Siggia [BMS94, MS95] ont proposé une relation de force/extension analytique s'ajustant notamment aux mesures expérimentales de Smith, Finzi et Bustamante [SFB92]. Cette relation reproduit exactement les limites de faible et grand étirements. Elle interpole le régime intermédiaire avec environ 10% d'écart à la solution exacte, que nous venons de déterminer (figure 4.1).

¹Qui est nulle en l'absence de force, par isotropie (la limitation à $z > 0$ par le substrat n'est pas formellement considérée).

²Le comportement aux longueurs intermédiaires $L \simeq A$ est repris au chapitre 6.

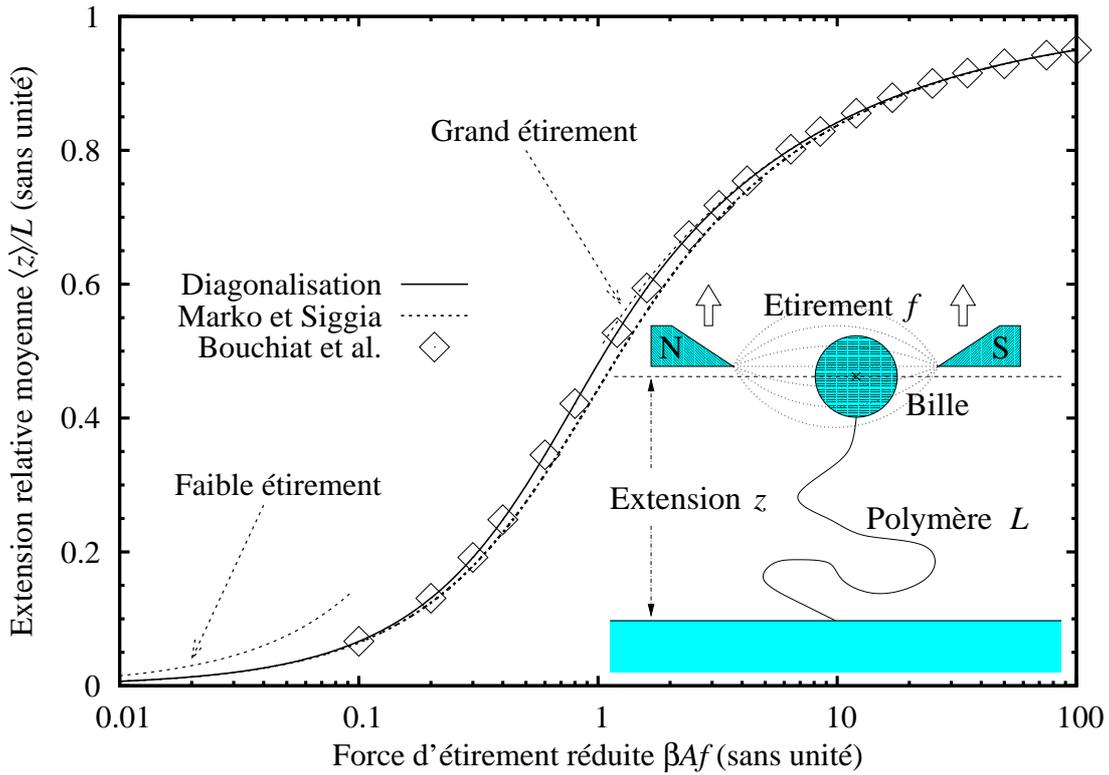


FIG. 4.1 – La relation de force/extension spécifie comment une force extérieure $f\vec{e}_z$ modifie l’extension moyenne $\langle z \rangle$ d’un polymère, dans la direction d’étirement. La formule proposée par Marko et Siggia (pointillés) décrit exactement les limites de faible et de grand étirement. Pour des force intermédiaires elle interpole la solution exacte du WLC (ligne pleine) avec moins de 10% d’écart. Notre diagonalisation numérique du hamiltonien du rotateur rigide sous champ, reproduit bien les données obtenues par Bouchiat (symboles \diamond) par résolution numérique de l’équation de Schrödinger correspondante.

Elle s’écrit

$$\beta A f = \frac{\langle z \rangle}{L} + \frac{1/4}{(1 - \langle z \rangle / L)^2} - \frac{1}{4} \simeq \begin{cases} \frac{3}{2} \frac{\langle z \rangle}{L} & \frac{\langle z \rangle}{L} \text{ et } \beta A f \ll 1, \\ 1 - \frac{1}{\sqrt{4\beta A f}} & \frac{\langle z \rangle}{L} \lesssim 1 \text{ et } \beta A f \gg 1. \end{cases}$$

Le deux limites s’évaluent de manière analytique, à partir respectivement du modèle gaussien (§ 2.3) et d’un traitement perturbatif du WLC analogue à celui de Wilhelm et Frey (§ 3.3) [MS95]. Nous voyons que le polymère commence à s’étirer sous l’action d’une force typique $\beta A f \simeq 1$, soit moins d’un dixième de piconewton pour un brin d’ADN à température ambiante.³ Le comportement élastique d’un tel ADN est en fait bien décrit par le WLC, pour des forces allant jusqu’à environ 10 pN. Au-delà la molécule présente une extensibilité, nécessitant la relaxation de la contrainte d’inextensibilité du WLC (§ 3.2, voir par exemple

³Rappelons que dans ces conditions $k_B T = 1/\beta = 4.1 \text{ pN} \cdot \text{nm}$.

le travail réalisé par Odijk [Odi95] en 1995). Pour des forces dépassant 50 pN, l'ADN change de configuration provoquant une élongation subite d'environ 70%, mais c'est une autre histoire [Mar98, LRS⁺99, SN03, CYL⁺04].

Pour finir, d'autres formules interpolatives ont été proposées depuis le travail de Marko et Siggia. Elle demeure toutefois une bonne « première approximation », à laquelle des corrections peuvent être apportées d'une manière systématique. Voir par exemple le travail de Bouchiat en 1999 exploitant les données de Wang, Allemand, Strick, Block et Croquette [BWA⁺99] : l'écart au WLC y est réduit à moins de 0.01% (figure 4.1). L'obtention d'une relation de force/extension du WLC, analytique et précise pour une large échelle de force ne constitue pas qu'un challenge théorique. Une méthode permettant de déterminer les longueurs totale et de persistance d'un brin d'ADN, consiste à ajuster une telle relation aux données expérimentales d'étirement de la molécule.

4.2 Un calcul indirect et numérique du facteur de cyclisation J

4.2.1 Principe de l'algorithme

Nous proposons une évaluation numérique exacte du facteur de cyclisation du WLC, *en l'absence* de force appliquée par un opérateur extérieur. Nous procédons en plusieurs étapes par détermination successive des diverses distributions en extension du WLC, dans l'esprit de la fin de section 2.3. L'évaluation produite est alors *indirecte*, mais elle permet de vérifier à chaque étape la *fiabilité* de nos résultats, ainsi que d'estimer l'importance d'une *fermeture exacte* des configurations de boucles. Son principe exploite :

- l'analogie avec la relation de force/extension, résolue par diagonalisation ;
- l'isotropie du problème, due à sa symétrie sphérique en l'absence de force.

La distribution *marginale* en extension z – ne contenant donc aucune information statistique sur les composantes x et y de la distance bout à bout⁴ – s'exprime comme le propagateur contraint ($f = 0$)

$$P(z; L, 0) = \frac{1}{Z(L, 0)} \int \mathcal{D}\vec{t} \delta \left[z - \int_0^L \cos \theta(s) ds \right] \exp \left\{ -\beta E \left[\vec{t}; L, \vec{0} \right] \right\}.$$

La représentation spectrale de la distribution de Dirac (§ D.2) réduit le calcul de la distribution à celui de la TF du propagateur d'un WLC (§ 2.1), sous *force complexe* (ik_z/β) \vec{e}_z

$$P(z; L, 0) = \frac{1}{Z(L, 0)} \int_{-\infty}^{\infty} \frac{dk_z}{2\pi} e^{ik_z z} Z(L, -ik_z/\beta).$$

⁴Voir l'exemple pédagogique de la chaîne gaussienne (§ 2.3).

Ces expressions sont des *cas particuliers* de la distribution en distance bout à bout \vec{r} écrite en début du chapitre 3. Tout d'abord nous sommes ici à force extérieure nulle $f = 0$. Enfin nous calculons une distribution marginale, obtenue par sommation des composantes x et y de \vec{r} . Alternativement, cela se traduit par une force complexe \vec{k} (multiplicateur de Lagrange fixant l'extension) selon \vec{e}_z , c'est-à-dire de composantes $k_x = k_y = 0$. Pour la suite nous omettons le plus souvent la dépendance explicite des diverses distributions en f , quand cette dernière est nulle.

Nous avons vu dans la précédente section que la relation de force/extension du WLC se dérive de sa fonction de partition, calculable par diagonalisation numérique du hamiltonien du rotateur rigide sous champ. Ne reste donc plus qu'à réaliser sa TF? Deux remarques s'imposent :

- seules les configurations flexibles $L \gg A$ ont été étudiées. Ainsi, seul l'état fondamental du hamiltonien a été calculé ;
- la « force » fixant l'extension est complexe. Son traitement ne peut pas être identique à celui d'une vraie force, appliquée par un opérateur extérieur.

À terme, nous voulons modéliser des boucles d'ADN dont la longueur est de l'ordre de celle de persistance. Nous pouvons toutefois obtenir le spectre entier du hamiltonien⁵ sans grand effort supplémentaire, ne nous interdisant ainsi aucune longueur particulière de polymère. Il existe en effet des méthodes numériques *robustes* permettant la *diagonalisation* (valeurs et vecteurs propres) des matrices réelles, tridiagonales et symétriques [PTVF02]. Dans le cas complexe, c'est l'hermiticité et non la symétrie que requièrent de telles méthodes. Le hamiltonien correspondant à la force complexe, fixant l'extension dans le calcul de la distribution $P(z; L)$, est symétrique mais pas hermitien. Une alternative possible est le calcul direct de son exponentielle, intervenant dans le propagateur.

4.2.2 Exponentielle de matrice

L'action d'une *exponentielle de matrice* sur un vecteur $|u\rangle$ particulier

$$|v(t)\rangle = e^{-t\hat{A}} |u\rangle \quad \text{est solution de} \quad \frac{d}{dt}|v(t)\rangle + \hat{A}|v(t)\rangle = 0, \quad |v(0)\rangle = |u\rangle,$$

et nous n'avons donc plus qu'à choisir une des nombreuses bibliothèques permettant le calcul d'un tel système d'équations différentielles ordinaires. Ces bibliothèques sont largement déboguées et rompues à une grande diversité de systèmes (grand nombre d'équations, coefficients non constants, non linéarité, etc.). Le nôtre n'est d'ailleurs pas trop complexe : le système est linéaire et de coefficients constants. Bien que précise, une telle méthode est toutefois à proscrire de par son inefficacité en termes de temps de calcul, et ce même en simplifiant les algorithmes déjà

⁵En contrôlant l'erreur liée à sa troncature inévitable.

existant.⁶ L'idée de calculer l'action de l'exponentielle de matrice sur un vecteur particulier, plutôt que ses composantes est en revanche à retenir. Par exemple \hat{A} peut être creuse sans que son exponentielle ne le soit pour autant : mieux vaut alors éviter son calcul. Une autre définition fiable d'une exponentielle de matrice est son développement de Taylor

$$|v(t)\rangle = |u\rangle + (-t\hat{A})|u\rangle + \frac{1}{2}(-t\hat{A})^2|u\rangle + \dots$$

dont la convergence est assurée. En pratique nous devons faire face aux erreurs d'arrondis, et aussi décider d'un critère de convergence de la somme. Il s'agit de deux troncatures, mais d'origine différente. La première, numérique, est par exemple déterminante si deux termes successifs de la somme sont du même ordre de grandeur mais de signes opposés. L'augmentation de la précision numérique, ou un regroupement pertinent des termes⁷ peuvent résoudre ce problème, bien qu'ils ne soient pas systématiquement possibles. Gardant cela à l'esprit, nous trouvons dans la littérature mathématique plusieurs critères de convergence précisant quand tronquer la somme, mais leur efficacité en termes de temps de calcul n'a là encore rien de systématique. L'exemple du développement de Taylor illustre bien la situation dans laquelle le sujet s'est longtemps trouvé : en théorie, les manières d'attaquer le calcul ne manquent pas, mais en pratique presque aucune de ces méthodes n'est systématiquement implémentable. Leur stabilité dépend fortement du *conditionnement* de la matrice⁸, elles ne tiennent pas compte des *erreurs d'arrondis* ou sont encore inefficaces en termes de *temps de calcul*.

Réaliser une *approximation de Padé* sur l'exponentielle de matrice apparaît comme un bon compromis, à condition toutefois que la matrice soit bien conditionnée. L'approximation de Padé d'ordre (p, q) d'une fonction de matrice (pour nous une exponentielle) se définit comme pour une fonction usuelle : c'est une autre fonction rationnelle⁹ dont le développement de Taylor coïncide avec celui de la fonction originelle jusqu'à l'ordre $(p + q)$. Le conditionnement de la matrice peut s'assurer par

$$\exp(-t\hat{A}) = \left[\exp\left(\frac{-t}{2^\lambda}\hat{A}\right) \right]^{2^\lambda} \quad \lambda \text{ entier positif ou nul.}$$

On parle de « *scaling and squaring* » en anglais. Une fois déterminé le λ permettant le conditionnement, l'approximation de Padé se réalise sur l'exponentielle de la matrice *mise à l'échelle*. La *quadrature* du résultat est ensuite réalisée par

⁶La tendance actuelle est même parfois inverse : en calculant l'exponentielle de matrice, par exemple avec l'algorithme présenté plus bas, on résout le système d'équations différentielles.

⁷Les négatifs d'un côté, les positifs de l'autre par exemple.

⁸Une matrice « peu conditionnée » peut avoir un spectre très variable lorsqu'on perturbe, même faiblement, un ou plusieurs de ses éléments. C'est pour cette raison qu'une matrice quelconque n'est pas systématiquement diagonalisable numériquement.

⁹C'est-à-dire le rapport de deux polynômes, l'un de degré p l'autre q .

multiplications successives. Cette méthode peut paraître lourde mais à titre de comparaison elle permet d'obtenir une meilleure précision que la résolution du système d'équation différentielles mentionnée au début, et ce pour un temps de calcul inférieur. Néanmoins, pour que ce dernier reste raisonnable nous devons nous cantonner aux petites matrices. Notons que nous accédons aux composantes de l'exponentielle de matrice, et non à son action sur un vecteur particulier.

Pour de grandes matrices nous pouvons en fait réduire le problème au calcul de l'exponentielle d'une matrice de taille moindre, en contrôlant l'erreur ainsi commise, pour laquelle nous utilisons la méthode précédente. Cette approche est d'autant plus pertinente que la matrice originelle \hat{A} est creuse, et que c'est l'action de son exponentielle sur un vecteur particulier qui nous intéresse. Supposons que nous disposons d'une base orthonormée $|m\rangle$ avec $m = 1, \dots, M < N$ la taille de \hat{A} , telle que $|1\rangle = |u\rangle$.¹⁰ Le projecteur $\hat{P} = \sum_{m=1}^M |m\rangle\langle m|$ (orthonormal) sur le sous-espace définit par cette base, permet d'écrire [HH05, HL97]

$$|v(t)\rangle = \hat{P} \left[\hat{P}^\dagger e^{-t\hat{A}} \hat{P} \right] \underbrace{\hat{P}^\dagger |u\rangle}_{=|1\rangle} \simeq \hat{P} \exp\left(-t \hat{P}^\dagger \hat{A} \hat{P}\right) |1\rangle$$

où l'on reconnaît la restriction $\hat{P}^\dagger \hat{A} \hat{P}$ de \hat{A} au sous-espace de dimension M . Remarquablement la précision du calcul peut être conservée même en diminuant significativement la dimension M par rapport à la taille N .

Étonnamment, un grand nombre de physiciens déconsidère le calcul direct d'exponentielles de matrice, jugé peu précis et élégant par rapport à une diagonalisation préalable de la matrice. Depuis plusieurs années des bibliothèques permettant de tels calculs *contrôlés*, sont pourtant disponibles. C'est, dans certaines mesures, le cas de la plupart des logiciels de calcul formel tels **Mathematica** ou **Matlab**¹¹. Nous avons choisit la bibliothèque **Expokit**, écrite en **Fortran** et dont les sources sont disponibles gratuitement sur internet [Sid98]. L'algorithme décrit ci-dessus est en fait un survol du principe de ses routines. C'est en particulier l'une des seules bibliothèques adressant systématiquement les problèmes d'arrondis, de temps de calcul (nous avons vu comment traiter optimalement de petites matrices et intelligemment certaines grandes matrices) et de stabilité selon les propriétés de la matrice. Elle a été récemment (2003) plébiscitée par Moler et Van Loan, s'intéressant au problème depuis presque 30 ans. Nous recommandons d'ailleurs au lecteur la revue écrite par ces auteurs [MvL03].

¹⁰Une telle base peut typiquement se construire en tridiagonalisant la matrice, par un algorithme de Arnoldi ou de Lanczos.

¹¹Mais aussi **MuPad**, **Scilab**, etc., pour n'oublier personne.

4.2.3 Distributions P et S : discussion et utilisation

Pour revenir au problème qui nous intéresse ici, nous calculons donc grâce à Expokit les composantes de

$$\exp\left[-\frac{L}{A} \widehat{H}(-ik_z/\beta)\right] |0, 0\rangle$$

dans la base des harmoniques sphériques $|\ell, m\rangle$. Pour chaque valeur de k_z nous ne retenons alors que la première composante $|0, 0\rangle$, qui fournit en effet l'amplitude du vide $Z(L; -ik_z/\beta)$ recherchée (§ 2.1). Comme pour la relation de force/extension (bien que les algorithmes ne soient pas les mêmes) nous devons en pratique tronquer le hamiltonien. Afin de minimiser l'erreur liée à la coupure ℓ_{\max} nous procédons pour chaque longueur réduite au test de convergence suivant : partant d'une valeur arbitraire (ni trop petite, ni trop coûteuse en temps de calcul : p. ex. $\ell_{\max} = 10$) la coupure est graduellement augmentée jusqu'à ce que son effet ne soit plus observable sur le résultat. Une ultime vérification peut être faite en considérant une coupure largement plus grande que la dernière retenue. Notons qu'il est plus commode d'observer cet effet sur $P(z; L)$ que sur $Z(L, -ik_z/\beta)$ (l'étape de TF est décrite plus bas). En considérant des longueurs de plus en plus faibles, nous avons constaté que la coupure devait être significativement augmentée. En couvrant l'échelle de longueurs $0.65 \leq L/A \leq 20$, c'est-à-dire $100 \lesssim L \lesssim 3000$ bp, le compromis¹² suivant s'est dégagé : une coupure de $\ell_{\max} = 50$ suffit amplement pour la majeure partie $L \gtrsim A$, tandis qu'aux plus faibles longueurs $L \ll A$ elle peut être augmentée par sécurité jusqu'à $\ell_{\max} = 100$.¹³ Pour terminer, signalons que l'algorithme utilisant Expokit s'est montré systématiquement stable. Cela ne doit pas nous étonner vu les propriétés du hamiltonien (tridiagonalité et symétrie).

Concernant l'échantillonnage en k_z , il doit se faire en vue d'une *transformée de Fourier rapide* ou FFT [PTVF02], de l'anglais « Fast Fourier Transform », dont le principe est détaillé en annexe F. Le lecteur retiendra que l'inextensibilité du WLC limite l'amplitude du vide $Z(L, -ik_z/\beta)$ en bande, c'est-à-dire que sa TF est telle que $P(|z| > L; L) = 0$. Cette propriété permet une évaluation convenable de la TF au moyen de l'algorithme de FFT à condition d'échantillonner Z au pire, c'est-à-dire au maximum avec un pas π/L (constant), fixé par la *fréquence critique* $z_c = L$ (§ F.1). L'annexe F montre en effet comment, lors d'une FFT la partie $|z| > z_c$ du résultat $P(z; L)$ se retrouve *repliée* sur sa partie $|z| \leq z_c$.¹⁴

¹²Si cela n'est pas trop coûteux en temps de calcul, nous aimerions ne pas avoir à changer ℓ_{\max} pour chaque longueur (l'idéal serait de pouvoir n'utiliser qu'une valeur de ℓ_{\max}).

¹³En fait $\ell_{\max} = 50$ permet parfaitement de descendre à $L/A \geq 0.65$. Nous citons $\ell_{\max} = 100$ en vue du chapitre 5 où nous diminuons encore L .

¹⁴Ce phénomène de repliement est bien connu et détaillé dans la littérature portant sur la FFT. Il est pourtant encore responsable de nombreuses évaluations numériques erronées de TF. Nous encourageons donc le lecteur à lire l'annexe F.

C'est pourquoi le choix minimum $z_c = L$ permet une évaluation convenable de la TF de $Z(L, -ik_z/\beta)$. Notons qu'il est en pratique d'usage de doubler la fréquence critique. Cette sécurité supplémentaire constitue un coût numérique¹⁵ négligeable pour la FFT, mais pas pour le calcul de l'exponentielle de matrice qui reste l'étape limitante de notre algorithme. Nous n'avons en fait constaté sur $P(z; L)$ aucune différence entre l'utilisation de L ou $2L$ (ce quelque soit la longueur) : cette sécurité s'est donc révélée superflue. Dans les conditions d'utilisation d'Expokit précisées ci-dessus, nous avons pu considérer un échantillonnage sur 4096 points, soient 2048 points pour la partie $0 \leq z \leq L$.¹⁶ Précisons pour finir qu'une utilisation optimale de l'algorithme utilise les propriétés suivantes de $P(z; L)$, qui est réelle et paire en z .

Nous avons ainsi pu calculer la distribution exacte en extension z du WLC, *sans limitation*¹⁷ dans la longueur L du polymère (figure 4.2, gauche). Elles sont

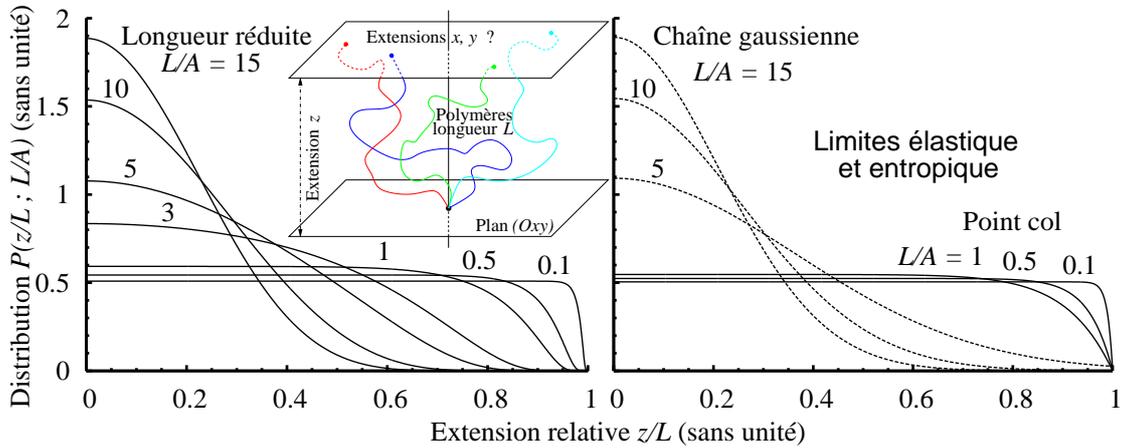


FIG. 4.2 – Calcul numérique exact de la distribution *marginale* en extension z du WLC, en l'absence de force (figure gauche). Son comportement *entropique* se manifeste dès cinq longueurs de persistance. Chaîne *gaussienne* et WLC s'accordent *quantitativement* aux plus grandes longueurs $L/A \gtrsim 10$ (figure droite, pointillés). La *rigidité* élastique se manifeste en dessous d'une longueur de persistance. Le *point col* du WLC permet une description *qualitative*, d'autant plus pertinente que la longueur est faible $L \ll A$ (figure droite, lignes pleines).

paire en z et montrent toutes un maximum en $z = 0$, reflet de l'*isotropie* du problème. Elles s'annulent de manière continue en atteignant l'*étirement maximal* $z \lesssim L$, y compris pour de faibles longueurs $L \ll A$. Dans ce cas particulier $P(z; L)$ tend vers la distribution uniforme, correspondant à la limite du *bâtonnet rigide* : $S(r; L) = 4\pi r^2 \times \delta(r - L)/(4\pi L^2)$. Une estimation *qualitative* est possible à partir du point col (§ 2.2), en négligeant les fluctuations d'origine

¹⁵Conserver le même pas en z , qu'avec $z_c = L$, requiert de doubler le nombre de points.

¹⁶Nous sommes restreints aux puissances de deux pour la FFT (§ F.1).

¹⁷Nous descendons en dessous de la longueur de persistance.

entropique (§ 2.4).¹⁸ Nous reconstruisons

$$\begin{aligned} P(z; L) &= \int \delta(r \cos \vartheta - z) Q(\vec{r}, L) d\vec{r} \\ &= \int_0^\infty dr \int_{-1}^1 d(\cos \vartheta) \frac{1}{2} S(r; L) \frac{1}{r} \delta(\cos \vartheta - z/r) \\ &= \frac{1}{2} \int_{|z|}^L \frac{S(r; L)}{r} dr. \end{aligned}$$

Le caractère qualitatif de l'estimation n'empêche toutefois pas de constater clairement la *domination* de la configuration de point col aux faibles longueurs (figure 4.2, droite/lignes pleines). D'autre part, le comportement *gaussien* (§ 2.3) est déjà *manifeste* pour des longueurs approchant cinq fois celle de persistance. Les deux modèles – WLC et chaîne gaussienne – présentent un accord *quasi-parfait* à partir de dix longueurs de persistance (figure 4.2, droite/pointillés). Pour des longueurs intermédiaires, nos résultats reproduisent *fidèlement* ceux obtenus par Samuel et Sinha [SS02a] en 2002.

La symétrie sphérique du problème (rappelons que $f = 0$) nous permet alors de relier analytiquement $P(z; L)$ à la distribution isotrope en distance bout à bout, c'est-à-dire en extension r (annexe D, ou bien précédente relation)

$$S(r; L) = 4\pi r^2 Q(r \vec{e}_r; L) = -2r \frac{dP}{dz}(r; L).$$

Pour chaque longueur de polymère, $S(r; L)$ s'annule de manière continue en approchant l'étirement maximal $r \lesssim L$, ou l'extension nulle $r \gtrsim 0$ (figure 4.3, haut/gauche). Comme nous allons le voir, le comportement limite pour $r \rightarrow 0$ joue un rôle central dans notre évaluation du facteur de cyclisation. La distribution montre également un unique maximum réalisé pour une extension approchant l'étirement total $r \lesssim L$ aux faibles longueurs de polymère¹⁹, mais s'identifiant au contraire à l'extension la plus probable $\sqrt{4AL/3}$ d'une chaîne gaussienne (§ 2.3), pour des longueurs dépassant dix longueurs de persistance. Le comportement aux longueurs intermédiaires reproduit, là encore les données de Samuel et Sinha [SS02a].

Nous avons également comparé nos résultats au champ moyen développé par Ha et coll. puis reformulé par Winkler (§ 3.2), ainsi qu'au développement de Wilhelm et Frey (§ 3.3) prolongé par Kleinert et Chervyakov (§ 3.4). Ce dernier (W&F) réalise un meilleur accord aux résultats exacts, pour des longueurs n'excédant pas deux longueurs de persistance (figure 4.3, haut/droite). La correction de Kleinert et Chervyakov ne modifie pas visiblement ce comportement. Le champ moyen ne permet qu'une estimation qualitative de la distribution entière : l'accord s'améliore aux grandes longueurs²⁰ mais reste qualitatif (figure 4.3, bas/gauche).

¹⁸ Leur prise en compte par le biais de $C(0; L + 2r)$ ne produit pas d'effet notable.

¹⁹ $S(r; L)$ tend vers la distribution piquée en $r = L$ caractéristique d'un bâtonnet rigide.

²⁰ Rappelons que l'expression simple de S est obtenue par développement en $1/L$.

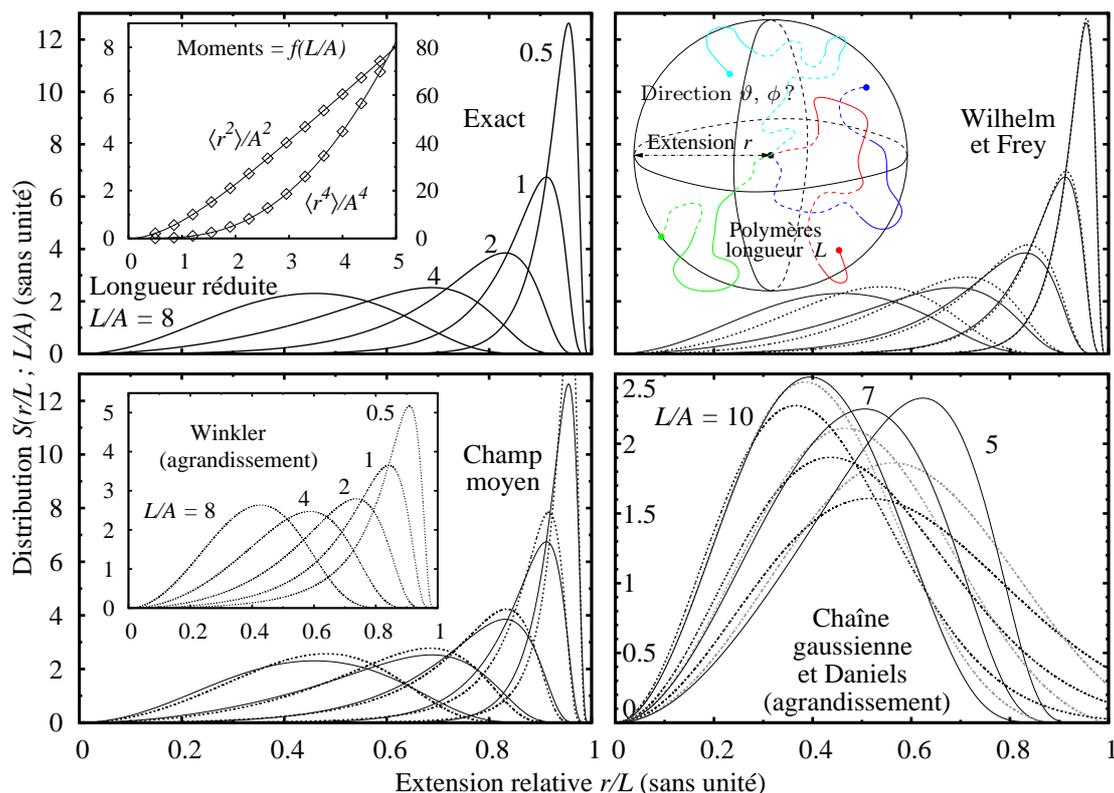


FIG. 4.3 – Calcul numérique exact de la distribution *isotrope*, en extension r , du WLC (figure haut/gauche, lignes pleines) et comparaison de ses deux premiers moments avec les résultats analytiques exacts (médaillon, symboles \diamond et lignes pleines respectivement). Ces résultats sont repris dans les autres figures, en fines lignes pleines tandis que les approximations sont en pointillés. Le développement de Wilhelm et Frey (§ 3.3) décrit les faibles longueurs $L \ll A$, allant jusqu'à deux longueurs de persistance (figure haut/droite). L'accord qualitatif du champ moyen (§ 3.2) tend à s'améliorer aux grandes longueurs (figure bas/gauche). L'estimation de l'extension la plus probable est correcte à toutes les longueurs. L'expression proposée par Winkler est dans tous les cas moins pertinente (médaillon). La chaîne gaussienne est d'autant plus adéquate que la longueur est grande $L \gg A$, en partant d'environ dix longueurs de persistance (figure bas/droite). Aux mêmes longueurs, le développement de Daniels au premier ordre (§ 3.1, en gris) réalise un meilleur accord.

L'expression de Ha et coll. produit un meilleur ajustement que celle de Winkler (médaillon), mais surtout elle permet une estimation *quantitative* de l'extension réalisant le maximum de la distribution indépendamment de la longueur considérée. D'autre part, le développement de Daniels (§ 3.1) ne permet pas une prise en compte significative des effets élastiques : il ne devient pertinent que pour $L/A \gtrsim 10$. Toutefois, à longueur fixée il réalise déjà au premier ordre un accord nettement plus précis que la chaîne gaussienne (figure 4.3, bas/droite).

Pour finir nous avons également testé la fiabilité de nos résultats par calcul explicite des deux premiers moments de la distribution, pour lesquels nous disposons d'expressions analytiques exactes (§ 2.1). Notre évaluation numérique ne montre aucune déviation, aux résultats théoriques, avec la longueur du polymère (médaillon de la figure 4.3, haut/gauche).

Nous pouvons donc calculer numériquement la distribution isotrope $S(r; L)$ exacte d'un WLC de longueur *a priori* quelconque. Comme nous l'avons remarqué pour la chaîne gaussienne (§ 2.3), l'estimation du rapport $S(r; L)/(4\pi r^2)$ dans la limite des faibles extensions $r \rightarrow 0$, à longueur L fixée, doit alors par définition nous fournir le facteur de cyclisation $J(L) = Q(\vec{0}; L)$ du WLC. Par soucis de cohérence avec les résultats obtenus [RvHL95, MRKL98, KML98] par Merlitz, Rippe, von Hippel, Klenin et Langowski au moyen de simulations de dynamique brownienne (1995–1998) d'une part et par Podtelezhnikov et Vologodskii [PV00] au moyen de simulations Monte Carlo (2000) d'autre part, nous calculons également le facteur de cyclisation comme limite du rapport

$$K(r; L) = \text{Prob}(r; L)/V(r) \quad \text{avec} \quad V(r) = \frac{4}{3}\pi r^3.$$

Nous reconnaissons la probabilité $\text{Prob}(r; L)$ de trouver la distance bout à bout d'un polymère de longueur L dans une boule de rayon r , précédemment définie pour la chaîne gaussienne (§ 2.3). Nous évaluons $J(L)$ de manière indirecte, mais nous pouvons alors étudier l'effet d'une non fermeture exacte des configurations de boucles en jouant sur le rayon $r \gtrsim 0$ de la boule. Le problème biologique correspondant est l'influence de la taille finie des protéines, fixant les boucles d'ADN observables dans une régulation génétique à l'étape de transcription. Intuitivement aucun changement significatif dans l'énergétique d'une telle cyclisation n'est attendu à moins que la taille des protéines en question ne soit de l'ordre de grandeur de celle de la séquence d'ADN cyclisée, c'est-à-dire $r \lesssim L$.

4.3 Pertinence des effets de tailles selon la longueur de boucle

Utilisant les distributions $S(r; L)$ précédemment calculées, nous avons évalué la distribution $Q(r\vec{e}_r; L)$ pour diverses extensions et le rapport $K(r; L)$ pour divers rayons de boule. Avant de discuter les résultats, examinons les possibles sources d'erreurs s'ajoutant à l'utilisation d'Expokit et de la FTT. En section précédente, nous avons obtenu $S(r; L)$ par dérivée de $P(z; L)$. Ce calcul est apparu fiable, soutenant la comparaison à diverses approximations ou résultats exacts. Ceci étant dit, l'évaluation numérique d'une dérivée doit systématiquement être considérée comme un sujet délicat. Dans le cas du rapport K nous pouvons

toutefois omettre cette étape de dérivation en intégrant par parties

$$\text{Prob}(r; L) = -2 \int_0^r r' \frac{dP}{dz}(r'; L) dr' = 2 \left[\int_0^r P(r'; L) dr' - r P(r; L) \right],$$

ce qui constitue une occasion supplémentaire de juger *a posteriori* notre calcul de S . En fait, la dérivation n'apparaît pas directement comme l'étape limitante de notre algorithme. Que ce soit pour Q ou K c'est la définition en z de P , c'est-à-dire l'échantillonnage en k_z de l'amplitude du vide Z qui est critique. Cela est particulièrement vrai aux faibles r que ce soit pour évaluer P , sa dérivée ou encore son intégrale.

Les résultats sont visibles en figure 4.4. Dans sa colonne de droite nous sommes intéressés par le comportement global du facteur de cyclisation $J(L)$. Il est tracé sur une échelle de longueurs incluant à la fois les régimes *élastique et entropique*. Cette représentation nous permet de juger de la pertinence des principales approximations exposées au chapitre 3. La colonne de gauche se concentre en revanche sur l'échelle de longueurs²¹ allant de 100 bp à 500 bp, et concerne plus particulièrement les comportements de la distribution Q ou du rapport K , pour $r \geq 0$ (échelle semi-logarithmique). Le graphe du haut montre que ces *deux quantités* convergent (différemment) vers J quand l'extension/rayon r est diminué. Par commodité considérons le rapport K (lignes pleines) : l'influence du rayon r augmente quand la longueur L diminue. Nous ne constatons aucun effet sur K en sa borne supérieure $L = 500$ bp : il reste du même ordre de grandeur. Cela est en fait vrai dès que $L \gtrsim 300$ bp, ou 100 nm. Il en va différemment pour la borne inférieure $L = 100$ bp. Partons de $r = 10$ nm : en diminuant de moitié le rayon ($r = 5$ nm) nous observons une baisse d'un ordre de grandeur pour K . Si nous divisons encore le rayon d'un facteur 5 pour finir à $r = 1$ nm le rapport K diminue également, mais la baisse constatée représente moins d'un ordre de grandeur (environ un tiers).

Cette étude met donc également en évidence une convergence *rapide* du rapport K (resp. de la distribution Q) vers le facteur de cyclisation $J(L)$, quand le rayon (resp. l'extension) r est diminué. Nous pouvons formuler le critère suivant : la convergence est atteinte dès que r est inférieur à L d'au moins *un ordre de grandeur*. Une évaluation quantitative de $J(L)$ jusqu'à une longueur de persistance (150 bp, ou 50 nm) est en effet possible à partir de $K(5 \text{ nm}; L)$, comme le montre le graphe de droite correspondant. Nous obtenons le profil de J attendu : un unique maximum définissant une longueur de boucle *la plus probable* d'à peu près 500 bp – entre trois et quatre longueurs de persistance – pour laquelle effets élastique et entropique sont tous deux importants. Les longueurs à la fois de plus en plus courtes ou importantes, sont alors de moins en moins susceptibles d'être observées : la rigidité élastique rend difficile la *courbure* de petits polymères (courtes séquences) tandis que le comportement entropique rend

²¹Rappelons que la longueur de persistance de l'ADN est de 150 bp, ou 50 nm.

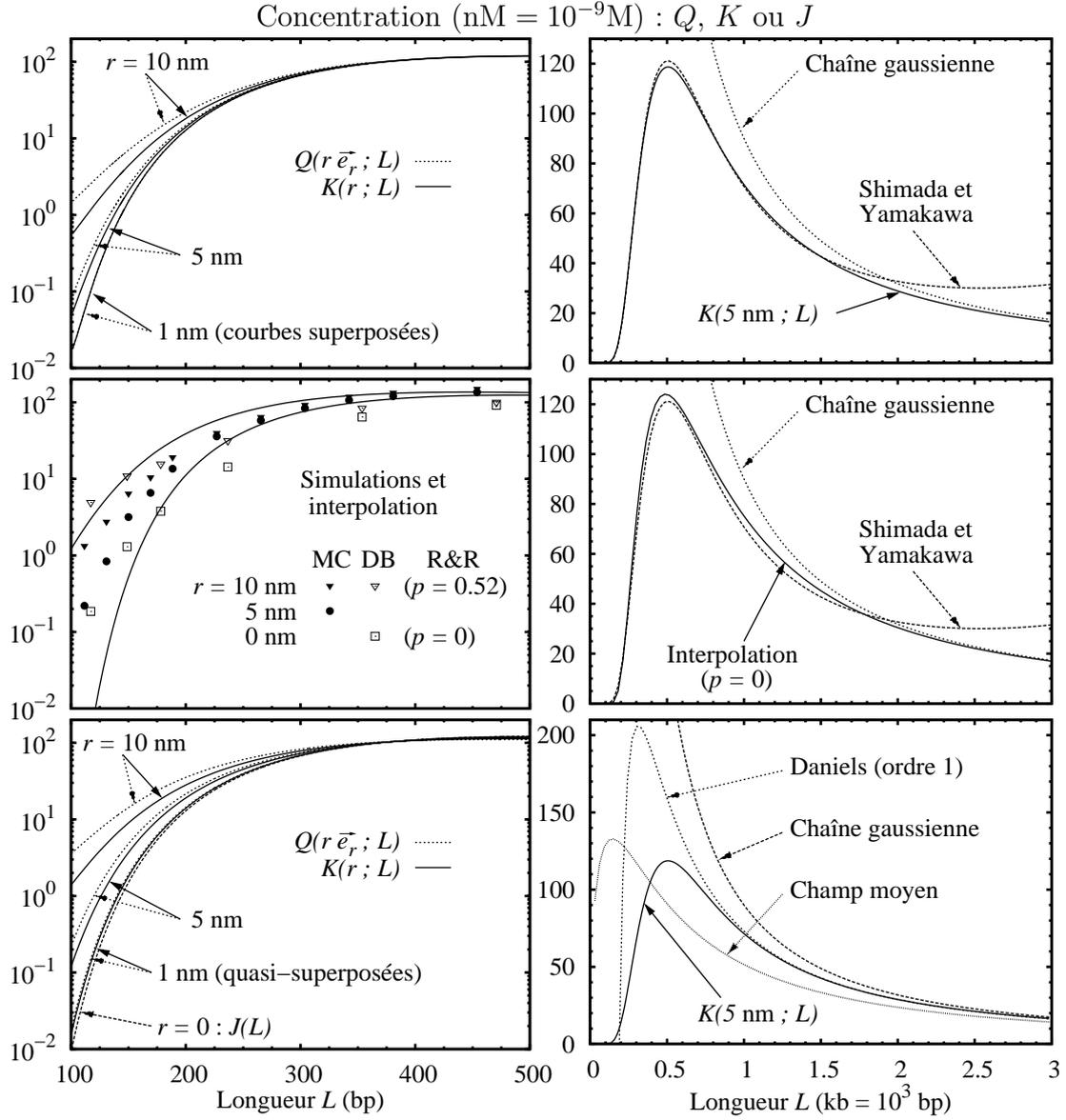


FIG. 4.4 – Calcul numérique du facteur de cyclisation J par sa distribution en extension (directement Q ou par le rapport K). Un accord quantitatif est obtenu dès que l’extension est inférieure d’un ordre de grandeur à la longueur considérée (figures du haut). Ces résultats vérifient (figure centre/gauche) de précédentes simulations Monte Carlo (MC) ou de dynamique brownienne (DB). Ils suppriment les divergences respectives des expressions de Shimada et Yamakawa, et du polymère gaussien (figure haut/droite) qui se complètent remarquablement d’autre part (recouvrement vers $L/A \simeq 13$). Ils permettent enfin de juger de la pertinence d’approximations mentionnées au chapitre 3 (figure bas/droite). Les figures du centre montrent également (lignes pleines) la formule interpolative (3.1) de R&R. La figure bas/gauche montre un accord quantitatif, aux faibles extensions relatives r/L , de l’adaptation (2.4) de S&Y. S’il devient qualitatif quand r/L augmente, il reste acceptable pour les valeurs étudiées après ($r \simeq 10$ nm).

peu probable la *coïncidence* des deux extrémités de longs polymères (grandes séquences). Notons toutefois que notre algorithme ne permet jamais d'atteindre les configurations de boucles à proprement parler, obtenues pour un rayon exactement nul $r = 0$. Étant donné les faibles rayons considérés le seul effet observé est un évènement fortement piqué, traduisant les configurations de bâtonnets quasi-rigides $L \gtrsim r$.²² L'échelle $L \geq 100$ bp que nous considérons n'inclut pas cet évènement : par exemple pour le plus grand rayon que nous considérons (c'est-à-dire $r = 10$ nm), si l'évènement correspondant commence à s'étaler il reste néanmoins piqué autour de $L \gtrsim 30$ bp.

Comme le montre le graphe centre/gauche, nos résultats s'accordent aux précédentes simulations disponibles comme celles de *dynamique brownienne* réalisées par Merlitz et coll. (symboles blancs) pour des rayons $r = 0, 10$ nm ou encore celles de *Monte Carlo* produites par Podtelezhnikov et Vologodskii (symboles noirs) pour des rayons $r = 5, 10$ nm. Les données de Merlitz et coll. doivent être comparées au rapport K en gardant à l'esprit qu'elles incluent la rigidité de torsion de l'ADN, et sa nature électrostatique.²³ Chacun de ces deux types de simulation présente aussi son lot de *difficultés* : les configurations dont l'extension se trouve dans une boule de rayon r sont des *évènements rares*, ne se réalisant qu'au bout d'un *temps long*. Ce constat s'accroît d'autant plus que r est petit (voir le critère ci-dessus) ou que le polymère est rigide (longueur approchant la persistance). Un des auteurs du premier groupe (Rippe) a récemment proposé [RCA⁺99, Rip01] un ajustement de ses propres données au moyen de la formule (3.1) *interpolant* entre régimes élastiques et entropiques. Rappelons qu'elle doit permettre la prise en compte de $r \geq 0$ au moyen d'un paramètre p . Nous avons reproduit cet ajustement, afin d'illustrer la précision satisfaisant les auteurs de ces simulations.

Pour finir nous testons l'expression (2.4) de $Q(r \vec{e}_r; L)$ que nous avons proposé à partir du facteur de cyclisation $J(L)$ obtenu par Shimada et Yamakawa. Comme attendu, l'accord aux résultats numériques passe de quantitatif à qualitatif quand l'extension relative r/L augmente (graphe bas/gauche). Prenons la borne inférieure $L = 100$ bp : pour $r = 1$ nm, ou $r/L = 3\%$ les courbes se superposent mais pour $r = 5$ nm, ou $r/L = 15\%$ la description n'est déjà plus que qualitative. Pour la borne supérieure, en fait dès $L \gtrsim 300$ bp cette évaluation fournit le bon ordre de grandeur pour une large gamme d'extensions r .

Revenons au comportement global de J (colonne de droite). Sur le graphe du haut nous constatons la pertinence de l'expression de S&Y (§ 2.4) jusqu'à dix longueurs de persistance. Au-delà, la chaîne gaussienne (§ 2.3) décrit bien le WLC dont le comportement est en effet essentiellement entropique. Le développement de Daniels (§ 3.1) permet d'obtenir un meilleur accord pour ces mêmes longueurs

²²Rappelons que le WLC est inextensible : Q et K ne sont définis que pour $L \geq r$.

²³Nous développons un peu plus en section 5.4.

mais il échoue à prendre convenablement en compte les effets de rigidité élastique du WLC, comme nous pouvons le voir dans le graphe du bas. Ce graphe confirme également la non pertinence des autres schémas d'approximation, tel le champ moyen (§ 3.2) ou le développement de Wilhelm et Frey (§ 3.3, non montré). Finalement, la formule (3.1) de $p = 0$ réalise une interpolation convenable entre les régimes élastique et entropique (graphe du centre).

4.4 Taille du répresseur LacR de l'opéron Lac

Nous nous proposons d'illustrer les résultats de la précédente section sur l'exemple de la répression de l'opéron Lac, déjà évoquée en section 1.1. L'expression des protéines permettant à la bactérie *E. Coli* de réaliser le métabolisme du sucre lactose, peut être stoppée à l'étape de transcription par cyclisation de deux séquences distinctes, la plus courte incluant le promoteur de l'opéron. Nous notons $O_1O_3 = 76$ bp et $O_1O_2 = 385$ bp les deux boucles d'ADN correspondantes. Rappelons que les opérateurs $O_{1,2,3}$ sont en fait les courtes séquences spécifiques, auxquelles le répresseur LacR s'accroche fixant ainsi les boucles.²⁴ Une répression efficace repose sur la coopération des deux processus : bien que O_1O_3 contienne le promoteur de l'opéron sa cyclisation est moins probable que celle de O_1O_2 , de part sa longueur (voir le facteur de cyclisation déterminé précédemment, figure 4.4).

Par résolution numérique des équations déterminant la configuration générale d'un WLC, Balaëff, Mahadevan et Schulten ont estimé l'énergie élastique de la boucle O_1O_3 autour de $23 k_B T$ [BMS99, BMS03]. Le modèle étudié par ces auteurs est en fait beaucoup *plus général* que notre WLC. Il prend en compte la rigidité de torsion de l'ADN et utilise des orientations initiale et finale de la boucle s'accordant aux structures cristallographiques disponibles pour le LacR (voir note en bas de page 64). Une anisotropie dans la rigidité de courbure de l'ADN est également considérée pour traduire sa structure microscopique (grand et petit sillon). Ils proposent finalement un traitement qualitatif de la nature électrostatique de l'ADN. Les auteurs attribuent 81% des $23 k_B T$ à la *courbure* de l'ADN soit environ $19 k_B T$. Les $4 k_B T$ restants correspondent à sa torsion, l'effet de l'anisotropie ne pouvant par ailleurs pas être isolé.²⁵ En

²⁴Ce sont les deux dimères composant le tétramère LacR qui concrètement reconnaissent les opérateurs et s'y accrochent. C'est un produit du métabolisme du lactose qui induit le changement configurationnel des dimères (pas besoin d'ATP), empêchant cet accrochage.

²⁵Ces auteurs ont également exploré le rôle d'un accrochage coopératif du LacR et de la protéine « activatrice » CAP, sur la structure de la boucle [BMS04]. Un article écrit avec Koudella [BKMS04] reprend et complète ces résultats, au moyen de simulations de dynamique brownienne et de Monte Carlo sur un WLC identique à celui que nous utilisons. La structure du complexe LacR/ O_1/O_3 est dans ces travaux une donnée fixe, mais Villa a récemment complété la description de Balaëff et Schulten par simulations de dynamique moléculaire de ce

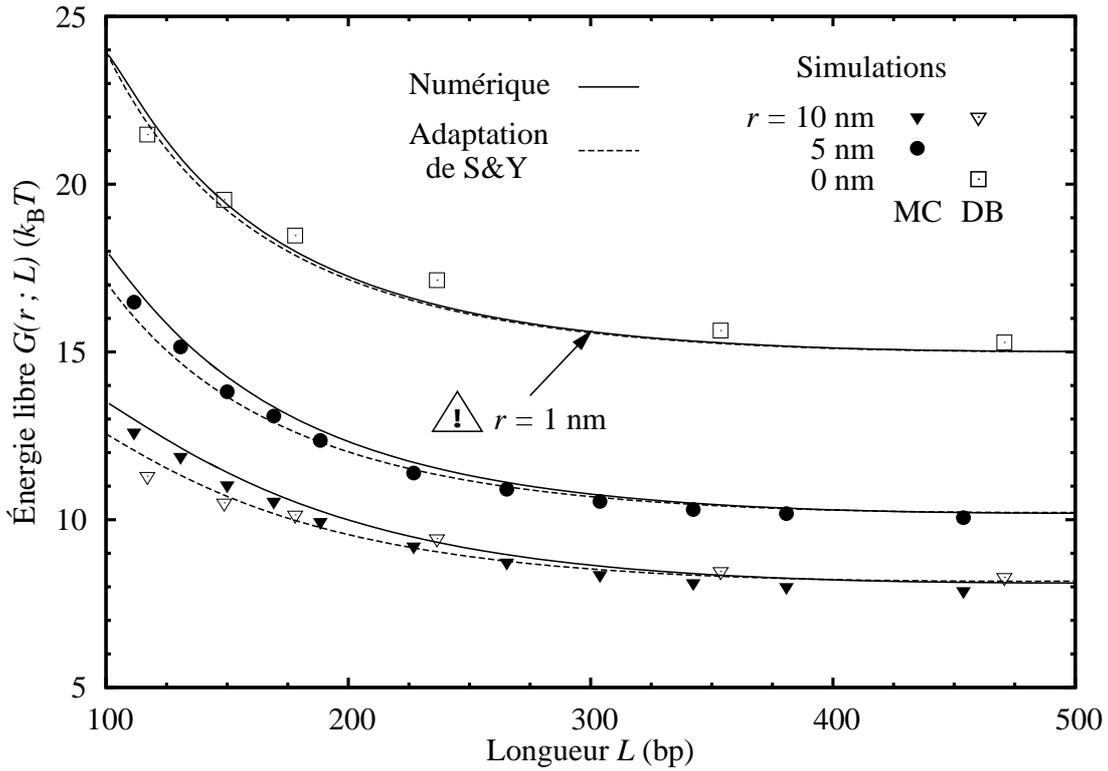


FIG. 4.5 – Variation avec la longueur L du polymère, de l'énergie libre associée aux configurations d'extensions se trouvant dans une boule de rayon r . Elle se déduit en $k_B T$, de $-\ln [K(r; L) \times V(r)]$ avec le volume $V(r) = \frac{4}{3}\pi r^3$ (cf fin § 2.3). Les données de la figure 4.4 sont reprises : notre évaluation numérique (lignes pleines), l'adaptation (2.4) du résultat de S&Y (pointillés) ainsi que les simulations de Monte Carlo (MC) et de dynamique brownienne (DB).

revanche, cette étude est purement *mécanique* : elle néglige les effets d'origine entropique. Les auteurs justifient ce choix du fait que $O_1 O_3$ ne représente qu'une demie-longueur de persistance. Ils valident alors l'énergie obtenue par comparaison à l'énergie libre d'environ $20 k_B T$ calculables des affinités du LacR mesurées en 1987 par Hsieh, Whitson, Matthews et Wells [HWMW87] (TABLE I).²⁶ En toute rigueur, s'il est vrai que pour des longueurs aussi faibles la configuration de point col (limite élastique) domine, les effets entropiques ne peuvent pas *a priori* être négligés. Leur estimation doit donc être comparée à l'énergie du point col.

complexe [VBMS04, VBS05]. On retiendra que ces considérations *ne modifient pas* l'ordre de grandeur de l'énergie élastique de la boucle, autour de $20 k_B T$.

²⁶L'affinité (en fait son inverse) du LacR avec une séquence incluant O_1 et O_3 (espacés de 92 bp) est évaluée entre $3.4\text{--}6.2 \times 10^{-12}$ M. La différence d'énergie libre associée à cet accrochage est alors (via le logarithme) d'environ $-26 k_B T$. Si la séquence ne contient que O_1 l'affinité du LacR varie entre $9.8\text{--}1.1 \times 10^{-11}$ M. Pour O_3 seul, une affinité 100 fois inférieure (en fait supérieure) fournit le bon ordre de grandeur. Correspondent à ces accrochages des différences d'énergie libre de $-25 k_B T$ et $-21 k_B T$. Une énergie libre de cyclisation peut alors être calculée par la différence $\simeq 26 - (-25 - 21) = 20 k_B T$.

Au risque d'être répétitif notons que ce dernier correspond aux orientations limites minimisant l'énergie élastique (§ 2.2), qui sont *a priori* différentes de celles utilisées par Balaëff et coll. obtenues des diverses structures cristallographiques du LacR. De même, les boucles qu'il obtiennent ne sont *a priori* pas planaires contrairement au point col. De $\varepsilon_*(0) = 14.055$ (figure 2.3) nous obtenons une énergie de $27.7 k_B T$, pour $L = 76$ bp ($22.9 k_B T$ pour les 92 bp de Hsieh et coll.). Connaissant la structure cristallographique du LacR nous pouvons estimer sa taille typique.²⁷ Suivant Brenowitz, Pickar et Jamison [BPJ91] nous retenons la valeur d'environ 13 nm, ce qui représente plus de 50% de la longueur de la boucle. L'énergie diminue alors significativement à $10.9 k_B T$ (figure 2.3) : dans les deux cas, l'accord avec Balaëff et coll. est assez mauvais. Si l'ensemble des différences précédemment évoquées entre les deux modèles peut être suspecté, notons tout d'abord que la taille du LacR est de l'ordre de grandeur de la longueur de la boucle ($r/L \simeq 50\%$). Nous attendons donc une *forte dépendance* de l'énergie à la taille du LacR retenue. Ainsi pour récupérer une énergie d'environ $19 k_B T$ une extension relative $r/L = 25\%$ suffit (nous lisons $\varepsilon_*(0.25) = 9.4$ sur la figure 2.3), c'est-à-dire $r \simeq 6$ nm. Cette estimation reste *raisonnable* compte tenu du fait que notre modèle n'inclut *aucun détail* géométrique de la fermeture de la boucle (complexe LacR/boucle d'ADN) : nous n'avons donc aucune raison d'être catégorique sur le choix $r = 13$ nm, c'est pourquoi nous utilisons dans le reste de ce travail une *taille typique* $r = 10$ nm pour les protéines fixant les boucles étudiées. Pour O_1O_3 (donc $r/L \simeq 40\%$) nous retenons donc une énergie de $14.1 k_B T$. Qu'en est-il de l'énergie libre associée à cette cyclisation ? Les figures 4.4 (gauche) et 4.5 couvrent l'échelle de longueurs $100 \leq L \leq 500$ bp. En anticipant sur la section 5.4 (voir en particulier la note de bas de page 81), nous pouvons en fait descendre jusqu'à $L \gtrsim 50$ bp. Nous calculons $K(10 \text{ nm}; 76 \text{ bp}) = 0.2255 \pm 0.0085$ nM auquel nous associons l'énergie $G(10 \text{ nm}; 76 \text{ bp}) = 14.38 \pm 0.55 k_B T$. Nous pouvons compléter la comparaison avec notre adaptation (2.4) de S&Y, pour laquelle $K = 0.601$ nM et $G = 13.4 k_B T$. Étant donné la grande extension relative considérée ($r/L \simeq 40\%$) nous n'attendons pas d'accord quantitatif entre cette dernière et le résultat numérique : nous observons en effet un facteur 3 entre les rapports K , bien qu'en termes d'énergie cela ne signifie qu'une différence de $\ln 3 \simeq 1 k_B T$. Finalement, l'accord avec l'estimation qualitative de Balaëff et coll. est là encore assez mauvais. Toutefois, comme les énergies élastique et libre sont très proches nous retenons que, pour une longueur aussi petite ($L/A \simeq 0.5$), les *effets entropiques* peuvent bien être négligés.

Nos résultats peuvent également être comparés aux expériences réalisées par Brenowitz et coll. (1991) sur une séquence de 114 bp (correspondant à la répression de

²⁷Voir par exemple le travail [LCH⁺96] de l'équipe : Lewis, Chang, Horton, Kercher, Pace, Schumacher, Brennan et Lu (1996). La structure incluant le répresseur et la boucle d'ADN n'est pas connue. Le brin ne « survit » pas à la cristallisation. Au mieux nous disposons de la structure du LacR avec un morceau de séquence à chacun de ses deux dimères.

l'opéron Gal), contenant deux opérateurs possédant tous deux une forte affinité avec le LacR [BPJ91]. En mesurant les fractions de molécules d'ADN bouclées et non bouclées, ils calculent une énergie libre de cyclisation de $12.2 \pm 0.2 \text{ kcal} \cdot \text{mol}^{-1}$, c'est-à-dire $20.6 \pm 0.4 k_B T$, à laquelle ils associent un facteur de cyclisation de 0.8 nM. Plus précisément, les auteurs disposent à la fin de leur étude de plusieurs estimations : le choix que nous citons correspond en fait à une séquence de 167 bp. Brenowitz et coll. avaient déjà proposés à l'époque d'*inclure* la taille du LacR dans la longueur de la boucle : les 53 bp ($\simeq 17.7 \text{ nm}$) additionnels correspondent en effet *qualitativement* aux 13 nm avancés plus haut. Gardant notre estimation $r \simeq 10 \text{ nm}$, la figure 4.4 (haut/gauche) confirme cette estimation du facteur de cyclisation : $K(10 \text{ nm}; 114 \text{ bp}) = 1.03 \pm 0.09 \text{ nM}$. Nous lui associons une énergie libre de cyclisation de $G(10 \text{ nm}; 114 \text{ bp}) = 12.86 \pm 0.09 k_B T$ (figure 4.5, ligne pleine). Quand à notre adaptation (2.4) de S&Y, elle produit $K = 5.44 \text{ nM}$ (figure 4.4, bas/gauche) et $G = 11.20 k_B T$ (figure 4.5, pointillés). Comme pour les 76 bp, le mauvais accord pour K (imputable à $r/L \simeq 26\%$) apparaît atténué pour G . Le mauvais accord également obtenu avec les $20.6 k_B T$ de Brenowitz et coll. peut en revanche sembler *contradictoire* avec l'accord *quantitatif* constaté pour le facteur de cyclisation. Nous avons supposé que la boucle se forme (fixation par le LacR) dès que sa distance bout à bout (entre O_1 et O_3) pénètre dans une boule de rayon typique 10 nm : ce processus autorise en fait de nombreuses configurations ne se concrétisant pas *biologiquement* par une boucle (orientations relatives des opérateurs, disposition du LacR, etc.).²⁸ Notre choix de $V(r) = \frac{4}{3}\pi r^3$ est simpliste mais il constitue une bonne *première approximation* compte tenu du manque d'informations systématiques dont nous disposons sur l'interface LacR/boucle d'ADN (voir note en bas de page 64, [VBMS04, VBS05]). Pour finir, l'énergie élastique du point col est de $12.1 k_B T$: les *effets entropiques* sont plus importants que pour la boucle de 76 bp, mais ils peuvent encore être négligés.

Pour la boucle $O_1 O_2 = 385 \text{ bp}$, la taille du LacR ($r/L \simeq 8\%$) joue un moindre rôle. Comme constaté en figure 4.4 (haut/gauche), dès $L \geq 300 \text{ bp}$ le rapport $K(r; L)$ est déjà convergé pour $r \leq 10 \text{ nm}$. Pour $L = 385 \text{ bp}$, nous lisons en effet $K(10 \text{ nm}; L) = 104.6 \pm 1.1 \text{ nM}$, tandis que $K(1 \text{ nm}; L) = 103.7 \pm 1.2 \text{ nM}$. Les énergies libres de cyclisation associées ne diffèrent donc que par *le volume* $V(r)$

$$G(1 \text{ nm}; L) - G(10 \text{ nm}; L) \simeq k_B T \times 3 \ln(10/1) \simeq 6.91 k_B T.$$

La figure 4.5 confirme ce résultat : $G(10 \text{ nm}; L) = 8.24 \pm 0.01 k_B T$ tandis que $G(1 \text{ nm}; L) = 15.15 \pm 0.01 k_B T$. L'adaptation (2.4) de S&Y reproduit fidèlement ces résultats étant données les faibles extensions relatives r/L mises en jeu. Notons également que pour une telle longueur, les *effets entropiques* doivent être considérés : l'énergie élastique du point col ($r = 10 \text{ nm}$) n'est que

²⁸Il ne prend pas non plus en compte les affinités du LacR avec les opérateurs.

de $4.85 k_B T$.

Rappelons pour finir que les expériences de cyclisation à proprement parler utilisent des enzymes type *ligases* plutôt que des protéines comme le LacR (§ 1.2). Ces enzymes attachent de manière irréversible les extrémités des séquences d'ADN dès que ces dernières coïncident. Dans ce cas la formation de boucles nécessite alors une extension $r = 0$, *quelque soit* la longueur de la séquence. De plus, aux faibles longueurs la forme favorisée des boucles est *circulaire*, et non celle du point col.²⁹ Les travaux auxquels nous avons comparé nos résultats ci-dessus considèrent tous la formation de boucle par l'accrochage d'une protéine, le LacR. Cette remarque permet néanmoins de souligner l'importance des orientations limites de la boucle. Par exemple, il est raisonnable de penser que la rigidité de torsion considérée par Balaeff et coll. ou encore Merlitz et coll. produit des différences significatives, particulièrement si les extrémités de la boucle sont fixées. Nous attendons néanmoins plus encore de la rigidité de torsion si cette fixation par la protéine, disons le LacR, requiert un alignement particulier avec chaque opérateur. Une manifestation typique sont les oscillations observées par Baldwin et Shore (1981/1983) dans le facteur de cyclisation [SLB81, SB83], de période la torsion intrinsèque de l'ADN.

²⁹Cette circularité a récemment été remise en cause lors d'expériences de cyclisation de courts fragments [CW04, CW05], avec des ligases. Ces expériences sont certes controversées [DSS⁺05], mais cela plaide quelque peu pour l'*abus de langage* adopté dans ce manuscrit. Nous n'avons pas distingué les phénomènes de cyclisation de ceux de formation d'une boucle par une protéine, sauf quand cela prêtait à confusion.

Chapitre 5

Perte de rigidité de l'ADN : cassure effective à mi-longueur

Au chapitre 4 nous avons modélisé au moyen du WLC (chapitre 2) la formation de boucles dans l'ADN, notamment observables lors d'une régulation de l'expression génétique à l'étape de transcription. La considération de la *taille finie* des protéines fixant ces boucles s'est révélée d'autant plus pertinente que leur *séquence* est courte, et donc difficile à courber : plus précisément la taille de ces protéines peut être considérée comme nulle dès qu'elle représente moins de 10% de la longueur de la séquence qu'elles ferment (§ 4.3).¹ C'est le cas des séquences de plus de deux longueurs de persistance (c'est-à-dire 300 bp, ou 100 nm) considérant une taille typique de 10 nm pour ces protéines, comme c'est le cas pour le répresseur de l'opéron Lac (§ 4.4).

Cette modélisation reste naïve par de multiples aspects. Elle ne prend notamment en compte qu'un seul type de protéines : celles fixant la boucle. C'est une situation expérimentale réalisable sans problème *in vitro*, mais largement idéale *in vivo* où les molécules d'ADN sont constamment courbées ou parcourues par de *multiples* protéines. De tels accrochages peuvent modifier significativement la *rigidité* effective (localement) de la molécule d'ADN [RB99]. Cette perte (ou gain) de rigidité peut aussi bien être une simple *conséquence* de l'accrochage qu'une *fonction* particulière des protéines : voir la liste succincte en section 1.5. Rappelons ici le cas de la protéine HU sans laquelle le GalR ne peut maintenir stablement la boucle mise en jeu lors de la répression de l'opéron Gal.

Une signature expérimentalement accessible de la réponse élastique de l'ADN est sa relation de force/extension (§ 4.1). Si la molécule contient une boucle nous pouvons aussi mesurer son facteur de cyclisation (§ 4.3). Ces deux quantités ont en effet été pressenties comme *indicateurs* de tels accrochages, mais leurs effets ne peuvent pas être *systématiquement* intuités, motivant alors les études

¹Cette étude concerne la taille et non les contraintes que ces protéines imposent aux extrémités de la séquence.

quantitatives que nous résumons dans la prochaine section introductive.

5.1 Modélisation : faire le tri

Intuitivement une perte de rigidité, induite par l'accrochage de protéines, doit pouvoir être prise en compte par l'introduction d'une longueur de persistance *effective* [YM03, KML⁺05]. Cela est simple dans la mesure où la quasi-totalité de nos résultats ne dépendent que de la longueur réduite L/A . En revanche une telle longueur de persistance devient un *indicateur* peu pertinent de la rigidité de la molécule d'ADN. Réciproquement elle est également peu informative sur le nombre d'accrochages, leurs amplitudes, etc. Bref cette méthode ne requiert pas une compréhension de la physique de ces mécanismes. Il est par ailleurs vraisemblable qu'elle ne soit valable que pour certaines conditions.² En sa faveur, les mesures de force/extension ou du facteur de cyclisation sont systématiquement ajustées aux expressions du type M&S (§ 4.1) ou S&Y (§ 2.4) n'incluant pas ces accrochages. Pouvoir préciser une longueur de persistance effective, même si nous disposons d'un modèle plus élaboré, permet une comparaison directe à ces travaux [KML⁺05, BBS⁺00].

D'autre part, une approche « tout analytique » n'a aucune chance d'aboutir : nous ne disposons pour le WLC, d'expressions exactes ni pour sa relation de force/extension, ni pour son facteur de cyclisation. Nous généralisons donc les méthodes (semi-analytique et numérique) précédemment utilisées : ne reste qu'à préciser le mécanisme que nous voulons décrire *en pratique*.

Permettons nous avant cela une parenthèse. Ces mécanismes sont en fait étudiés depuis longtemps dans un cadre plus général que celui de l'accrochage de protéines à la molécule d'ADN (cyclisée ou non), incluant notamment les effets de *séquence* (contenu en paires de bases) du brin ou bien encore l'apparition de *défauts* dans sa structure en double hélice [LC86, BMRJ88, HR90]. La modélisation d'un accrochage par une *cassure* rigide dans le polymère date de la même époque. Des expériences de cyclisations réalisées par Cloutier et Widom [CW04, CW05] ont récemment relancé l'intérêt d'une description théorique de tels défauts. Nous ne détaillons pas les mesures, qui se sont *controversées* (voir l'article de Du, Smith, Shiffeldrim, Vologodskaja et Vologodskii [DSS⁺05]), mais elles suggèrent à leurs auteurs une remise en question du WLC.³ Tandis que les premiers travaux utilisaient des modèles *microscopiques*, étudiés par simulations type Monte Carlo, les développements récents consistent à inclure les mêmes défauts dans un

²Voir par exemple la fonction de corrélation déduite en 1998 par Rivetti, Walker et Bustamante [RWB98]

³Les résultats montrent un J significativement plus important que celui du WLC aux faibles longueurs tout en confirmant qualitativement ses prédictions aux grandes longueurs. La dépendance à la torsion de l'ADN est aussi plus importante que celle attendue. Ces expériences ne contiennent pas de protéine susceptible de s'accrocher au brin.

modèle *mésoscopique* du type WLC.

Les travaux théoriques qui ont suivi ces expériences se sont en effet concentrés sur une formulation du WLC ne reposant pas sur l'hypothèse de *linéarité*, usuellement posée en élasticité. Reprenons l'exemple pédagogique typique de la courbure d'une paille (en plastique, pour boire). Tant que la courbure imposée est faible, elle se répartit le long de la paille dont le profil reste alors continu (différentiable). Si la courbure s'amplifie, elle va en revanche brusquement se localiser en un point de la paille qui par conséquent se plie, se casse (transition produisant une discontinuité du profil). Citons les articles de : Chakrabarti et Levine [CL05]; Wiggins, Phillips et Nelson [WPN05]; Popov et Tkachenko [PT05a]. Ces modèles reposent *tous* à un moment sur une résolution numérique. Ils ne diffèrent que par la *nature* des non linéarités, des « défauts », susceptibles d'apparaître tout au long du polymère.

La rigidité de la cassure est en fait imposée par celle du complexe, formé par l'accrochage de la protéine à l'ADN. Ces cassures peuvent donc aussi bien être rigides [PT05a] que flexibles [WPN05] : elles sont *a priori* semi-flexibles, avec une rigidité élastique plus ou moins grande que celle de l'ADN nu. Au contraire des cassures rigides où l'angle est fixé, celles semi-flexibles peuvent privilégier ou non un certain angle. Enfin, leur localisation sur la molécule d'ADN peut aussi bien être spécifique en sites ou non. Elles peuvent être permanentes ou au contraire apparaître (et disparaître) à la manière d'excitations thermiques, entraînant dans les deux cas un coût énergétique à préciser. En 2003 Yan et Marko ont étudié [YM03], toujours numériquement, les effets respectifs d'une telle diversité de cassures sur la réponse élastique d'un WLC (discret) [DA00], par calcul de sa relation de force/extension. Ils ont récemment (2005) généralisé leur méthode [YKM05], avec l'aide de Kawamura, au calcul du facteur de cyclisation.⁴ Ces quelques considérations s'imaginent particulièrement bien par l'accrochage de protéines à l'ADN, mais rappelons au lecteur qu'il possède une certaine liberté quand à l'interprétation de ces cassures. Suite aux expériences de Cloutier et Widom, Yan et Marko ont ainsi postulé l'ouverture de petites *bulles* de dénaturation ($\simeq 3$ bp) le long de la molécule d'ADN [YM04]. Cette hypothèse est reprise dans l'article de Ranjith, Sunil Kumar et Menon [RSKM05] dans le cas d'une unique bulle à mi-longueur de la boucle.⁵ Même peu probable, l'apparition de tels joints flexibles *faciliterait* significativement la courbure de courtes séquences. L'énergie libre typique associée à l'ouverture de ces hypothétiques bulles ($\simeq 10 k_B T$) les rendent *peu plausibles* aux yeux des biologistes, mais cette

⁴Ils traitent par ailleurs diverses orientations limites à la fermeture de la boucle ($r = 0$).

⁵ $J(L)$ y est la somme de l'expression de S&Y (§ 2.4) pour les configurations sans bulle, avec l'intégrale sur les configurations composées des moitiés du polymère séparées par la bulle (dont l'ouverture a un coût énergétique). La distribution en extension de chaque moitié est celle du champ moyen (§ 3.2), tandis qu'une intégration numérique permet aux auteurs de corriger *a posteriori* une évaluation analytique de type point col.

énergie reste tout de même significativement inférieure à celle de cyclisation d'un court WLC. Pour finir, Wiggins et Nelson [WN05] ont récemment mis l'accent sur une *classe* de modèles, linéaires comme le WLC mais moins rigide que celui-ci.⁶ Aux *grandes* longueurs ces modèles (WLC inclu) montrent tous le *même* comportement, bien qu'ils puissent *différer* significativement les uns des autres aux *faibles* longueurs (classe d'universalité du WLC, § 3.4).

Dans ce chapitre nous considérons le cas d'une unique cassure, *fixe* en amplitude *et* position, à mi-longueur du WLC. Notre approche se révèle similaire à celle d'un travail récemment publié par Sankararaman et Marko [SM05b], mais nous conservons pour la boucle d'ADN la possibilité de n'être pas exactement fermée. Nous montrons dans un premier temps (§ 5.2) comment modifier la limite élastique (point col) ainsi que l'adaptation (2.4) de S&Y. Nous modifions ensuite (§ 5.3) l'algorithme numérique qui nous sert là encore de référence. L'effet de l'amplitude de la cassure est étudié en section 5.4. Nous y discutons brièvement sa pertinence comme modèle *effectif*, de mécanismes entraînant une perte de rigidité de l'ADN, c'est-à-dire sans tenir compte de leurs *détails* microscopiques. Dès que nous voudrions discuter spécifiquement des effets dus à la cassure, nous fixerons une extension de 10 nm (§ 4.4). Soulignons pour finir le rôle central que peut jouer la rigidité de torsion de l'ADN, en particulier quand l'accrochage de la protéine nécessite l'ouverture locale du double brin afin de s'y insérer [LBC⁺03, LGA99]. Nous négligeons cette torsion comme nous l'avons fait jusqu'ici. Là encore, cette approche se justifie dans un premier temps par l'étude de situations expérimentales dans lesquelles les extrémités de la molécule, et par conséquent sa torsion, ne sont pas contraintes (p. ex. fixées à un substrat et une bille).

5.2 La limite élastique du modèle du ver change

Pour prendre en compte une cassure à mi-longueur, la théorie développée en section 2.2 doit être généralisée à une orientation à mi-longueur *a priori* quelconque $\omega(L/2^-) = a/2$. Le polymère ainsi décrit est le collage (en la cassure) de deux WLC, de mêmes longueurs de contour $L/2$ et de persistance A . Autrement dit s'il reste symétrique, il n'est plus différentiable en $L/2$ où se produit la discontinuité⁷

$$\Delta = \omega(L/2^+) - \omega(L/2^-) = 2\pi - a.$$

Nous fixons ensuite cette discontinuité afin d'imposer l'amplitude de la cassure $\kappa = \pi - \Delta = a - \pi$. La situation sans cassure ($\Delta = 0$, $\kappa = \pi$) est bien

⁶Localement, leur énergie élastique dépend moins fortement de la courbure $|\dot{\vec{t}}(s)/ds|$ que celle du WLC (quadratiquement). Leur attention semble avoir été attirée par les résultats de [vNvdHdJ⁺03].

⁷Qui fait de A un indicateur alors peu pertinent de sa rigidité globale.

retrouvée par le choix de $a/2 = \pi$.

L'intégrale première du mouvement est inchangée mais son intégration sur $0 \leq s \leq L/2$ produit maintenant l'équation intégrale⁸

$$\xi = \int_b^\beta \frac{d\Omega}{\sqrt{1 - k^2 \sin^2 \Omega}} = -F(b, k) + F(\beta, k) \quad (5.1)$$

avec les « orientations » initiale $\beta = \Omega(0)$ et à mi-longueur $b = \Omega(L/2)$. De même, on obtient l'analogie de la seconde équation intégrale (2.3)

$$[2 - k^2(1 - r/L)] \xi = 2 \int_b^\beta \sqrt{1 - k^2 \sin^2 \Omega} \, d\Omega = 2 [-E(b, k) + E(\beta, k)]. \quad (5.2)$$

Les notations sont les mêmes qu'en section 2.2. L'amplitude de la cassure est contenue dans $b = (\pi - \kappa)/4$. En l'absence de cassure ($\kappa = \pi$) les intégrales elliptiques $F(0, k) = E(0, k) = 0$, par définition.⁹ Le paramètre k caractérise comme avant, la solution générale de l'équation du mouvement (annexe C). Fixant les valeurs de l'extension relative r/L , de l'amplitude de la cassure b et de l'orientation initiale β , le k cherché doit simultanément satisfaire les deux équations intégrales (5.1) et (5.2), soit

$$[2 - k^2(1 - r/L)] [-F(b, k) + F(\beta, k)] = 2 [-E(b, k) + E(\beta, k)].$$

Le *point col* est la configuration d'orientation *initiale* réalisant l'énergie élastique minimale. L'introduction de la cassure ne modifie pas l'expression formelle de ε . En revanche celle de ξ change, dans sa borne inférieure d'intégration $b \geq 0$. Cette dernière est indépendante de β (elle est fixée) et la dérivée de ε par rapport à l'orientation initiale est donc elle aussi inchangée, exprimée en fonction de ξ . La condition initiale α_* réalisant le minimum de l'énergie réduite ε est alors toujours reliée au k_* du point col, par

$$k_* \sin \beta_* = 1 \quad \text{et donc} \quad k_* > 1.$$

Une transformation de Jacobi réelle (§ 2.2) nous permet de considérer de tels $k = 1/x > 1$. Le paramètre x de la configuration recherchée est alors solution de

$$(1 + r/L) \{-F[B_{L/2}(x), x] + F[B_0(x), x]\} = 2 \{-E[B_{L/2}(x), x] + E[B_0(x), x]\}$$

avec les fonctions $B_0(x) = B(x, 0)$ et $B_{L/2}(x) = B(x, L/2)$ issues de $B(x, s) = \arcsin[\sin \Omega(s)/x]$. Son énergie élastique réduite se réécrit (en explicitant ξ)

$$\varepsilon(r/L; x, \alpha, \kappa) = 4 \{-F[B_{L/2}(x), x] + F[B_0(x), x]\}^2 [2x^2 - (1 - r/L)].$$

Le point col, réalisant l'énergie élastique minimale

$$\varepsilon_*(r/L; \kappa) \equiv \varepsilon(r/L; x_*, \alpha_*, \kappa)$$

⁸Nous considérons comme en section 2.2 les configurations de type (i).

⁹Le lecteur pourra vérifier la cohérence des prochains résultats, avec ceux de la section 2.2.

vérifie comme avant $\sin \beta_*/x_* = 1$, soit $\omega_*(0) = \alpha_* = 2 \arccos x_*$. Alors les F et E contenant $B_0(x_*) = \frac{\pi}{2}$ deviennent les intégrales elliptiques de Legendre *complètes* du premier et deuxième type respectivement, simplifiant quelque peu les expressions ci-dessus. Là encore, une configuration de type (i) correspond à une orientation initiale $\alpha < \alpha_*$, tandis que $\alpha > \alpha_*$ produit un cas (ii).

Pour une amplitude de cassure fixée κ , l'extension maximale que le polymère peut réaliser est $L \sin(\kappa/2) \leq L$. C'est une conséquence de l'inextensibilité des deux WLC le composant ainsi que de la rigidité du complexe polymère/protéine induisant la cassure. Cette configuration la *plus* étirée a une orientation initiale $\omega_*(0) = \omega(L/2) = (\pi + \kappa)/2$ (ou $x_* = \sin[(\pi - \kappa)/4]$), et elle est d'énergie nulle $\varepsilon_*(\sin(\kappa/2); \kappa) = 0$: c'est le collage en la cassure de deux WLC totalement étirés. La figure 5.1 confirme ce comportement et montre plus généralement une réduction importante de l'énergie élastique du polymère par l'introduction de la cassure. Cette dépendance s'exprime simplement pour les configurations de

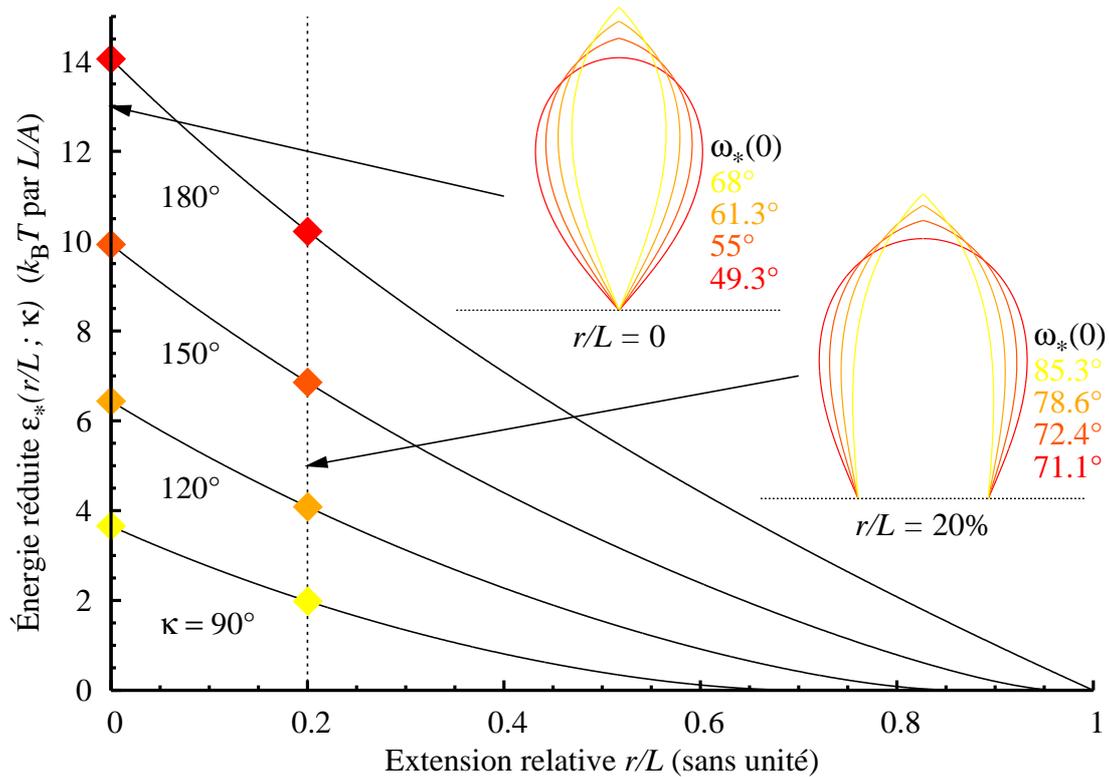


FIG. 5.1 – L'introduction d'une cassure à mi-longueur diminue significativement l'énergie élastique du point col. Son amplitude κ fixe les orientations du polymère en $L/2$, et pour une distance bout à bout r/L fixée le point col possède alors l'orientation initiale $\omega_*(0)$ (et finale) réalisant le minimum de l'énergie élastique. Sont ici tracées $\kappa = 90^\circ, 120^\circ, 150^\circ$ et $\kappa = 180^\circ$ soit une absence de cassure (§ 2.2). La configuration la plus étirée $r/L = \sin(\kappa/2)$ est d'énergie nulle. L'énergie s'exprime en $k_B T$ à partir de l'énergie réduite $\frac{\varepsilon_*(r/L; \kappa)}{L/A}$.

boucle $r = 0$: sur l'échelle de cassures « utiles » $\pi/2 \leq \kappa \leq \pi$, une régression linéaire de $\varepsilon_*(0; \kappa)$ produit un accord quantitatif (moins de 15% d'écart relatif¹⁰)

$$\varepsilon_{\text{lin}}(\kappa) = -7.287 + 6.639 \times \kappa$$

avec κ en radians. La comparaison est disponible en figure 5.2 où κ est exprimé en degrés, auquel cas $\varepsilon_{\text{lin}}(\kappa) = -7.287 + 0.116 \times \kappa$. Rappelons au lecteur peu

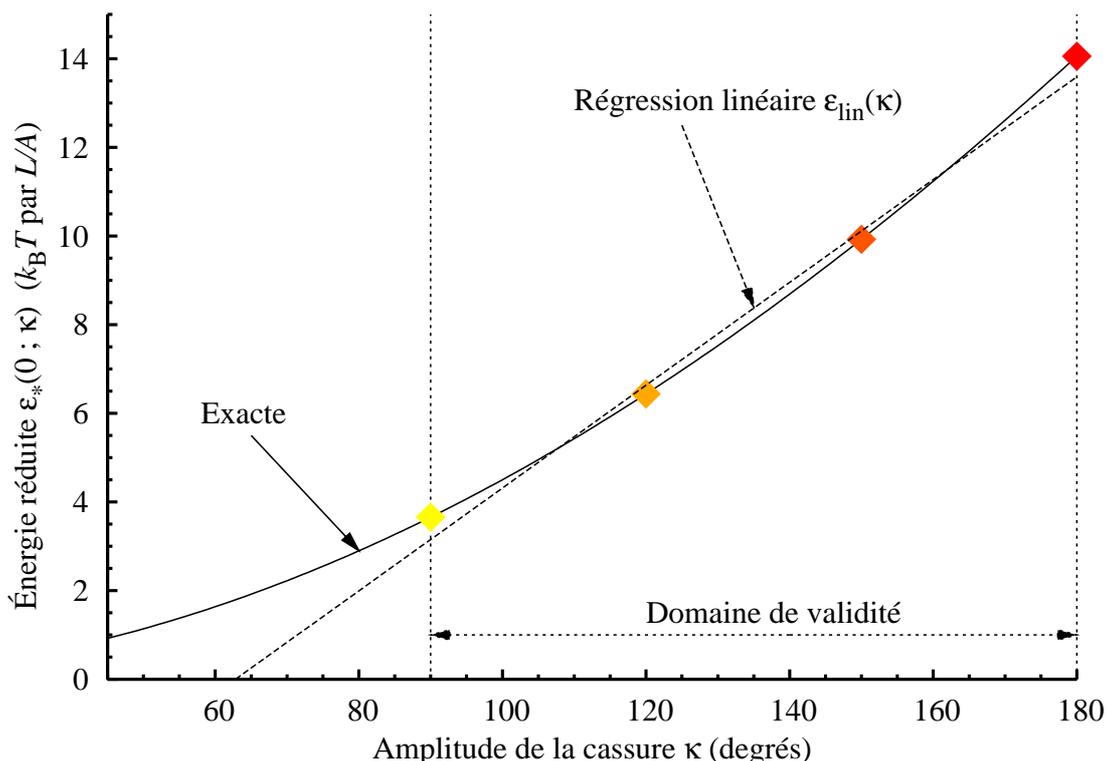


FIG. 5.2 – Pour les configurations de boucles ($r = 0$), l'énergie élastique dépend quasi-linéairement de l'amplitude $90^\circ \leq \kappa \leq 180^\circ$ de la cassure à mi-longueur. La régression linéaire $\varepsilon_{\text{lin}}(\kappa) = -7.287 + 0.116 \times \kappa$ produit moins de 15% d'écart relatif à $\varepsilon_*(0; \kappa)$, voir moins de 5% en se limitant à $\kappa \geq 100^\circ$ (avec la même régression).

familier des intégrales elliptiques qu'il peut trouver en annexe E le modèle du lignes & cercles ne faisant intervenir que des fonctions trigonométriques usuelles, et pouvant constituer une bonne première approximation du point col que nous venons de décrire (voir par exemple la figure 5.3).

Pour finir cette section nous proposons une expression analytique de la distribution en distance bout à bout $Q(\vec{r}; L, \kappa)$ du polymère, prenant en compte l'effet de la cassure à mi-longueur. Son calcul exact est impossible (§ 5.1) qu'il y ait une

¹⁰Moins de 5% si l'on se restreint à $\kappa \geq 100^\circ$, en gardant la même régression.

cassure ou non. Dans ce dernier cas ($\kappa = \pi$), nous avons proposé (§ 2.4) d'appliquer l'expression du facteur de cyclisation $J(L) = Q(\vec{0}; L)$ obtenue par Shimada et Yamakawa¹¹ à des configurations de boucles factices, dont la longueur $\bar{L}(r)$ prend en compte l'extension r des configurations réellement considérées. La longueur additionnelle empirique $\bar{L} - L = 2 \times r$ produit la formule (2.4), valide aux faibles extensions relatives ($r/L \lesssim 5\%$, § 4.3). Nous avons motivé ce choix simple par l'étude systématique des configurations (rigides) de boucles susceptibles d'approcher au mieux le point col de la situation $r/L \geq 0$ considérée (revoir les figures 2.5 et 2.6).

Dans le cas d'une cassure non nulle ($\kappa \neq \pi$), nous pouvons facilement reproduire une telle étude à condition de garder son amplitude κ fixe.¹² La figure 5.3 (droite) montre que nous pouvons conserver le choix empirique $\bar{L} - L = 2r$, et ce quelque soit κ . Pour avancer une expression de

$$Q(r \vec{e}_r; L, \kappa) \simeq \frac{1}{A^3} C(r; L, \kappa) \exp \left[-\frac{\varepsilon_*(r/L; \kappa)}{L/A} \right],$$

à partir du résultat de Shimada et Yamakawa (§ 2.4) nous devons néanmoins faire une hypothèse supplémentaire, à savoir : négliger le rôle de la cassure dans les fluctuations autour du point col

$$\begin{aligned} C(r; L, \kappa) &\simeq C(0; L + 2r, \kappa) \stackrel{?}{=} C(0; L + 2r, \pi) \\ \text{soit } Q(r \vec{e}_r; L, \kappa) &\simeq \frac{1}{A^3} C(0; L + 2r, \pi) \exp \left[-\frac{\varepsilon_*(r/L; \kappa)}{L/A} \right]. \end{aligned} \quad (5.3)$$

Pour $r = 0$, cette expression se réduit à celle obtenue [SM05b] par Sankararaman et Marko. Comme en section 2.4 nous comparons l'énergie réduite du vrai point col (ligne pleine) à celle¹³ $\frac{\varepsilon_*(0; \kappa)}{1+2r/L}$ de la boucle factice (pointillés). En figure 5.3 nous considérons des amplitudes $\kappa = \pi/2, 2\pi/3$ et $5\pi/6$ (soient $\kappa = 90^\circ, 120^\circ$ et 150°) pour lesquelles nous pouvons également utiliser la régression linéaire $\varepsilon_*(0; \kappa) \simeq \varepsilon_{\text{lin}}(\kappa)$ (pointillés gris). Avec la boucle factice, l'extension r (du vrai point col) n'intervient qu'à travers la longueur additionnelle : elle n'est donc pas limitée à $L \sin(\kappa/2)$, valeur indiquée par une flèche dans la figure 5.3 (gauche). C'est notamment pourquoi l'accord entre les deux énergies se dégrade quand la cassure augmente (c.-à-d. quand κ diminue). En effet, pour que leur écart relatif reste inférieur à 5% nous devons nous limiter aux extensions relatives $r/L \lesssim 5\%$, 12% ou 22% pour $\kappa = 90^\circ, 120^\circ$ ou 150° tandis qu'en l'absence de cassure ($\kappa = 180^\circ$) nous couvrions jusqu'à $r/L \lesssim 34\%$ (figure 2.6). De même, cet écart relatif atteint environ 25% en $r/L \simeq 17\%, 28\%, 40\%$ ou 51% pour $\kappa = 90^\circ, 120^\circ, 150^\circ$ ou 180° .

Pour finir, la régression linéaire produit l'accord quantitatif attendu, le « plus

¹¹En fait, l'expression des fluctuations $C(0; L)$ seule : l'énergie $\varepsilon_*(r/L)$ est connue (§ 2.2).

¹²Les boucles factices de longueur $\bar{L}(r, \kappa)$ pourraient aussi avoir une cassure $\bar{\kappa}(r, \kappa) \neq \kappa$.

¹³Par rapport à la vraie longueur L .

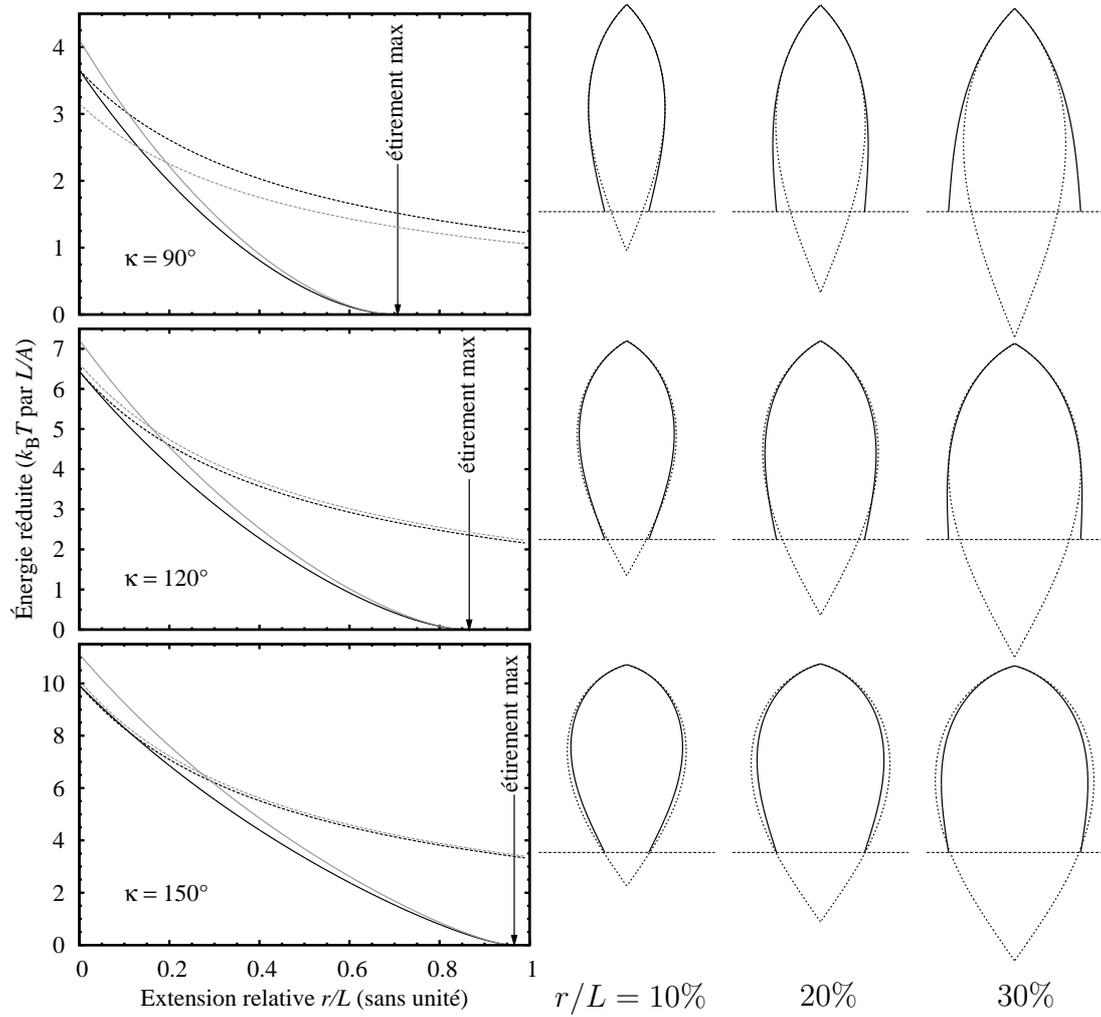


FIG. 5.3 – Droite : le point col de longueur L , d’extension r et de cassure κ à mi-longueur (ligne pleine) est qualitativement approché par une boucle factice (celle de point col également) de longueur $(L + 2r)$, identiquement cassée (pointillés). Gauche : les énergies réduites de la boucle $\frac{\varepsilon_*(0;\kappa)}{1+2r/L}$ et du point col $\varepsilon_*(r/L; \kappa)$ (pointillés et ligne pleine) s’accordent aux faibles extensions relatives r/L d’autant plus que la cassure est faible (κ grand). Voir le texte pour les détails. Sont repris *en gris*, la régression linéaire $\varepsilon_*(0; \kappa) \simeq \varepsilon_{\text{lin}}(\kappa)$ (pointillés) et le modèle du lignes & cercles (ligne pleine) de l’annexe E. Cette étude motive l’évaluation des fluctuations $C(r; L, \kappa)$ de $Q(r \vec{e}_r; L, \kappa)$ comme celles $C(0; L + 2r, 180^\circ)$ du $J(L)$ de S&Y.

mauvais » ajustement étant réalisé pour $\kappa = 90^\circ$, comme en figure 5.2. Dans l’expression de $Q(\vec{r}; L, \kappa)$ proposée, seules les fluctuations (5.3) sont approximées. L’énergie réduite $\varepsilon_*(r/L; \kappa)$ est en effet exactement connue. Nous venons de voir dans quelles mesures cette énergie s’accorde à celle de la boucle factice. Leur substitution n’apparaît alors pertinente que couplée à la régression linéaire $\varepsilon_*(0; \kappa) \simeq \varepsilon_{\text{lin}}(\kappa)$, elle même pertinente sur l’échelle de cassures $90^\circ \leq \kappa \leq 180^\circ$, afin d’obte-

nir pour Q une expression ne demandant ni résolution d'équations, ni évaluation d'intégrales elliptiques

$$\varepsilon_*(r/L; \kappa) \simeq \frac{\varepsilon_*(0; \kappa)}{1 + 2r/L} \simeq \frac{\varepsilon_{\text{lin}}(0; \kappa)}{1 + 2r/L}$$

soit $Q(r \vec{e}_r; L, \kappa) \simeq \frac{1}{A^3} C(0; L + 2r, \pi) \exp\left[-\frac{0.116 \times \kappa - 7.287}{(L + 2r)/A}\right]$. (5.4)

Afin de juger quantitativement les approximations sur lesquelles sont basées les formules (5.3) et (5.4), nous montrons maintenant comment adapter l'algorithme numérique de la section 4.2 afin qu'il prenne également en compte la cassure à mi-longueur.

5.3 Comment modifier l'algorithme numérique ?

Pour résumer, en section 4.2 nous avons proposé d'évaluer le facteur de cyclisation $J(L)$ d'un WLC de longueur L , en l'absence de force appliquée par un opérateur extérieur, à partir de la probabilité $\text{Prob}(r_B; L)$ qu'a sa distance bout à bout \vec{r} de se trouver dans une boule de rayon r_B . Par isotropie cette probabilité est simplement reliée à $P(z; L)$ sa distribution marginale en extension z , une composante quelconque de \vec{r} . Nous avons alors vu comment cette distribution se réduit à la TF de la fonction de partition du WLC, d'orientations limites quelconques et soumis à une force complexe $-ik_z/\beta \vec{e}_z$ qui correspond au multiplicateur de Lagrange fixant z . Finalement la fonction de partition s'interprète comme le propagateur du rotateur rigide quantique sous champ extérieur (§ 2.1). Son hamiltonien n'est diagonalisable qu'en champ nul, c'est pourquoi nous choisissons un calcul direct et numérique des éléments de matrice de son exponentielle. L'introduction de la cassure ne change que la *fonction de partition* dans le raisonnement précédent : essentiellement, l'isotropie n'est pas brisée comme nous travaillons toujours à force réelle nulle. Comme posé en section 5.2, le polymère est le collage en la cassure de deux WLC de mêmes longueurs de contour $L/2$ et de persistance A . Nous ne faisons pas d'autre hypothèse sur ses configurations afin de ne plus négliger les contributions entropiques, au contraire du point col (§ 5.2).¹⁴ Ce polymère n'est plus différentiable en $L/2$ où se produit la discontinuité

$$\begin{aligned} \Delta\varphi &= \varphi(L/2^+) - \varphi(L/2^-) = \varphi_+ - \varphi_- \\ \Delta\theta &= \theta(L/2^+) - \theta(L/2^-) = \theta_+ - \theta_- \end{aligned}$$

Au risque d'être un peu lourd, nous présentons des calculs détaillés dans la suite de cette section. Supposons dans un premier temps $\Omega_- = (\varphi_-, \theta_-)$ et $\Omega_+ = (\varphi_+, \theta_+)$ fixés. Le propagateur, dont la TF donne la distribution en extension z

¹⁴En particulier, ses configurations ne sont ni symétriques, ni planaires.

prenant en compte cette discontinuité, se calcule grâce au *principe de superposition* (annexe A)

$$Z(\Omega_L|\Omega_+; L/2, -ik_z/\beta) \times Z(\Omega_-|\Omega_0; L/2, -ik_z/\beta).$$

Vérifions le explicitement : nous devons convoluer les distributions en extension z , des deux WLC de longueur $L/2$ et dont les orientations limites sont – contrairement à la section 4.2 – spécifiées

$$\int_{-\infty}^{\infty} p_+(z-z') p_-(z') dz' = \int_{-\infty}^{\infty} dz' \left[\int_{-\infty}^{\infty} \frac{dk_2}{2\pi} e^{ik_2(z-z')} \zeta_+(-ik_2/\beta) \right] \times \\ \times \left[\int_{-\infty}^{\infty} \frac{dk_1}{2\pi} e^{ik_1 z'} \zeta_-(-ik_1/\beta) \right].$$

En réarrangeant les intégrations nous faisons en effet apparaître la TF, du produit des propagateurs annoncé plus haut

$$\dots = \int_{-\infty}^{\infty} \frac{dk_2}{2\pi} e^{ik_2 z} \zeta_+(-ik_2/\beta) \int_{-\infty}^{\infty} dk_1 \zeta_-(-ik_1/\beta) \underbrace{\int_{-\infty}^{\infty} \frac{dz'}{2\pi} e^{iz'(k_1-k_2)}}_{=\delta(k_1-k_2)},$$

où nous avons utilisé les abréviations $\zeta_+(-ik_z/\beta) = Z(\Omega_L|\Omega_+; L/2, -ik_z/\beta)$ et $p_+(z) = P(z; \Omega_L|\Omega_+, L/2)$.¹⁵ Cette expression considère encore des orientations limites Ω_0 et Ω_L particulières. Comme en section 2.1, la décomposition des propagateurs sur les harmoniques sphériques $\langle \Omega|\ell, m \rangle = Y_\ell^m(\Omega)$ permet de réécrire leur produit

$$\zeta_+(-ik_z/\beta) \times \zeta_-(-ik_z/\beta) = \sum_{\ell, m} \sum_{\ell', m'} Y_\ell^m(\Omega_L) \tilde{Y}_{\ell'}^{m'}(\Omega_0) \langle \ell, m | \hat{\Delta} | \ell', m' \rangle$$

où nous avons temporairement posé¹⁶

$$\hat{\Delta} = \exp \left[-\frac{L/2}{A} \hat{H}(-ik_z/\beta) \right] |\Omega_+ \rangle \langle \Omega_- | \exp \left[-\frac{L/2}{A} \hat{H}(-ik_z/\beta) \right].$$

Comme en section 2.1, la sommation sur les orientations limites sélectionne¹⁷ les composantes $|0, 0 \rangle$ de $\hat{\Delta}$

$$Z_{\Omega_+, \Omega_-}(L, -ik_z/\beta) = \int \frac{d\Omega_0 d\Omega_L}{4\pi} \zeta_+(-ik_z/\beta) \times \zeta_-(-ik_z/\beta) = \\ = \langle 0, 0 | \exp \left[-\frac{L/2}{A} \hat{H}(-ik_z/\beta) \right] |\Omega_+ \rangle \langle \Omega_- | \exp \left[-\frac{L/2}{A} \hat{H}(-ik_z/\beta) \right] |0, 0 \rangle$$

¹⁵Réciproquement p_- et ζ_- correspondent aux orientations limites Ω_- et Ω_0 .

¹⁶La notation $|\Omega \rangle$, pour le moment évitée dans le texte principal, est expliquée en annexe A. Elle nous permet de procéder par étape et ainsi éviter d'écrire un quadruple développement sur les harmoniques sphériques.

¹⁷Reconnaitre $Y_0^0(\Omega) = \frac{1}{\sqrt{4\pi}}$ dans $\int \frac{d\Omega_0 d\Omega_L}{4\pi}$, puis utiliser l'orthonormalité des $Y_\ell^m(\Omega)$ (§ B.2).

Les éléments de matrices du hamiltonien \widehat{H} (§ 2.1) montrent que

$$\exp\left[-\frac{L/2}{A}\widehat{H}(-ik_z/\beta)\right]|0,0\rangle = \sum_{\ell} \langle \ell, 0 | \exp\left[-\frac{L/2}{A}\widehat{H}(-ik_z/\beta)\right]|0,0\rangle \times |\ell, 0\rangle,$$

c'est-à-dire ne possède des composantes *a priori* non nulles que pour $m = 0$.¹⁸ Pour les orientations Ω_{\pm} en la cassure, la décomposition sur les harmoniques sphériques se réduit donc à

$$\begin{aligned} Z_{\Omega_+, \Omega_-}(L, -ik_z/\beta) &= \frac{1}{4\pi} \sum_{\ell'} \sum_{\ell} \mathbb{P}_{\ell'}(\cos \theta_+) \mathbb{P}_{\ell}(\cos \theta_-) \sqrt{(2\ell' + 1)(2\ell + 1)} \times \\ &\times \langle 0, 0 | \exp\left[-\frac{L/2}{A}\widehat{H}(-ik_z/\beta)\right] |\ell', 0\rangle \langle \ell, 0 | \exp\left[-\frac{L/2}{A}\widehat{H}(-ik_z/\beta)\right] |0, 0\rangle, \end{aligned}$$

avec les polynômes de Legendre $\mathbb{P}_{\ell}(\cos \theta) = Y_{\ell}^0(\Omega)/\sqrt{\frac{2\ell+1}{4\pi}}$ (§ B.2). Pour finir, si le hamiltonien n'est pas hermitien il est en revanche symétrique, et par conséquent

$$\langle 0, 0 | \exp\left[-\frac{L/2}{A}\widehat{H}(-ik_z/\beta)\right] |\ell', 0\rangle = \langle \ell', 0 | \exp\left[-\frac{L/2}{A}\widehat{H}(-ik_z/\beta)\right] |0, 0\rangle.$$

Le propagateur, indépendant des orientations limites Ω_0 et Ω_L mais avec celles en la cassure Ω_{\pm} arbitrairement fixées, prend donc la forme

$$Z_{\Omega_+, \Omega_-}(L, -ik_z/\beta) = \frac{1}{4\pi} \sum_{\ell', \ell} \underbrace{\mathbb{P}_{\ell'}(\cos \theta_+) \mathbb{P}_{\ell}(\cos \theta_-)}_{\text{cassure}} \underbrace{\mathbb{Z}_{\ell, \ell'}(L/2, -ik_z/\beta)}_{\text{configurations}}$$

où nous avons défini la matrice

$$\begin{aligned} \mathbb{Z}_{\ell, \ell'}(L/2, -ik_z/\beta) &= \sqrt{(2\ell + 1)(2\ell' + 1)} \langle \ell, 0 | \exp\left[-\frac{L/2}{A}\widehat{H}(-ik_z/\beta)\right] |0, 0\rangle \times \\ &\times \langle \ell', 0 | \exp\left[-\frac{L/2}{A}\widehat{H}(-ik_z/\beta)\right] |0, 0\rangle. \end{aligned}$$

Elle est symétrique et indépendante des orientations Ω_{\pm} en la cassure. Elle est d'autre part la seule source, dans le propagateur, de dépendance en le multiplicateur de Lagrange k_z fixant l'extension z du polymère, ainsi qu'en sa longueur L . Pour Ω_{\pm} , la dépendance est uniquement en θ_{\pm} *via* les polynômes de Legendre. Insistons enfin sur la présence des composantes $|\ell \neq 0, 0\rangle$, au contraire de l'expression type amplitude du vide précédemment obtenue en section 2.1. Elles traduisent le choix d'orientations particulières en la cassure, contrairement aux orientations limites Ω_0 et Ω_L dont nous nous sommes affranchis.

Nous inspectons maintenant comment imposer de manière systématique une amplitude κ de cassure à mi-longueur. La discontinuité des orientations Ω_{\pm} en la

¹⁸En d'autres termes $|\ell, m\rangle$ est stable en m sous l'action de la sous matrice $H_{\ell, \ell'}^{(m)}$.

cassure (voir $\Delta\varphi$ et $\Delta\theta$ ci-dessus) est fixée de la manière la plus générale au moyen du produit scalaire des vecteurs tangents \vec{t}_{\pm} correspondants¹⁹

$$\cos \Delta\theta - (1 - \cos \Delta\varphi) \sin(\theta_-) \sin(\theta_- + \Delta\theta) = \cos(\pi - \kappa).$$

Le propagateur ci-dessus est indépendant de φ_{\pm} et pour le point col (§ 5.2) nous avons décrit une cassure plane. Cela suggère de se réduire à la contrainte

$$\Delta\varphi = 0 \quad \text{et} \quad |\Delta\theta| = \pi - \kappa.$$

Sa prise en compte produit le propagateur que nous cherchions en fait : indépendant des orientations limites Ω_0 et Ω_L , et prenant en compte l'ensemble des orientations en la cassure Ω_{\pm} réalisant l'amplitude κ

$$Z_{\kappa}(L, -ik_z/\beta) = \sum_{\ell', \ell} \mathbb{C}_{\ell', \ell}(\kappa) \mathbb{Z}_{\ell, \ell'}(L/2, -ik_z/\beta).$$

C'est la trace du produit de la matrice \mathbb{Z} contenant l'information configurationnelle des deux WLC de longueur $L/2$ collés en la cassure, par la matrice \mathbb{C} qui en spécifie l'amplitude κ .²⁰ L'écriture de \mathbb{C} est lourde bien que son contenu soit simple : c'est l'intégration des polynômes de Legendre, sur les Ω_{\pm} respectant la contrainte qui fixe κ . Comme ces polynômes ne dépendent pas de φ_{\pm} , nous intégrons au préalable $\int_0^{2\pi} d\varphi_- \int_0^{2\pi} d\varphi_+ \delta(\varphi_+ - \varphi_-) = 2\pi$. Nous écrivons le résultat sous la forme condensée

$$\mathbb{C}_{\ell, \ell'}(\kappa) = \frac{1}{2} \int_{-1}^1 dx_- \int_{-1}^1 dx_+ \delta(x_+, x_-, \kappa) \mathbb{P}_{\ell}(x_-) \mathbb{P}_{\ell'}(x_+)$$

avec la contrainte (symétrisée) sur $x_{\pm} = \cos \theta_{\pm}$

$$\begin{aligned} \delta(x_+, x_-, \kappa) \equiv & \delta \left\{ \cos(\theta_+) - \cos[\theta_- - (\pi - \kappa)] \right\} \\ & + \delta \left\{ \cos(\theta_+) - \cos[\theta_- + (\pi - \kappa)] \right\}. \end{aligned}$$

Nous conservons temporairement la déflexion $c = (\pi - \kappa)$ produite par la cassure plutôt que son amplitude κ pour le reste du calcul.²¹ Après intégration sur x_+ il reste deux intégrales du type recouvrement des $\mathbb{P}_{\ell}(x)$.

$$\begin{aligned} \mathbb{C}_{\ell, \ell'}(\pi - c) &= \frac{1}{2} \left[R_{\ell, \ell'}(c) + R_{\ell, \ell'}(-c) \right] \\ R_{\ell, \ell'}(c) &= \int_{-1}^1 \mathbb{P}_{\ell}(\cos \theta_-) \mathbb{P}_{\ell'}[\cos(\theta_- - c)] d(\cos \theta_-). \end{aligned}$$

Sachant que $\mathbb{P}_{\ell}(-x) = (-1)^{\ell} \mathbb{P}_{\ell}(x)$ (§ B.2) le calcul d'une de ces intégrales seulement est nécessaire. Pour voir cela, remarquons que l'intégration sur $-1 \leq x_- \leq 1$

¹⁹Nous utilisons l'expression de § A.2, entre \vec{t}_n et \vec{t}_{n+1} .

²⁰Une expression similaire se retrouve dans les travaux de Yan et coll. [YM03].

²¹Les résultats sont un peu plus simples à résumer (voir plus bas).

provient de celle sur $0 \leq \theta_- \leq \pi$ et donc $\sin \theta_- = (1 - x_-^2)^{1/2} > 0$. Dans les $R_{\ell, \ell'}$ nous pouvons donc écrire

$$\begin{aligned} \cos(\theta_- - c) &= x_- \cos c + (1 - x_-^2)^{1/2} \sin c \\ \cos(\theta_- + c) &= x_- \cos c - (1 - x_-^2)^{1/2} \sin c \\ &= - \{ (-x_-) \cos c + [1 - (-x_-)^2]^{1/2} \sin c \} \end{aligned}$$

et en posant $y = -x_-$ nous voyons alors que

$$\begin{aligned} R_{\ell, \ell'}(-c) &= \int_{-1}^1 (-1)^\ell \mathbb{P}_\ell(y) (-1)^{\ell'} \mathbb{P}_{\ell'}[y \cos c + (1 - y^2)^{1/2} \sin c] dy \\ &= (-1)^{\ell + \ell'} R_{\ell, \ell'}(c). \end{aligned}$$

Nous obtenons donc la règle de sélection suivante

$$\mathbb{C}_{\ell, \ell'}(\kappa) = \begin{cases} R_{\ell, \ell'}(\pi - \kappa) & \text{quand } \ell \text{ et } \ell' \text{ sont de même parité,} \\ 0 & \text{sinon.} \end{cases}$$

L'intégrale de recouvrement se calcule analytiquement, en explicitant les polynômes de Legendre, pour quelques valeurs particulières de ℓ et ℓ' . À mesure qu'ils augmentent, cette méthode devient toutefois rapidement difficile à implémenter. Nous évaluons en fait la majorité des éléments de \mathbb{C} numériquement, en prenant soin de diviser l'intervalle d'intégration entre les $(\ell + \ell')$ zéros du produit des polynômes de Legendre, qui est en effet d'autant plus rapidement oscillant que ℓ ou ℓ' augmente. Notons que nous ne devons faire ce travail qu'une fois par amplitude κ car \mathbb{C} ne dépend que de la cassure.

Nous évaluons d'autre part $\mathbb{Z}_{\ell, \ell'}(L, -ik_z/\beta)$ en calculant au moyen d'Expokit [Sid98]

$$\exp \left[-\frac{L/2}{A} \widehat{H}(-ik_z/\beta) \right] |0, 0\rangle$$

dans la base des harmoniques sphériques $|\ell, m\rangle$ comme en section 4.2. Pour chaque valeur de k_z nous ne sommes alors intéressés que par les composantes $|\ell, 0\rangle$. En pratique nous devons tronquer le hamiltonien, mais nous choisissons la coupure ℓ_{\max} en s'assurant au préalable de la convergence du résultat. Comme en section 4.2 nous devons augmenter ℓ_{\max} si nous voulons descendre à de faibles longueurs réduites $L/A \ll 1$. Nous avons pu couvrir $0.3 \leq L/(2A) \leq 1.7$, c'est-à-dire $100 \lesssim L \lesssim 500$ bp en nous limitant à $\ell_{\max} = 100$ aux plus faibles longueurs²²; coupure pour laquelle le temps de calcul reste raisonnable. Le lecteur attentif aura remarqué que la composante $|0, 0\rangle$ peut être réutilisée pour le cas sans cassure, en interprétant ci-dessus $L/(2A)$ comme L/A .²³ En section 4.3 nous

²²Nous avons ponctuellement poussé ce test de convergence jusqu'à $\ell_{\max} = 250$, sans observer de différence.

²³L'inverse n'est pas possible : en l'absence de cassure (§ 4.2) nous n'avons stocké que la première composante $|0, 0\rangle$.

pouvions donc tracer le rapport $K(r; L)$ jusqu'à $L/A \geq 0.3$, soit $L \gtrsim 50$ bp. Nous nous sommes en fait volontairement restreint à l'intervalle $100 \lesssim L \lesssim 500$ bp afin de conserver une présentation unitaire du facteur de cyclisation le long de ce manuscrit.²⁴ Cela ne nous a pas empêché d'utiliser ces données, par exemple pour la boucle $O_1O_3 = 76$ bp de l'opéron Lac (§ 4.4). Rappelons pour finir que la matrice \mathbb{Z} est symétrique et indépendante de l'amplitude κ de la cassure. Elle ne doit donc être échantillonnée qu'une fois par longueur L .

L'échantillonnage en k_z se fait en vue de la FFT du propagateur $Z_\kappa(L, -ik_z/\beta)$, pour laquelle nous utilisons les propriétés suivantes de la distribution $P(z; L, \kappa)$, qui est : réelle, paire et assurément nulle pour des extensions $z \geq L \sin(\kappa/2)$. Comme pour le cas sans cassure (4.2), les deux premières relèvent d'une formulation convenable du problème visant à une utilisation optimale de l'algorithme de FFT (§ F.1). La dernière permet en théorie de gagner en temps de calcul par rapport au cas sans cassure en diminuant, sans danger de repliement (§ F.1), la fréquence de coupure $L \sin(\kappa/2) \leq L$. Cela représente toutefois un gain négligeable : l'étape limitante reste le calcul de l'exponentielle de matrice. En pratique nous avons donc utilisé L , dont le doublement s'est là encore montré superflu (4.2). Nous avons également conservé l'échantillonnage produisant 2048 points sur $0 \leq z \leq L$.

5.4 Importance de mécanismes entraînant une perte de rigidité

Comme en section 4.2 nous pouvons alors calculer $P(z; L, \kappa)$ sans limitation dans la longueur L , pour diverses amplitudes κ de la cassure. Nous pouvons en déduire $S(r; L, \kappa)$: l'étude de ces deux distributions montre à longueur fixée, une perte de rigidité consécutive à l'introduction de la cassure (diminution de κ). Les comportements observés apparaissent raisonnables bien qu'il n'existe pas (à notre connaissance) dans la littérature, de résultat systématique (analytique ou numérique) prenant en compte l'effet d'une telle cassure (ou même de plusieurs cassures, § 5.1) pour ces deux grandeurs [DvML⁺05]. À ce niveau nous faisons donc confiance à notre algorithme, précédemment éprouvé dans le cas sans cassure (§ 4.2) qui est par ailleurs reproduit en prenant $\kappa = 180^\circ$. C'est pourquoi nous ne montrons pas ici ces grandeurs pas plus que nous ne tentons une étude de l'effet de la cassure, au moyen d'une longueur de persistance effective par exemple. Nous concentrons notre attention sur le facteur de cyclisation (figure 5.4). La perte de rigidité entraînée par l'introduction de la cassure, favorise globalement la formation de boucles. Cet effet est toutefois nettement plus prononcé

²⁴Si nous étions descendu jusqu'à 50 bp, pourquoi ne pas le faire également pour la cassure, c'est-à-dire $L/(2A) = 0.15$, etc.

aux faibles longueurs $L \lesssim A$. La figure 5.4 (haut/gauche) montre une augmentation d'à peu près un ordre de grandeur du rapport $K(10 \text{ nm}; 100 \text{ bp}, \kappa)$ par 10° de cassure supplémentaire. Plus précisément, cela est vrai jusqu'à $\kappa \gtrsim 120^\circ$: ensuite, l'effet reste significatif mais apparaît moins important. Nous avons vu

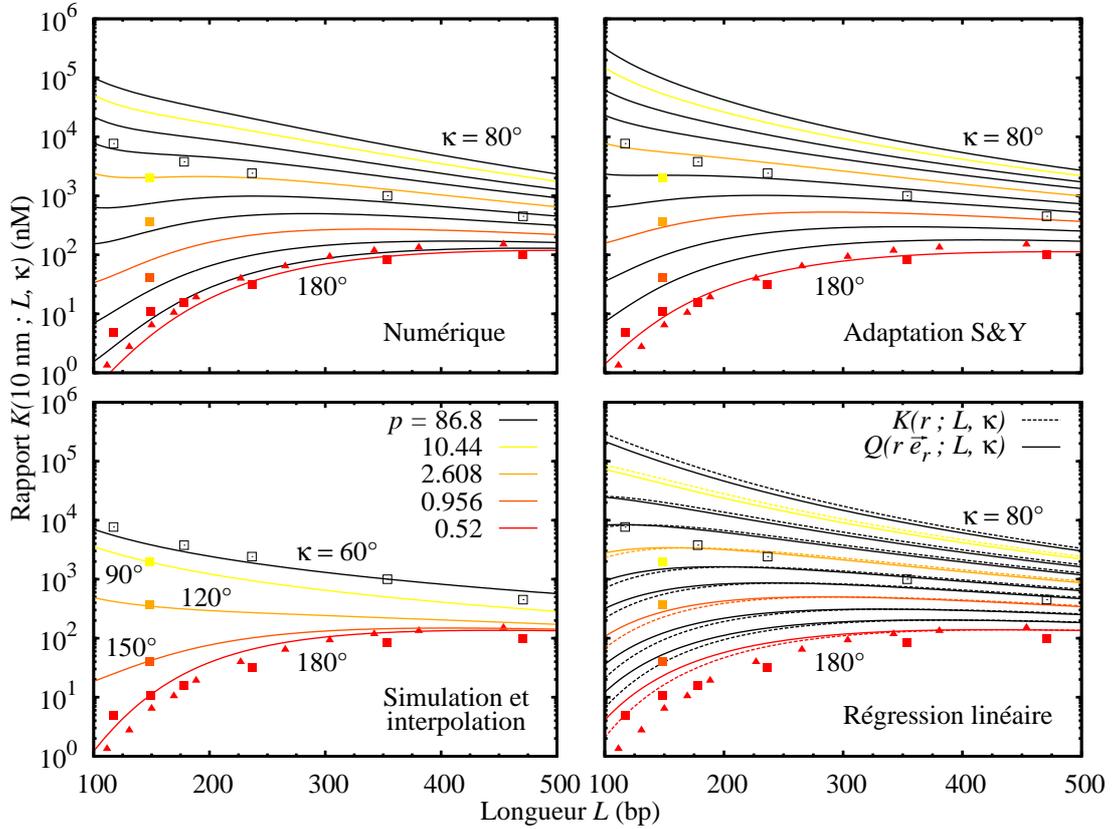


FIG. 5.4 – L'introduction d'une cassure à mi-longueur favorise la formation de boucles. Les cassures $\kappa = 180^\circ, 150^\circ, 120^\circ$ et 90° précédemment étudiées sont différenciées par leur couleur allant resp. du rouge au jaune. Haut/gauche : calcul numérique (ligne pleine) du rapport K pour $80^\circ \leq \kappa \leq 180^\circ$, par pas de 10° ($r = 10 \text{ nm}$). Comparaison aux données (reprises dans les autres figures) de dynamique brownienne (\square) disponibles pour $60^\circ \leq \kappa \leq 180^\circ$, par pas de 30° (même r), et de Monte Carlo en l'absence de cassure (\blacktriangle). Un accord quantitatif avec les premières n'est obtenu qu'en l'absence de cassure, mais les comportements sont similaires quand κ diminue. Bas/gauche : reproduction de l'ajustement (3.1) à ces données, réalisé par Rippe. Haut/droite : adaptation (5.3) de S&Y. Accord quantitatif pour une large gamme de cassures, dès que $L \gtrsim 300$. Bas/droite : l'adaptation (5.4) de S&Y produit un accord quasi-identique. L'intégration de la distribution Q (ligne pleine) pour calculer K (pointillés) peu raisonnablement être omise.

en section 4.3 que la courbure d'une telle longueur (100 bp) est difficile surtout de par sa rigidité élastique, les effets d'origine entropique ne jouant qu'un rôle

secondaire²⁵. Parmi les configurations de « boucle » possibles c'est alors celle de point col qui domine. Sa courbure est dans un premier temps facilitée par une fermeture non rigoureuse de la boucle ($r \gtrsim 0$). Nous avons alors constaté en section 5.2 comment, à distance bout à bout fixée (typiquement 10 nm), l'introduction d'une cassure peut d'autant plus faciliter cette courbure.

Les simulations de dynamique brownienne réalisées par Merlitz et coll. permettent également de considérer des cassures, comme nous fixes en amplitude et position [RvHL95, MRKL98, KML98, Rip01]. Quelques uns de leurs résultats sont publiés, pour le cas d'une cassure à mi-longueur d'amplitude $\kappa = 60^\circ$ et un rayon de boucle $r = 10$ nm. L'ajustement réalisé par l'un des auteurs (Rippe), déjà reproduit en section 4.3 pour l'effet de la taille finie des protéines fixant la boucle ($r \geq 0$), permet également de prendre en compte celui de cette cassure ($\kappa \leq 180^\circ$). Le même paramètre p est utilisé dans la formule (3.1). Nous reproduisons tels quels les ajustements qu'il propose, l'article [Rip01] ne montrant malheureusement les données de dynamique brownienne²⁶ qu'en l'absence de cassure ou bien quand $\kappa = 60^\circ$. L'accord quantitatif obtenu pour le cas sans cassure (§ 4.3) se dégrade quand cette dernière augmente (c'est-à-dire κ diminue), comme le montre la figure 5.4 (bas/gauche). L'écart constaté ne devient toutefois significatif (disons un ordre de grandeur) qu'à partir de $\kappa \lesssim 90^\circ$. Il doit être pondéré par la diminution de l'effet de la cassure, constaté à partir de $\kappa \lesssim 120^\circ$ et qui se produit également ici. Rappelons enfin que le modèle de Merlitz et coll. prend en compte la rigidité de torsion de l'ADN ainsi que sa nature de polyélectrolyte, toute deux censées rendre plus contraignant sa courbure. La majoration que nous constatons, de leurs données par nos résultats, est donc raisonnable. Nous n'avons pas soulevé ce point lors de l'étude du cas sans cassure (§ 4.3), où un croisement des données s'opère pourtant en dessous de $L \lesssim 250$ bp pour $r = 10$ nm ($L \lesssim 150$ bp pour $r = 1$ nm) ce qui est critiquable au premier abord. Soulignons néanmoins que ce même croisement est observé²⁷ avec le résultat original de Shimada et Yamakawa (§ 2.4), autour de $L \lesssim 150$ bp, qui n'inclut pas non plus d'effets de torsion ou d'interaction électrostatique. Même constat pour les simulations de Monte Carlo ($r = 10$ nm) réalisées par Podtezhnikov et Vologodskii [PV00] avec lesquelles nos résultats sont en bon accord. Nous reproduisons ces données dans les figures 5.4 (droite/haut et bas), principalement afin de guider l'oeil du lecteur. Une comparaison qualitative aux récents résultats obtenus par Yan, Kawamura et Marko [YKM05] est également possible, bien qu'ils traitent le cas de cassures semi-flexibles (dans notre cas κ est fixe, maintenant une cassure rigide). Pour une unique cassure centrale, de rigidité identique au reste de l'ADN et d'amplitude « au repos » 90° c'est alors

²⁵D'autant plus que la séquence est encore raccourcie

²⁶C'est-à-dire plus d'un point, afin d'évaluer la fiabilité de l'ajustement.

²⁷Voir les figures 4.4 (gauche/bas et milieu) ou la figure 1-A de [MRKL98].

leur facteur de cyclisation qui majore notre résultat, et ce malgré une fermeture rigoureuse de la boucle ($r = 0$, contre 10 nm pour nous).²⁸

Nous étudions maintenant l'expression (5.3) proposée pour Q en section 5.2. Pour la borne inférieure de la figure 5.4 (droite/haut), c'est-à-dire $L = 100$ bp nous n'attendons qu'un accord qualitatif aux résultats numériques, particulièrement en l'absence de cassure (§ 4.3) étant donné l'extension typique de $r = 10$ nm utilisée ($r/L = 30\%$). Toutefois la perte de rigidité entraînée par l'introduction de la cassure suggère intuitivement une amélioration de cette approximation (comparer p. ex. les configurations des figures 5.3 et 2.5 à r/L fixée). Une telle diminution avec κ de l'écart aux résultats numériques est en effet observable, mais l'accord reste qualitatif. Il devient quantitatif aux plus grandes longueurs (faibles r/L) : dès $L = 300$ bp la différence est minime pour $\kappa \lesssim 120^\circ$. À $L = 500$ bp, ce constat s'étend à l'ensemble des cassures envisagées. En figure 5.3 (droite/bas) nous avons testé l'expression (5.4) utilisant la régression linéaire pour l'énergie de la boucle factice. L'intégration nécessaire au calcul du rapport K (pointillés) n'apparaît alors pas indispensable : la distribution Q (ligne pleine) fournit une estimation fiable pour de larges échelles de longueur L et de cassure κ .²⁹ Nous pouvons donc calculer le facteur de cyclisation, en fait K ou Q , pour une extension $r \gtrsim 0$ et une cassure $90^\circ \lesssim \kappa \leq 180^\circ$ au moyen de la formule analytique (5.4).

Un effet particulièrement intéressant produit par la cassure est la réduction de la longueur de « boucle » (rappelons que $r = 10$ nm) la plus probable. En l'absence de cassure elle est d'environ 500 bp, mais pour une cassure de 120° elle n'est plus que d'à peu près 200 bp, comme le montre la figure 5.5 (gauche). L'adaptation (5.3) de S&Y (symboles \circ) reproduit qualitativement le comportement du calcul numérique (symboles \square). Notons qu'à partir de 120° (130° pour S&Y) nous ne pouvons plus discerner cette longueur « la plus probable », de l'évènement associé à la configuration la plus étirée $L = 30/\sin(\kappa/2)$ bp (§ 5.2). C'est alors cette borne inférieure (polymère inextensible) du rapport K qui constitue la longueur de boucle la plus probable (« transition » symbolisée par une flèche sur la figure). Comme nous l'avons remarqué en section 5.2 (figure 5.3 p. ex.), cette limitation ne concerne pas la boucle factice de longueur $(L + 2r)$. Pour $\kappa \lesssim 120^\circ$, la longueur la plus probable associée doit alors être considérée avec prudence. Toujours sur la figure 5.5 (gauche), l'omission de l'intégration de Q (ligne pleine) permettant d'obtenir K (symboles \blacktriangle) ne constitue qu'une approximation mineure.

Au rapport $K(10 \text{ nm}; L, \kappa)$ de la figure 5.4 nous pouvons associer, comme en section 4.4, une énergie libre de cyclisation. Elle est montrée comme fonction

²⁸Nous ne reproduisons pas leur figure 6 (carrés pleins et vides) : sans cassure leur $J(L)$ ne coïncide avec celui de S&Y exprimé en M, qu'après division par un facteur 10. Nous n'avons pas réussi à rétablir intelligiblement les unités. Le lecteur fera attention à une erreur de ce type dans [SM05b] (figure 4) où $J(L)$ est en nm^{-3} et non en M.

²⁹Incluant $\kappa = 80^\circ$, bien que la régression ait été réalisée sur $90^\circ \leq \kappa \leq 180^\circ$.

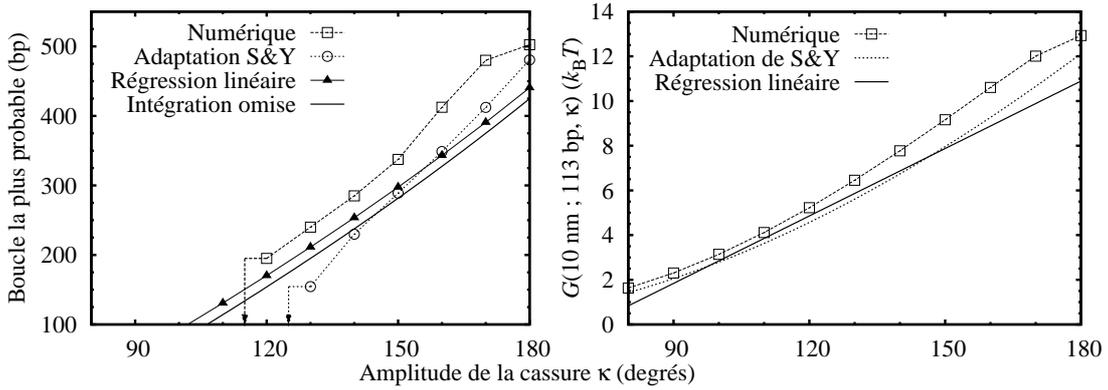


FIG. 5.5 – Effet de la cassure sur la longueur de boucle la plus probable (gauche) calculée pour $r = 10$ nm, et l'énergie libre de cyclisation d'une séquence de 113 bp (droite) identique à celle impliquée dans la régulation de l'opéron Gal. La boucle la plus probable diminue de longueur (erreur $\lesssim 10$ bp, plus petite que \square) avec κ , jusqu'à devenir la plus étirée de longueur ($30 \text{ bp} / \sin(\kappa/2)$) (« transition » symbolisée par une flèche). Les $12 k_B T$ mesurés par Lia et coll. correspondent à une très faible cassure ($\kappa \simeq 170^\circ$) induite par la protéine HU (voir texte). Notons que l'énergie libre déduite de (5.4), en omettant l'intégration, reste une fonction linéaire de κ (ligne pleine).

de la longueur L de la « boucle » ($r = 10$ nm), pour diverses amplitudes κ de cassure, en figure 5.6. Nous allons illustrer nos résultats au moyen de la répression de l'opéron Gal de la bactérie *E. coli*, déjà évoqué en section 1.1. Similairement au cas de l'opéron Lac, l'expression des protéines réalisant le métabolisme du sucre galactose peut être stoppée à l'étape de transcription en formant une boucle $O_E O_I = 113$ bp, comprenant les deux promoteurs de l'opéron. Les opérateurs O_E et O_I sont là encore les courtes séquences, auxquelles le répresseur GalR s'accroche spécifiquement fixant ainsi la boucle.³⁰ Toutefois, contrairement au LacR, les deux sous-unités (dimères) formant le GalR (tétramère) interagissent faiblement [MH98, GA02]. Comme la séquence à cycliser est courte (moins d'une longueur de persistance), la répression de l'opéron par le GalR seul est peu efficace au regard des résultats de la section 4.3. Il a été démontré *in vitro* que l'ajout de la protéine HU (pour « Heat Unstable ») permettait de retrouver une répression efficace [LBC⁺03]. Cette protéine, présente dans *E. coli* en un grand nombre de copies, est connue pour s'accrocher à la séquence d'ADN de manière non spécifique [SLZR03]. Les cassures ainsi provoquées peuvent produire une condensation significative de l'ADN, suggérant un rôle analogue aux histones des chromosomes chez les eucariotes [DG02, vNVG⁺04].³¹ Parmi les scénarii

³⁰C'est aussi un produit du métabolisme du galactose qui induit le changement configurationnel (sans consommation d'ATP) empêchant l'accrochage du GalR.

³¹Cela est vrai à faible concentration en HU. À forte concentration, la molécule peut être entièrement recouverte et se trouver au contraire rigidifiée en un bâtonnet. Pour la répression

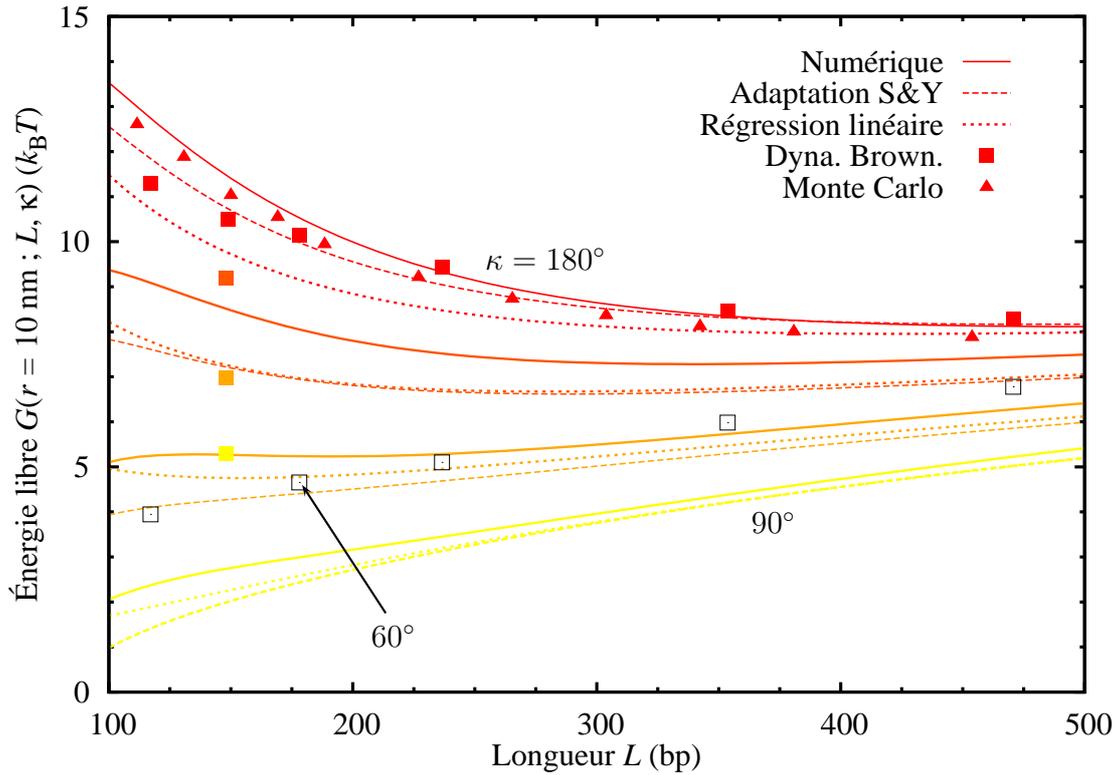


FIG. 5.6 – Variation avec la longueur L , de l'énergie libre associée aux configurations d'extensions se trouvant dans une boule de rayon $r = 10$ nm. Nous considérons à mi-longueur une cassure $90^\circ \leq \kappa \leq 180^\circ$, par pas de 30° (du jaune au rouge respectivement). L'énergie se déduit (en $k_B T$) de $-\ln [K(r; L) \times V(r)]$ avec le volume $V(r) = \frac{4}{3}\pi r^3$. Nous reprenons les données de la figure 5.4 : notre évaluation numérique (lignes pleines), l'adaptation (5.3) de S&Y (tirets) et son approximation supplémentaire (5.4) au moyen de la régression linéaire proposée en section 5.2 (intégration omise, pointillés). Sont également reproduites les simulations de dynamique brownienne (\square) et de Monte Carlo (\triangle).

imaginés d'interactions entre le GalR, HU et le brin d'ADN, [SVA05, SGLA02, GVZA01, VLK⁺03, ACA96, LSAA97, KA01, AA97, LGA99] il est maintenant admis le suivant : HU s'accroche sur la séquence³² entre les deux opérateurs, ce qui provoque une cassure rendant la cyclisation moins difficile et l'interaction des dimères du GalR suffisante. Il n'existe pas de résultats systématiques sur la rigidité du complexe HU/ADN, mais de sérieux arguments ont été avancés pour une grande flexibilité [vNVG⁺04]. La cassure que nous avons introduite est certes rigide, mais pourquoi ne pas exploiter nos résultats ? Avec un dispositif du même type que celui schématisé en figure 4.1 (médaillon), Lia, Bensimon, Cro-

de l'opéron Gal nous considérons la situation idéale d'une protéine HU par molécule d'ADN.

³²HU s'ajuste dans une bulle de dénaturation dont l'ouverture est facilitée par un surenroulement négatif de la séquence.

quette, Allemand, Dunlap, Lewis, Adhya et Finzi ont mesuré la force critique de 0.88 pN à laquelle la molécule d'ADN passe en moyenne autant de temps cyclisée que non [LBC⁺03]. Notons que la différence d'extension observée entre ces deux états est supérieure à la longueur de la boucle 55 nm > 37.7 nm ; mais nous reviendrons sur ce point dans le chapitre 6. Ils en déduisent une énergie libre de cyclisation d'environ $12 k_B T$ ce qui correspond pour nous à une très faible cassure ($\kappa \simeq 170^\circ$), comme le montre la figure 5.5 (droite). Cette estimation peut changer si la taille typique du GalR se révèle être inférieure à 10 nm, ou encore que nous incluons la torsion de l'ADN.

L'idée d'attribuer une fonction à la perte de rigidité (cassures) associée à l'accrochage des protéines n'est pas nouvelle en biologie. Nous disposons d'un large panel de protéines, qui ne cesse de grandir, dont nous connaissons bien certaines propriétés d'accrochage comme leur spécificité en sites, leur affinité, etc. Pour la plupart d'entre elles nous avons également de sérieux indices à propos de la possible fonction associée à leur accrochage. Nous en avons donné une liste succincte en section 1.5. En revanche, force est de constater que nous manquons de données systématiques sur le complexe protéine/ADN, au sujet par exemple de sa rigidité élastique ou de l'angle préférentiel qu'il peut induire. Dans ces conditions notre modélisation du phénomène biologique ne peut être qu'effective.

Au niveau théorique, les modèles remettant en cause la linéarité de l'élasticité du WLC (§ 5.1) semblent s'accorder sur la domination, aux faibles longueurs $L \lesssim A$, des configurations de boucles ne présentant qu'un seul défaut.³³ Ce constat a d'abord été fait par Wiggins et coll. [WPN05], puis il s'est rétrospectivement confirmé sur les travaux de Yan et coll. [YKM05]. En variant la position de ce défaut dans le voisinage de la mi-longueur, Ranjith et coll. [RSKM05] ont de plus confirmé l'intuition selon laquelle il s'agit bien de l'endroit le plus favorable pour son apparition, lors d'une cyclisation.³⁴ Bien que simpliste à première vue, la situation que nous avons considérée s'avère finalement une bonne première approche du problème.

³³Pour des orientations limites quelconques, ou encore des boucles en forme d'épingle à cheveux. Pour un « cercle », c.-à-d. les conditions limites pertinentes pour les expériences de cyclisation, ce sont les situations à deux défauts qui dominent aux mêmes longueurs.

³⁴Merlitz et coll. [MRKL98] mentionnent également une différence d'un ordre de grandeur pour le facteur de cyclisation d'un polymère de $L = 150$ bp présentant une amplitude $\kappa = 100^\circ$, selon que la cassure se trouve en $L/2$ ou $L/5$ (mais ils ne montrent pas leurs données).

Chapitre 6

Contrôle du brin d'ADN par une force extérieure

Dans les dispositifs expérimentaux permettant la manipulation de molécules d'ADN uniques, le brin est le plus souvent contrôlé au moyen d'une force extérieure appliquée par l'opérateur. Un dispositif typique ressemble à celui schématisé en figure 4.1 (médaillon), mais le substrat ainsi que le dispositif permettant d'appliquer la force peuvent varier. Pour la fin de cette étude nous pouvons conserver notre vue schématique, mais le lecteur intéressé pourra consulter les revues [BMW00, SDC⁺03] écrites par Bustamante, Macosko et Wuite (2000) ou par Strick, Dessinges, Charvin, Dekker, Allemand, Bensimon et Croquette (2003). Grâce à un tel montage Finzi et Gelles [FG95] furent les premiers (1995) à observer de manière directe la formation d'une boucle dans une molécule d'ADN (au moyen d'un LacR). Leur travail contient la mesure des temps moyens passés par l'ADN dans les états cyclisé ou non, en l'absence de force extérieure. En 2003, Lia et coll. [LBC⁺03] ont repris le même type d'étude avec force, dans le cadre de la répression de l'opéron Gal. Ces quelques remarques motivent la prise en compte dans notre algorithme d'une force extérieure f non nulle. Pour simplifier, nous revenons au cas sans cassure $\kappa = 180^\circ$ (omis dans les expressions).

6.1 Apparition d'une direction privilégiée

Quand le polymère est étiré par une force extérieure uniforme $f \vec{e}_z$, le système perd sa symétrie sphérique (isotropie). L'apparition d'une direction privilégiée portée par la force ne laisse en effet au système que la symétrie de révolution autour de cette direction, c'est-à-dire cylindrique. Notre méthode utilise la symétrie sphérique pour calculer la distribution isotrope $S(r; L, f = 0)$ en distance bout à bout à partir de celle, marginale, en extension $z : P(z; L, f = 0)$. Nous pouvons

facilement calculer cette dernière à force non nulle¹

$$\begin{aligned} P(z; L, f) &= \frac{1}{Z(L, f)} \int_{-\infty}^{\infty} \frac{dk_z}{2\pi} e^{ik_z z} Z(L, f - ik_z/\beta) \\ &= \frac{Z(L, 0)}{Z(L, f)} \exp(\beta f z) P(z; L, 0) \end{aligned}$$

mais elle ne permet pas le calcul de $S(r; L, f)$. Autrement dit, elle ne contient pas suffisamment d'information sur le système. À force nulle chacune des directions x , y ou z sont équivalentes et P est en fait valable pour l'une quelconque de ces directions (voir l'exemple pédagogique de la chaîne gaussienne § 2.3). À force non nulle en revanche, la direction d'étirement z se distingue de celles qui lui sont perpendiculaires (x ou y), sur lesquelles nous n'avons aucune information. Plus précisément nous ne sommes en fait plus intéressés par la distribution isotrope en extension $r = \sqrt{x^2 + y^2 + z^2}$ mais par celle spécifiant à la fois, en coordonnées cylindriques, les extensions z et $\rho = \sqrt{x^2 + y^2}$. Dans l'absolu nous aimerions pouvoir

1. calculer la distribution conjointes en extensions z et x (choisie arbitrairement, comme z à force nulle) $\mathcal{P}(x, z; L, f)$. C'est encore une distribution marginale, car nous n'avons aucune information sur l'extension y ;
2. trouver une relation analytique simple, en tirant profit de la symétrie cylindrique, entre cette distribution \mathcal{P} et celle spécifiant les extensions z et ρ , $\mathcal{S}(\rho, z; L, f)$.

Le calcul de $\mathcal{P}(x, z; L, f)$ est identique à celui de $P(z; L, f)$ en imposant la contrainte supplémentaire $x = \int_0^L \cos \varphi(s) \sin \theta(s) ds$, au moyen d'une deuxième distribution de Dirac. Leurs représentations spectrales exprime \mathcal{P} comme la double TF – modulo la normalisation par $Z(L, f)$ – de l'amplitude du vide²

$$\mathcal{Z}(L, f - ik_z/\beta, k_x) = \langle 0, 0 | \exp \left\{ -\frac{L}{A} \left[\widehat{H} (f - ik_z/\beta) + ik_x A \widehat{\cos \varphi \sin \theta} \right] \right\} | 0, 0 \rangle$$

avec le hamiltonien \widehat{H} du rotateur rigide sous champ (§ 2.1). Dans la base des harmoniques sphériques $|\ell, m\rangle$, le nouvel opérateur a les composantes (annexe B)

$$\begin{aligned} \langle \ell, m | \widehat{\cos \varphi \sin \theta} | \ell', m' \rangle &= -\frac{1}{2} \frac{1}{\sqrt{(2\ell+1)(2\ell'+1)}} \times \\ &\times \left\{ \left[\sqrt{(\ell'+m')(\ell'+m)} \delta_{m, m'-1} - \sqrt{(\ell'-m')(\ell'-m)} \delta_{m, m-1} \right] \delta_{\ell, \ell'-1} + \right. \\ &\left. + \left[\sqrt{(\ell+m)(\ell+m')} \delta_{m', m-1} - \sqrt{(\ell-m)(\ell-m')} \delta_{m, m'-1} \right] \delta_{\ell', \ell-1} \right\}. \end{aligned}$$

¹La fonction de partition $Z(L, f) = \langle 0, 0 | \exp \left\{ -\frac{L}{A} \left[\widehat{L}^2/2 - \beta A f \widehat{\cos \theta} \right] \right\} | 0, 0 \rangle$ assure la normalisation de la distribution. En pratique, elle compense le facteur $\exp(\beta f z)$.

²Qui se réduit bien à $Z(L, f - ik_z/\beta)$ pour $k_x = 0$.

Le calcul de la distribution en extensions x et z ne pose donc pas de problème en théorie. Malheureusement elle n'est pas fiable de manière simple à celle en extensions ρ et z , comme c'était le cas à force nulle entre les distributions P et S (annexe D). Nous pouvons en revanche calculer \mathcal{S} directement à partir de l'amplitude du vide \mathcal{Z} , par un raisonnement similaire à celui de l'annexe D. Nous omettons temporairement la dépendance en L , afin de gagner un peu en clarté. La symétrie cylindrique rend la distribution en distance bout à bout $Q(\vec{r}; f)$ indépendante en ϕ : c'est une fonction de ρ et z uniquement. Sa transformée de Fourier possède la même symétrie : c'est une fonction de k_ρ et k_z uniquement, c'est-à-dire indépendante de k_ϕ . Pour $\vec{k} = k_\rho \vec{e}_x + k_z \vec{e}_z$ elle se réécrit

$$\int d\vec{r} \exp\left\{-i[k_\rho(\vec{r} \cdot \vec{e}_x) + k_z(\vec{r} \cdot \vec{e}_z)]\right\} Q(\vec{r}; f) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dz \exp[-i(k_\rho x + k_z z)] \underbrace{\int \delta[(\vec{r} \cdot \vec{e}_x) - x] \delta[(\vec{r} \cdot \vec{e}_z) - z] Q(\vec{r}; f) d\vec{r}}_{\text{distribution } \mathcal{P}(x, z; f)}$$

Nous reconnaissons la double TF de la distribution $\mathcal{P}(x, z; f)$, à savoir l'amplitude du vide $\mathcal{Z}(f - ik_z/\beta, k_\rho)$ normalisée par $Z(f)$. La distribution $Q(\vec{r}; f)$ est obtenue par double TF inverse. En prenant $\vec{r} = \rho \vec{e}_x + z \vec{e}_z$, elle se réécrit

$$Q(\vec{r}; f) = \frac{1}{Z(f)} \int \frac{d\vec{k}}{(2\pi)^3} \exp\left\{i\left[\rho \underbrace{\vec{k} \cdot \vec{e}_x}_{k_\rho \cos k_\phi} + z \underbrace{\vec{k} \cdot \vec{e}_z}_{k_z}\right]\right\} \mathcal{Z}(f - ik_z/\beta, k_\rho).$$

En intégrant sur k_ϕ nous obtenons finalement

$$\mathcal{S}(\rho, z; L, f) = 2\pi\rho \times \frac{1}{Z(L, f)} \int_{-\infty}^{\infty} \frac{dk_z}{2\pi} \exp(ik_z z) \times \int_0^{\infty} \frac{dk_\rho}{2\pi} k_\rho J_0(\rho k_\rho) \mathcal{Z}(L, f - ik_z/\beta, k_\rho)$$

où $J_0(u) = \int_0^{2\pi} e^{\pm iu \cos \alpha} \frac{d\alpha}{2\pi}$ est la fonction (réelle et paire) de Bessel du premier type, d'ordre zéro. Comme pour la distribution en extension z nous pouvons isoler la contribution en force

$$\mathcal{S}(\rho, z; L, f) = \frac{Z(L, 0)}{Z(L, f)} \exp(\beta f z) \mathcal{S}(\rho, z; L, 0).$$

Nous reconnaissons en k_z une TF « usuelle » (§ F.1) pour laquelle nous utilisons l'algorithme de FFT [PTVF02]. L'intégrale en k_ρ est en fait une TF radiale en dimension 2, également appelée Transformée de Hankel (TH) ou encore transformée de Fourier/Bessel (§ F.2). Son calcul n'est pas du ressort de la FFT, et une intégration numérique directe est délicate : l'amplitude du vide \mathcal{Z} et la fonction de Bessel J_0 sont toutes deux des fonctions rapidement oscillantes de k_ρ . En section 5.3 nous avons intégré numériquement une autre fonction rapidement

oscillante, à savoir un produit de polynômes de Legendre, en divisant l'intervalle d'intégration entre les zéros de cette fonction. Le principe est à retenir, mais ici l'intégration sur k_ρ couvre $[0; \infty[$ en entier. En 1971, Lado proposa un algorithme permettant l'évaluation précise d'une TH [Lad71]. Il n'est implémenté dans aucune des bibliothèques de calcul numérique les plus connues et son usage dans la littérature scientifique semble limité, c'est pourquoi nous en détaillons le principe en annexe F. Pratiquement, le lecteur retiendra l'introduction d'une coupure dans l'intervalle d'intégration et un échantillonnage spécifique (à la manière d'une quadrature gaussienne) à pas non constant.

Il n'est pas évident à première vue que l'algorithme de Lado et celui de FFT n'interfèrent pas. C'est encore l'inextensibilité du polymère qui va nous sauver. Elle assure que $\mathcal{S}(\rho, z; L) = 0$ dès que $\rho^2 + z^2 \geq L^2$. Cela reste en particulier vrai $\forall z$, quand $\rho \geq L$. Notons

$$f(\rho, k_z; L) = \int_0^\infty \frac{dk_\rho}{2\pi} k_\rho J_0(\rho k_\rho) \mathcal{Z}(-ik_z/\beta, k_\rho).$$

Nous vérifions alors que si $\forall z$, on a pour $\rho \geq L$

$$\int_{-\infty}^\infty \frac{dk_z}{2\pi} \exp(ik_z z) f(\rho, k_z; L) = 0,$$

alors par transformée inverse $f(\rho, k_z; L) = 0$ dès que $\rho \geq L$, et ce $\forall k_z$. Autrement dit nous venons de montrer pour $f(\rho, z; L)$ l'existence d'une (vraie) coupure L en ρ , valable pour toute valeur de k_z . La coupure (approchée) en k_ρ correspondante³ qui permet le calcul de la TH par l'algorithme de Lado, n'introduit alors pas d'erreur⁴, et ce quelque soit k_z . Nous pouvons donc encore utiliser L comme fréquence de critique en z pour réaliser la FFT. En effet nous avons également

$$\frac{\mathcal{S}(\rho, z; L)}{2\pi\rho} = \int_{-\infty}^\infty \frac{dk_z}{2\pi} \exp(ik_z z) f(\rho, k_z; L) = 0$$

dans l'autre cas particulier où $z \geq L$, et ce $\forall \rho \in [0; L[$. L'algorithme de Lado et celui de FFT sont donc indépendants.⁵ Signalons pour finir que contrairement aux sections 4.2 et 5.3, nous avons doublé la fréquence critique par sécurité.

En fait, le paramètre limitant n'est pas la coupure introduite par l'algorithme de Lado mais le nombre de points choisis pour l'échantillonnage. L'amplitude du vide \mathcal{Z} qui donne la distribution \mathcal{S} par TH et TF successives doit en effet encore être échantillonnée. Ce n'est plus comme avant (§§ 4.2 ou 5.3) une fonction d'une, mais de deux variables k_z et k_ρ . Plus précisément il s'agit, dans la base

³Qui est inversement proportionnelle à L : $\Lambda = \mu_N/L$ (§ F.2).

⁴L'erreur commise correspond à la relation d'orthogonalité approchée assurant la réciprocity de la TH discrète (§ F.2).

⁵Au niveau discret, les deux transformées permutent en fait. Nous l'avons également vérifié en pratique lors du calcul de \mathcal{S} : $\text{TF}(\text{TH}) = \text{TH}(\text{TF})$.

des harmoniques sphériques $|\ell, m\rangle$, de la première des composantes (c'est-à-dire celle selon $|0, 0\rangle$) de

$$\exp \left\{ -\frac{L}{A} \left[\widehat{H}(-ik_z/\beta) + ik_x A \widehat{\cos \varphi \sin \theta} \right] \right\} |0, 0\rangle$$

que nous calculons au moyen d'Expokit [Sid98]. Contrairement aux sections 4.2 et 5.3 où seul le hamiltonien du rotateur intervenait, la matrice dont nous voulons connaître l'action sur $|0, 0\rangle$ inclut l'opérateur additionnel $\widehat{\cos \varphi \sin \theta}$, pour lequel $|\ell, m\rangle$ n'est pas stable en m (voir ses composantes en début de section). Comme à chaque valeur de ℓ correspond $(2\ell + 1)$ valeurs de m , imposer une coupure ℓ_{\max} nous laisse une matrice carrée de taille $(\ell_{\max} + 1)(2\ell_{\max} + 1)$, contre $(\ell_{\max} + 1)$ avant.⁶ La taille ℓ_{vieux} est maintenant obtenue pour

$$\ell_{\max} (2\ell_{\max} + 3) = \ell_{\text{vieux}} \quad \text{soit} \quad \ell_{\max} \sim \sqrt{\ell_{\text{vieux}}/2}$$

pour de grandes matrices. Ainsi $\ell_{\max} \simeq 7$ suppose déjà une matrice de taille 100. Notons toutefois que cette dernière reste creuse et possède une symétrie bien particulière. Expokit a notamment été conçu pour prendre en compte de grandes matrices, à condition que ces dernières soient essentiellement creuses. Au regard des sections 4.2 et 5.3 nous avons retenu une coupure $\ell_{\max} = 50$, c'est-à-dire une matrice 5151×5151 . Ajouté au double échantillonnage, un temps de calcul raisonnable impose alors une sévère limitation sur le nombre de points, qui se répercute directement sur l'échantillonnage en ρ de \mathcal{S} . Or c'est l'intégrale de cette distribution qui à terme nous intéresse (§ 6.2) : nous ne pouvons donc pas plus nous permettre de sacrifier la définition en ρ de \mathcal{S} , que celle en z . En équilibrant⁷ les échantillonnages de l'algorithme de Lado et de la FFT nous avons constaté que l'obtention d'une définition suffisante de \mathcal{S} (512 points selon k_z , 1000 points selon k_ρ) était encore trop coûteuse, en temps de calcul, pour une étude systématique en fonction de la longueur L . Nous nous sommes donc limités à une seule valeur, à savoir les 113 bp impliqués dans la répression de l'opéron Gal (§ 1.1).⁸

6.2 Étirement de l'ADN par la force

Nous commençons par l'effet de la force sur la distribution marginale en extension z . Après ce qui vient d'être dit (§ 6.1), nous n'attendons pas de $P(z; L, f)$ une information exhaustive sur les configurations du polymère. La figure 6.1 (gauche) traduit l'étirement du polymère, dans la direction d'application de la force.

⁶En effet $2 \sum_{\ell}^{\ell_{\max}} (2\ell + 1) = 2(\ell_{\max} + 1)(2\ell_{\max} + 1)$.

⁷Pour l'algorithme de Lado le choix de ce nombre est libre, contrairement à la FFT pour laquelle nous sommes limités aux puissances de deux (§ F.1).

⁸Soit $L/A \simeq 0.75$ d'où la coupure $\ell_{\max} = 50$ (voir note de bas de page 54).

En augmentant la force nous observons graduellement un transfert de probabilité vers l'évènement $z \lesssim L$, autour duquel $P(z; L, f)$ se pique de plus en plus.⁹ Le calcul de cette distribution ne constitue pas qu'une bonne illustration de la

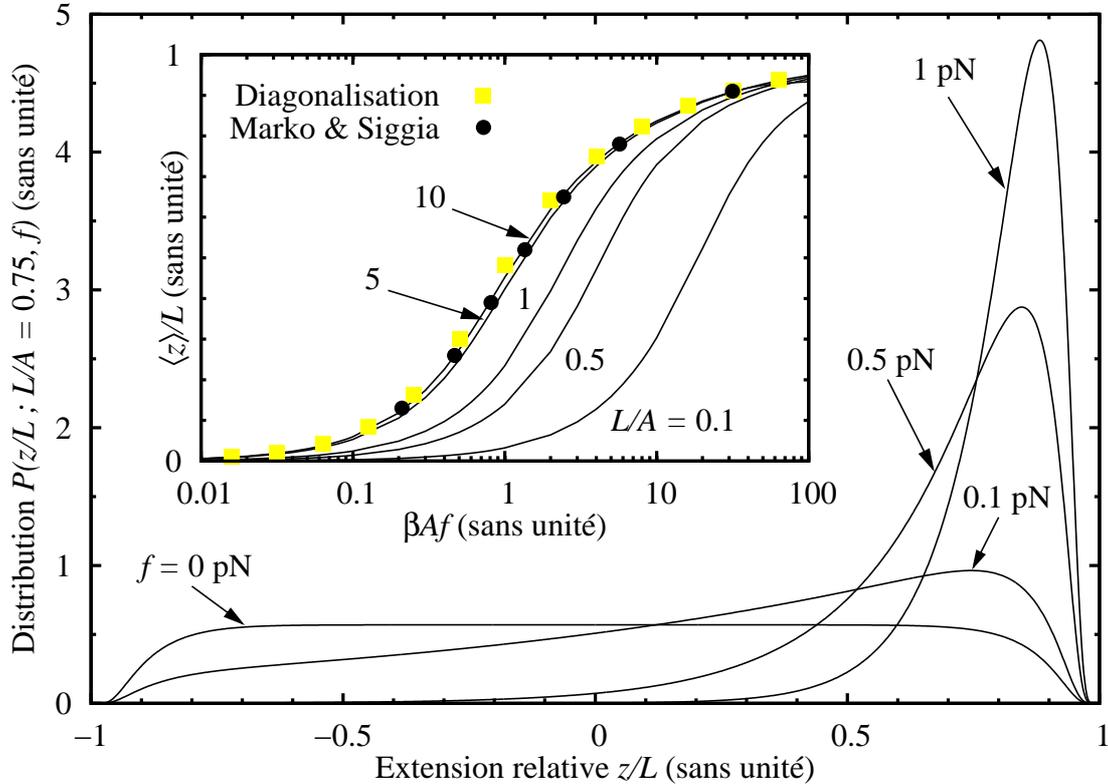


FIG. 6.1 – Apparition d'une direction privilégiée le long de de la force d'étirement ($z > 0$) illustrée sur la distribution marginale en z d'un polymère de longueur 113 bp ($L/A \simeq 0.75$). Médaillon : l'intégration directe de la distribution (lignes pleines) reproduit aux grandes longueurs $L \gg A$ la relation de force/extension obtenue par diagonalisation (■) en section 4.1. Nous reproduisons également la formule proposée par Marko et Siggia (●).

brisure de symétrie sphérique, par l'introduction de la force : $P(z; L, f)$ nous permet également de revisiter les résultats de la section 4.1. La figure 6.1 montre comment le polymère commence à s'étirer dès 0.1 pN pour se retrouver quasi-totalement étiré aux forces dépassant 2 pN. Bien que qualitatives, ces observations se confirment quantitativement en figure 4.1 : l'extension relative moyenne, dans la direction de la force, passe d'environ 50% à plus de 90% pour les forces réduites $\beta Af \simeq 1$ et 25 respectivement.

Nous tirons en fait des conclusions hâtives : en section 4.1 la relation de force/extension est calculée pour de long polymères $L \gg A$, tandis que la figure 6.1 montre la distribution d'une séquence de 113 bp, soit moins que la longueur de persistance. Comme la dépendance en force de $P(z; L, f)$ se sépare

⁹Cela reflète la non parité en z , du facteur exponentiel $\exp(\beta fz)$.

de $P(z; L, f = 0)$, nous pouvons en fait calculer la relation de force/extension pour une large gamme de longueurs, grâce aux résultats de la section 4.3. Nous évaluons l'extension z moyenne, par intégration numérique directe

$$\langle z \rangle = Z(L, 0) 2 \int_0^\infty \frac{\sinh(\beta f z)}{Z(L, f)} z P(z; L, 0) dz$$

où l'on a utilisé la parité de la distribution à force nulle. En figure 6.1 (médaillon) nous voyons que quelques longueurs de persistance suffisent pour retrouver la relation de force/extension de la section 4.1 (symboles). Plus précisément, si l'accord n'est encore que qualitatif à cinq longueurs de persistance, il devient quantitatif pour dix. Une différence significative n'est observable qu'aux faibles longueurs $L \ll A$ (p. ex. $L/A = 0.1$ sur la figure). Aux longueurs intermédiaires $L \simeq A$ la différence est moins prononcée.¹⁰ Une extension relative moyenne de 50% est ainsi provoquée par une force réduite $\beta A f \simeq 3$ (environ 0.25 pN) pour une longueur de persistance, contre $\beta A f \simeq 1$ (autour de 0.08 pN) aux grandes longueurs $L/A = 5$ ou 10.¹¹ Cette étude montre la bonne validité de la diagonalisation utilisée en section 4.1, pour les molécules d'ADN de plus d' 1 kb. Elle ne prétend néanmoins pas s'y substituer : la détermination du spectre entier du hamiltonien, du rotateur rigide sous champ (réel), est préférable au calcul de son exponentielle (avec un champ complexe), suivit d'une FFT puis d'une intégration.

Mentionnons pour finir une autre lecture possible de la relation de force/extension. La longueur de persistance de l'ADN dépend notamment de la concentration en ions Na^+ . Les 50 nm ou encore les 150 bp utilisés jusqu'ici correspondent en fait à la longueur de persistance observée à saturation en Na^+ , soit environ 100 mM. Les mesures réalisées par Baumann, Smith, Bloomfield et Bustamante [BSBB97] en 1997 s'ajustent à une longueur de persistance¹² inversement proportionnelle à la concentration en ions et saturant à la valeur communément admise

$$A - A_\infty = 64.8 / [\text{Na}^+] \quad \text{avec} \quad A_\infty = 50 \text{ nm, et } 1.86 \text{ mM} \lesssim [\text{Na}^+] \lesssim 586 \text{ mM.}$$

Ainsi, une longueur réduite donnée $L/A = \ell$ peut s'interpréter : à longueurs de contour L variable et de persistance $A = A_\infty$ fixe, comme en figure 6.1 (médaillon), ou réciproquement à longueurs de persistance A variable et de contour $L = L_0$ fixe. Dans ce dernier cas nous pouvons considérer que c'est la concentration en ions qui varie $[\text{Na}^+] \simeq 1.3 \times \frac{\ell}{L_0/A_\infty - \ell}$ mM. Nous ne retraçons pas la figure 6.1 (droite) : l'échelle de concentration ionique explorée correspond à une longueur réduite ℓ variant *grosso modo* entre 0.6 et une fois la longueur réduite à saturation L_0/A_∞ . Si on fixe $L_0 = 10A_\infty = 500$ nm, nous devons alors couvrir

¹⁰C'est le cas de la séquence de 113 bp pour laquelle $L/A \simeq 0.75$.

¹¹Attention aux unités réduites : $\langle z \rangle / L = 50\%$ correspond à $\langle z \rangle = A/2$ et $5A$, pour un polymère rigide ($L = A$) et flexible ($L = 10A$) respectivement.

¹²Dans l'article, elle est ponctuellement confondue avec l'élément de Kuhn.

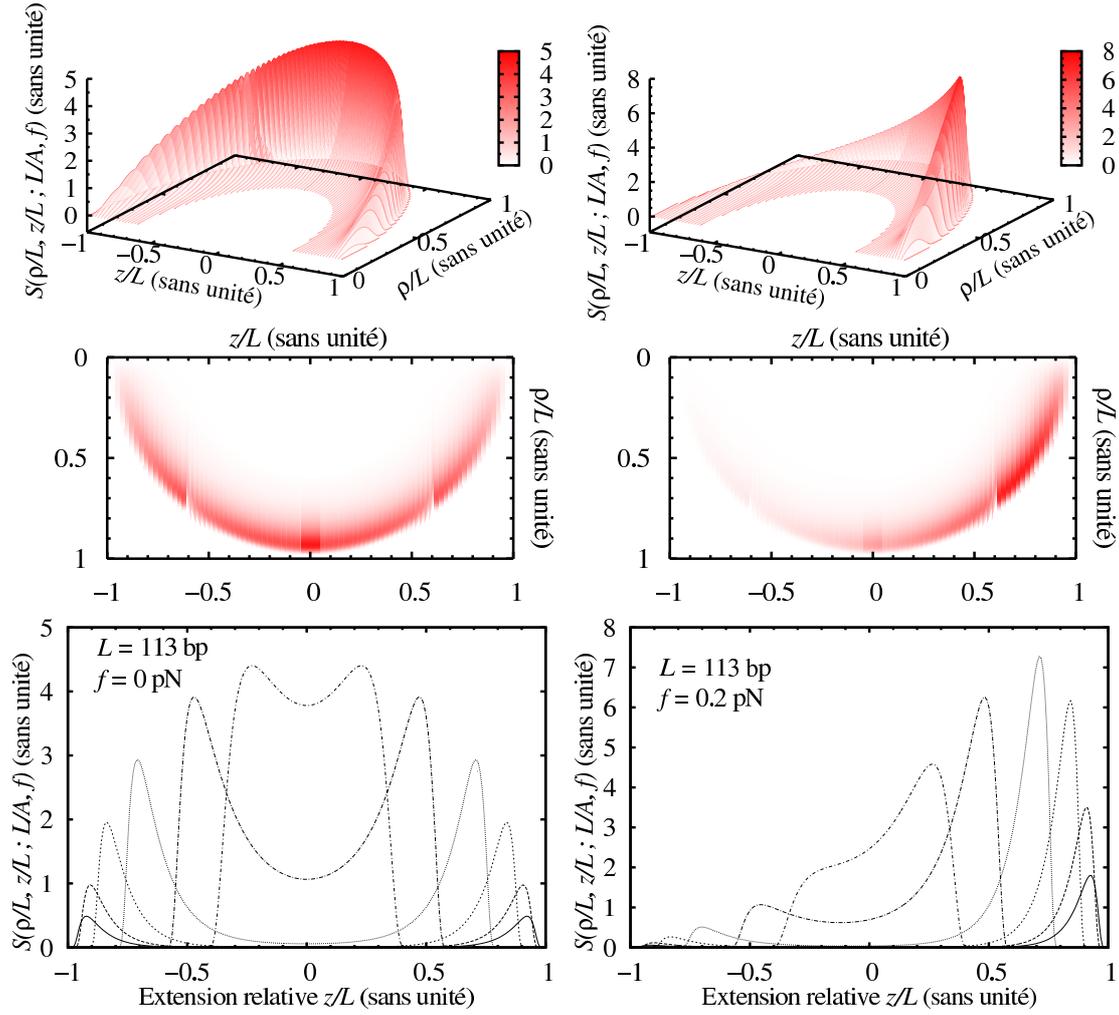


FIG. 6.2 – Calcul numérique de la distribution $\mathcal{S}(\rho, z; 113 \text{ bp}, f)$, en l'absence de force (gauche) et pour $f = 0.2 \text{ pN}$ (médaille). La direction d'application de la force ($z > 0$) est privilégiée, et la distribution perd sa parité en z . Sont montrés : l'ensemble des distributions (haut), leur projection sur le plan ($O\rho z$) (milieu) ainsi que six coupes à ρ fixé (bas), pour les valeurs $\rho/L = 0.1, 0.2, 0.4, 0.6, 0.8, 0.9$ (cette dernière est en ligne pleine).

l'échelle $6 \lesssim \ell \lesssim 10$ pour laquelle la figure 6.1 (droite) montre peu de différences dans la relation de force/extension¹³. Pour commencer à observer un étalement il faut descendre aux faibles longueurs $L_0 \leq A_\infty$, situation plutôt inhabituelle expérimentalement à moins d'étudier un mécanisme particulier, comme la cyclisation des 113 bp impliquées dans la répression de l'opéron Gal. Il est intéressant de noter qu'en pratique c'est l'inverse qui est réalisé : la relation de force/extension est mesurée pour diverses conditions de salinité, puis ajustée dans chaque cas à la formule de Marko et Siggia (§ 4.1, pouvant inclure une extensibilité de l'ADN) afin d'obtenir la dépendance en la salinité de la longueur de persistance.

¹³La force réduite correspond à $\beta A_\infty f$. Il est peut-être plus clair d'utiliser directement f .

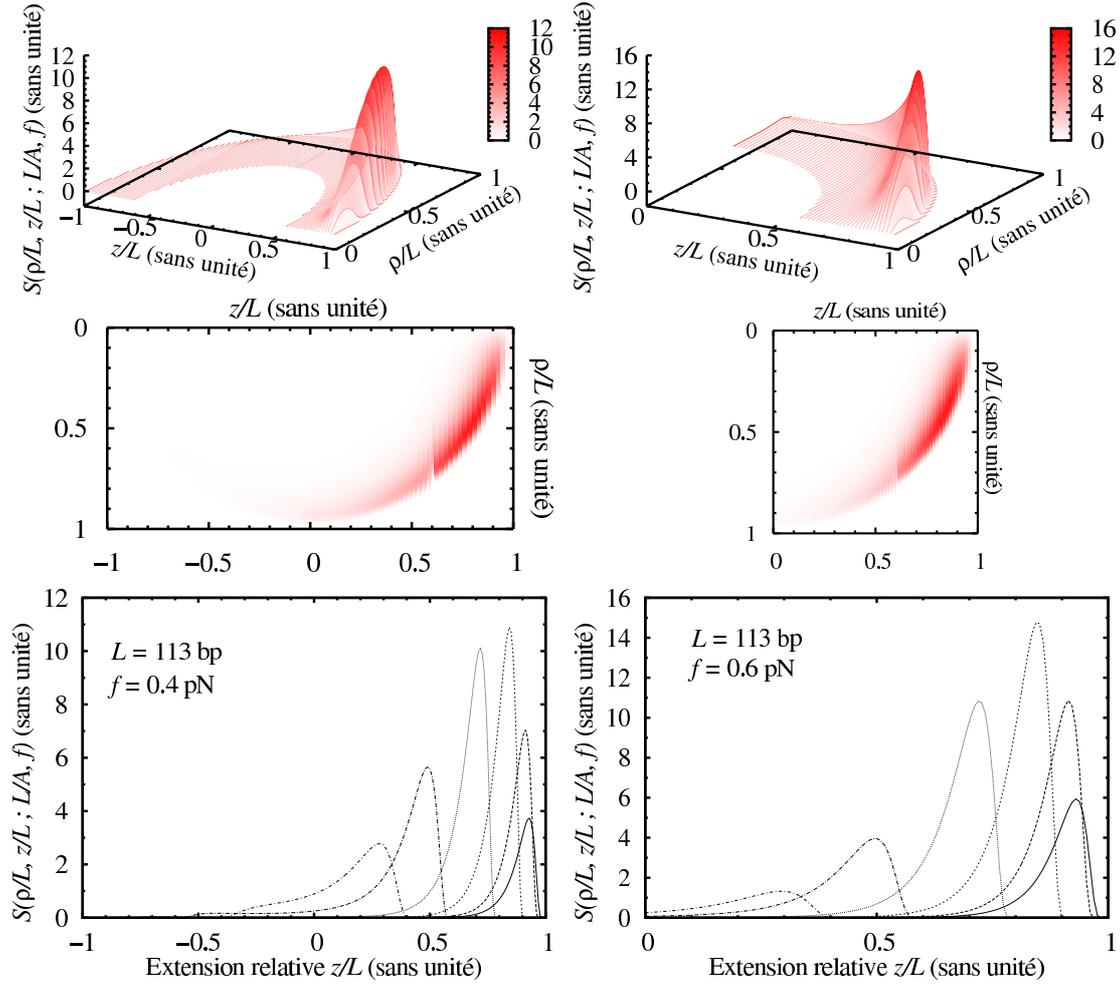


FIG. 6.3 – Idem figure 6.2, en considérant des forces $f = 0.4$ pN et 0.6 pN. Pour cette dernière, l'échelle est limitée à $0 \leq z/L \leq 1$.

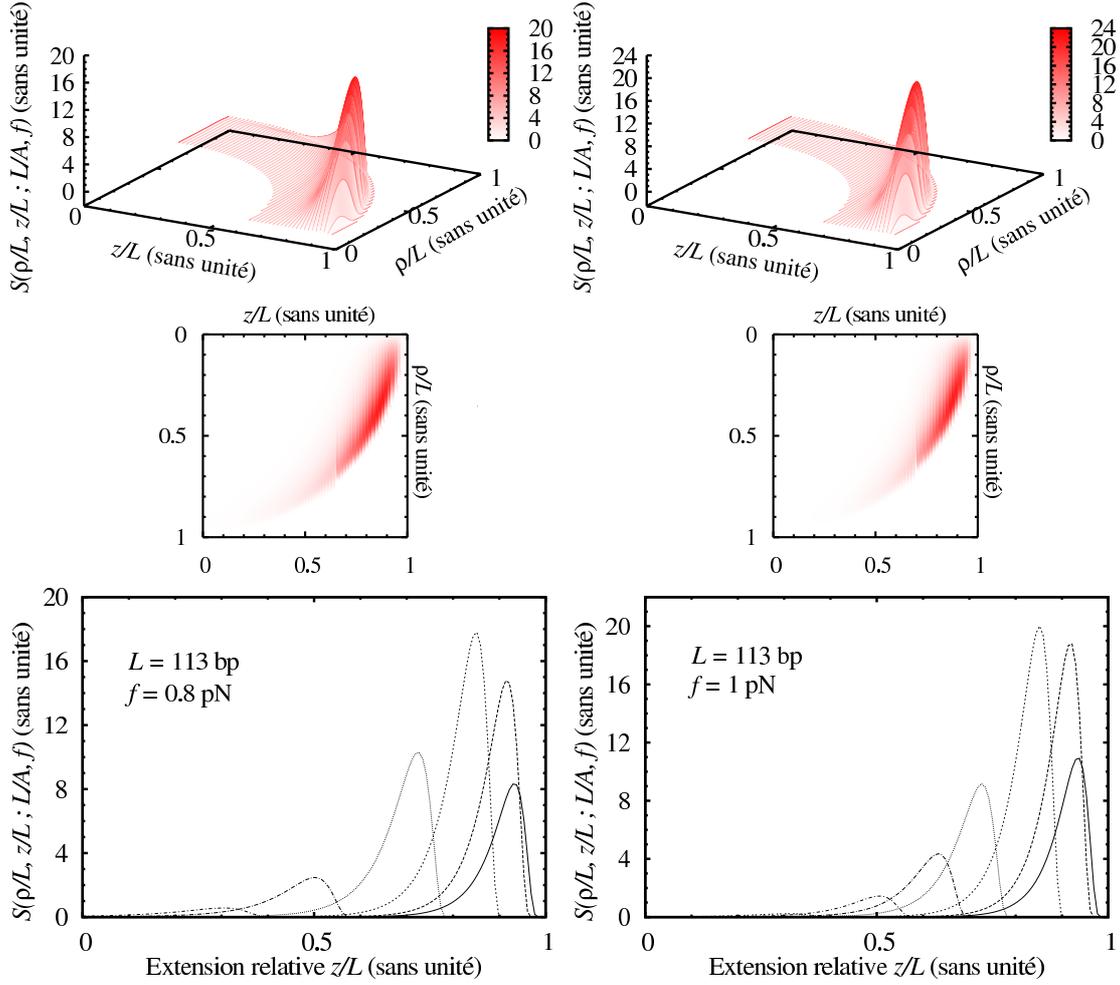
Les figures 6.2, 6.3 et 6.4 montrent également l'étirement du polymère dans la direction d'application de la force. Plus précisément, l'augmentation graduelle de la force se traduit par un transfert de probabilité vers l'évènement $z \lesssim L$ et $\rho \ll L$, autour duquel la distribution $\mathcal{S}(\rho, z; L, f)$ se pique de plus en plus. La distribution en extension $r = \sqrt{\rho^2 + z^2}$ se calcule alors par intégration

$$\begin{aligned} S(r; L, f) &= \int_{-\infty}^{\infty} dz \int_0^{\infty} d\rho \delta\left(r - \sqrt{\rho^2 + z^2}\right) \mathcal{S}(\rho, z; L, f) \\ &= r \int_{|z| < r} \frac{\mathcal{S}(\sqrt{r^2 - z^2}, z; L, f)}{\sqrt{r^2 - z^2}} dz. \end{aligned}$$

Numériquement, nous pouvons utiliser une quadrature de Gauss/Chebyshev¹⁴, ou bien encore prendre la limite de l'intégration usuelle

$$S(r; L, f) = \lim_{\epsilon \rightarrow 0^+} \operatorname{Re} \left[r \int_{|z| < r} \frac{\mathcal{S}(\sqrt{r^2 - z^2}, z; L, f)}{\sqrt{r^2 - z^2 + i\epsilon}} dz \right].$$

¹⁴Intégration sur $-1 < z/L < 1$ avec le poids $1/\sqrt{1 - (z/L)^2}$ [PTVF02].

FIG. 6.4 – Idem figure 6.2, pour des forces $f = 0.8$ pN et 1 pN.

Nous retrouvons à force nulle le comportement de la section 4.2. La figure 6.5 (gauche) montre une extension r la plus probable, tendant aux grandes forces vers l'étirement quasi-total $r \lesssim L$. Rappelons que nous sommes toujours à longueur fixée $L = 113$ bp et que c'est la force qui varie. Il est raisonnable de penser qu'aux plus grandes forces les distributions $P(z; L, f)$ et $S(r; L, f)$ soient très semblables, le polymère étant alors complètement étiré selon (Oz) . La figure 6.5 (droite) montre ces deux distributions sous 10 pN. Elles apparaissent en fait relativement différentes, malgré l'importante force appliquée (revoir la figure 4.1). Le médaillon confirme que les extensions les plus probables (lignes pleines) de ces deux distributions se rejoignent aux grandes forces, bien que leur comportement global demeure significativement différent. Pour l'illustrer, le médaillon montre aussi les maxima correspondants des deux distributions : la convergence aux grandes forces n'est pas triviale. Il serait plus intéressant de comparer leur intégrale

$$\int_{-r_B}^{r_B} P(z; L, f) dz \stackrel{f \gg 1}{\simeq} \int_0^{r_B} P(z; L, f) dz \quad \text{et} \quad \int_0^{r_B} S(r; L, f) dr = \text{Prob}(r_B; L, f).$$

Supposons qu'il existe une force f_0 au delà de laquelle ces deux quantités s'accordent quantitativement. Alors pour toute force $f \geq f_0$ nous pourrions évaluer le facteur de cyclisation, en fait le rapport $K(r_B; L, f)$, en utilisant la distribution $P(z; L, f)$ qui est simple à calculer comparée à $S(r; L, f)$. Malheureusement, il n'apparaît pas fiable de réaliser une nouvelle intégration numérique sur cette dernière : elle provient déjà d'un calcul d'exponentielle de matrice, ayant subi deux transformations intégrales pour être finalement de nouveau intégrée. Une telle comparaison ne serait de plus valable que pour les 113 bp considérés ici. Pour finir, revenons sur le travail de Lia et coll. [LBC⁺03]. La différence d'extension provoquée par la formation de la boucle de 113 bp, se trouve être plus importante que prévue. Les 55 nm (165 bp) mesurés dépasse la longueur de

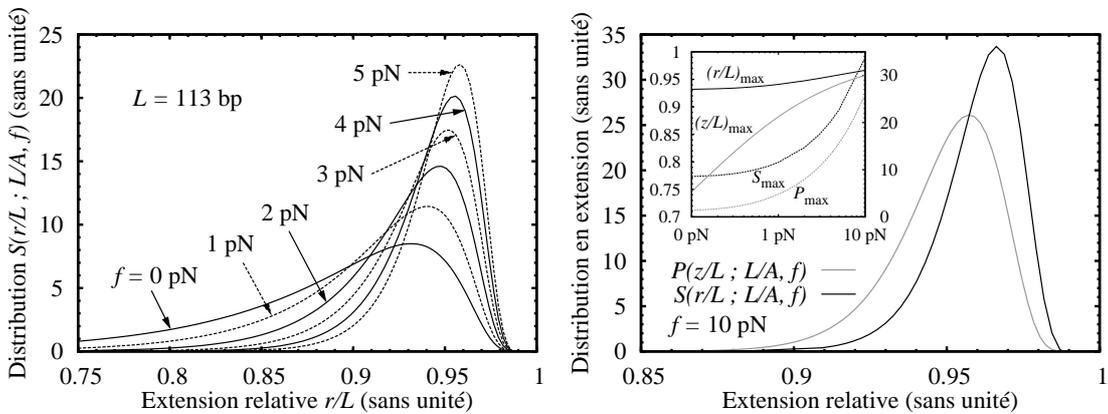


FIG. 6.5 – Effet de l'étirement illustré sur la distribution en extension r (gauche) d'une séquence de 113 bp (gauche). Pour 10 pN, elle diffère significativement de la distribution en extension z , c'est-à-dire le long de l'étirement (droite). Les extensions r et z les plus probables tendent toutes deux vers l'étirement maximum (droite, médaillon).

la séquence d'environ 2×26 bp. Pour les auteurs, cet excès est attribuable à une boucle en forme d'épingle à cheveux (figure 6.6). Nous avons jusqu'ici considéré la boucle indépendamment du reste de la molécule d'ADN où elle se forme. Cette approche est pertinente en l'absence de force : la seule contrainte sur la boucle – hormis de possibles cassures – est imposée par ses orientations limites, elles-mêmes fixées par le complexe ADN/protéine. En appliquant une force, les parties de la molécule éloignées de la boucle sont alors plus facilement étirées que celles de son voisinage direct. Selon la géométrie de la boucle, et donc selon le complexe ADN/protéine fixant ses orientations limites, la rigidité de l'ADN peut exiger quelques dizaines de paires de bases avant de récupérer la direction de l'étirement (figure 6.6). De récents travaux théoriques (2005) ont tenté de prendre en compte cet effet. Dans leur adaptation empirique du résultat de Shimada et Yamakawa, Sankararaman et Marko [SM05b] considèrent qu'une boucle de longueur L accapare en fait une longueur $(L + 2\epsilon)$. Si nous

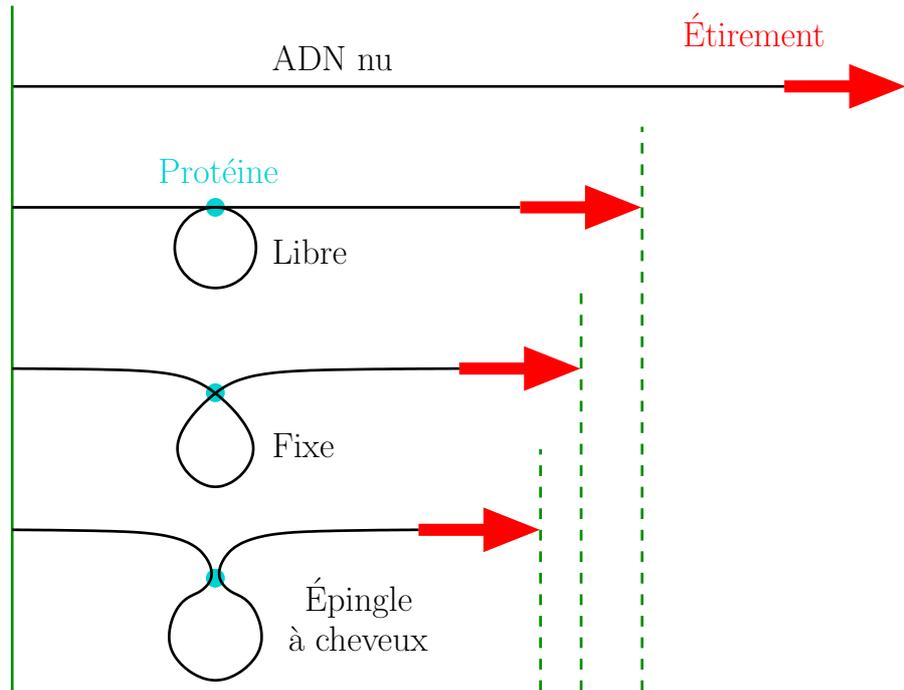


FIG. 6.6 – Effet de la rigidité élastique du complexe protéine/ADN sur l'étirement du brin d'ADN (illustration). Selon les orientations limites de la boucle, fixée par un complexe rigide, la persistance de l'ADN requiert une longueur plus ou moins grande pour que la molécule retrouve la direction de l'étirement (deux dernières situations). Dans le cas opposé d'un complexe flexible, les orientations limites sont relaxées par l'étirement : la longueur de la boucle seule est responsable de la différence d'extension observée (deux premières situations).

nous focalisons enfin sur les parties extérieures à la boucle, la molécule d'ADN peut être vue comme cassée en le complexe ADN/protéine.¹⁵ Blumberg, Tkachenko et Meiners [BTM05] ont estimé l'énergie libre associée à cette molécule¹⁶ et comparé les temps de vie correspondant aux données de Lia et coll. Kulić et coll. [KML⁺05] ont repris le calcul de Shimada et Yamakawa en stabilisant la boucle par un potentiel effectif. La relation de force/extension aux grandes forces est identique à celle sans cassure (§ 4.1) en introduisant une longueur de persistance effective [BBS⁺00]. Des données plus systématiques sur la géométrie de la boucle (complexe ADN/protéine) permettraient de lever la part d'arbitraire encore présente dans ces études, tout comme aux chapitres 4 et 5.

¹⁵La molécule est toutefois dans sa forme linéaire, au contraire du chapitre 5.

¹⁶Ils considèrent trois contributions : celle de la boucle (résultat de S&Y, négligeant l'affinité ADN/protéine), celle associée à l'étirement seul (intégration de la formule de Marko et Siggia) puis celle précisant l'étirement de la cassure (la fonction de corrélation en présence d'une cassure, dérivée par Rivetti et coll. [RWB98] est employée).

Chapitre 7

Vers des modèles de cyclisation de plus en plus complets ?

Un calcul analytique exact du facteur de cyclisation, ou plus généralement des diverses distributions en extensions du WLC n'est à ce jour pas encore possible. Les méthodes numériques sont un outil précieux car elles permettent l'obtention quasi-systématique de résultats quantitatifs. Elles peuvent en revanche s'avérer lourdes à mettre en place selon les phénomènes que nous voulons décrire : fermeture de la boucle, accrochage de protéines, étirement par une force, etc. Parmi les approximations analytiques possibles celle de point col semble prometteuse : son principe simple la rend en théorie systématiquement adaptable, par exemple aux phénomènes que nous venons de citer. L'expression semi-analytique du facteur de cyclisation obtenue par Shimada et Yamakawa montre un accord quantitatif avec les résultats numériques correspondant, pour les longueurs auxquelles effets élastique et entropique sont tous deux importants. Elle complète alors de façon remarquable le modèle du polymère gaussien. Obtenir une expression analytique grâce à cette méthode, autrement qu'en adaptant empiriquement le résultat de S&Y, demande objectivement une remise à plat du travail original [Yam97, SY84] de ces auteurs.¹ Cette voie a récemment été engagée par Kulić [Kul04, KML⁺05], soulignant sans ambiguïté les difficultés calculatoires (non triviales) ainsi que l'instabilité de la boucle dans le cas tridimensionnel. Le problème est alors de choisir une interaction stabilisant le système, qui soit pertinente expérimentalement.

7.1 Fermeture de la boucle

Tout au long de ce manuscrit nous avons souligné l'importance des orientations limites, en la fermeture de la boucle. Elles sont biologiquement im-

¹C'est l'une des références théoriques les plus citées dans la littérature scientifique, sans que personne n'arrive encore à en reproduire les résultats analytiques.

posées par le complexe ADN/protéine qui fixe la boucle. En l'absence de données expérimentales systématiques, ne serait-ce que pour l'opéron Lac, nous avons choisis de ne pas spécifier d'orientations particulières². La structure cristallographique de la protéine n'est pas exploitable telle quelle, car elle ne prend pas en compte les contraintes exercées par la boucle d'ADN, qui ne « survit » en effet pas à la cristallisation. L'algorithme numérique proposé par Villa, Balaëff et Schulten [VBMS04, VBS05] pourrait toutefois permettre de prédire des orientations vraisemblables à partir de ces structures.

Cela étant, nous pouvons déjà étudier les deux cas particuliers [SVA05] :

parallèle dont les orientations limites périodiques produisent des boucles cycliques. Il correspond par exemple aux expériences de cyclisation, dans le sens d'hybridation des extrémités par une enzyme [SLB81, SB83, DSS⁺05] ;

anti-parallèle dont les orientations limites opposées rappellent les boucles en forme d'épingle à cheveux, typiques des molécules d'ARN. Il pourrait correspondre à la boucle de 113 bp de l'opéron Gal [LBC⁺03, GVZA01].

Ces deux situations sont schématisées en figure 7.1. Nous avons déjà eu l'occasion, avec la cassure à mi-longueur, d'expliquer comment modifier notre algorithme numérique afin d'imposer des orientations particulières (§ 5.3). Quand au point col, il suffit de reprendre les résultats de $\omega(0) = \alpha/2$ quelconque, de la section 2.2 (§ 5.2 avec une cassure à mi-longueur). Pour de faibles longueurs $L \ll A$,

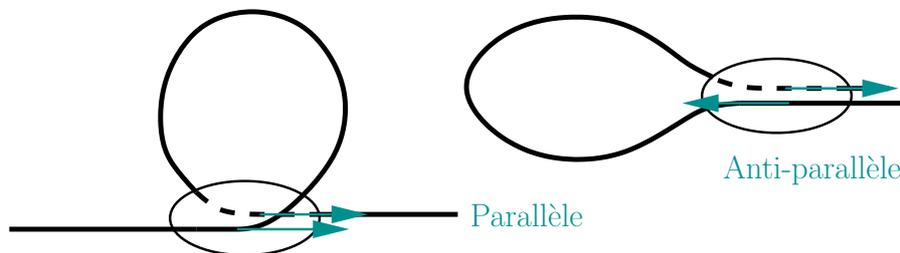


FIG. 7.1 – Orientations parallèles et anti-parallèles en la fermeture de la boucle.

la boucle en épingle à cheveux est plus difficile à former que le cercle. Le point col est par définition la boucle la moins difficile à former et nous attendons donc pour le facteur de cyclisation, la majoration : « anti-parallèle < parallèle < point col ». Aux grandes longueurs $L \gg A$, nous n'attendons au contraire aucun effet significatif de ces orientations.

Nous avons également mis en avant l'importance de ces orientations si la rigidité de torsion de l'ADN est considérée. Plus précisément, la fermeture peut alors exiger un certain alignement des extrémités.

²Pour le point col nous avons considéré les orientations réalisant l'énergie minimale, soit la situation la plus stable en l'absence de contrainte en la fermeture.

7.2 Rigidité de torsion

La rigidité de torsion de l'ADN s'introduit de manière intuitive en utilisant l'analogie entre le vecteur tangent décrivant le WLC, et un bâtonnet rigide (inextensible) subissant des rotations dans le temps, c'est-à-dire le rotateur rigide. Nous avons surtout insisté sur l'analogie quantique (§ A.2), n'utilisant explicitement le système classique que pour le point col (§ 2.2) bidimensionnel. Le cas tridimensionnel se trouve toutefois dans tout livre de physique des solides [Gol80]. Pour obtenir l'énergie cinétique de rotation du bâtonnet, nous devons exprimer son vecteur rotation³

$$\vec{\Omega}(s) = \dot{\varphi}(s) \vec{\Delta}_\varphi + \dot{\theta}(s) \vec{\Delta}_\theta + \dot{\psi}(s) \vec{\Delta}_\psi = \Omega_1(s) \vec{e}_1 + \Omega_2(s) \vec{e}_2 + \Omega_3(s) \vec{e}_3$$

dans ses axes principaux d'inertie \vec{e}_1 , \vec{e}_2 et $\vec{e}_3 = \vec{\Delta}_\psi = \vec{t}$; qui par définition diagonalisent son tenseur d'inertie. Les angles d'Euler (φ, θ, ψ) sont les trois degrés de liberté du rotateur. Le changement de bases correspond à la composition des trois rotations associées à ces angles. Nous obtenons

$$\Omega_1^2(s) + \Omega_2^2(s) = \dot{\varphi}^2(s) \sin^2 \theta(s) + \dot{\theta}^2(s) \quad \text{et} \quad \Omega_3(s) = \dot{\psi}(s) + \dot{\varphi}(s) \cos \theta(s)$$

grâce auxquelles l'énergie s'écrit

$$T_{\text{rot}} = \frac{1}{2} I [\Omega_1^2(s) + \Omega_2^2(s)] + \frac{1}{2} I' \Omega_3^2(s).$$

En effet, les rotations autour de tout axe appartenant au plan (\vec{e}_1, \vec{e}_2) sont équivalentes, ce qui explique l'unique moment d'inertie $I_1 = I_2 \equiv I$. Comme le bâtonnet n'a pas d'épaisseur il ne possède pas de moment d'inertie selon \vec{e}_3 , soit $I' \equiv I_3 = 0$. Si $I = A/\beta$, nous retrouvons alors l'énergie élastique de courbure du WLC car⁴ $\Omega_1^2(s) + \Omega_2^2(s) = [d\vec{t}(s)/ds]^2$. L'analogie quantique peut s'intuitiver en se rappelant la définition du moment cinétique. Dans les axes principaux $\vec{L}(s) = I [\Omega_1(s)\vec{e}_1 + \Omega_2(s)\vec{e}_2]$, et alors⁵

$$\beta T_{\text{rot}} = \beta \vec{L}^2(s)/(2I) = \beta^2 \vec{L}^2(s)/(2A).$$

Nous pouvons prendre en compte la torsion intrinsèque du WLC, en fait de l'ADN, en associant aux rotations autour de \vec{e}_3 un moment d'inertie I' non nul : bien que le bâtonnet n'ait toujours pas d'épaisseur, ces rotations (surenroulements) ont un coût énergétique non nul. Par analogie avec la longueur de persistance A , nous déduisons une longueur caractéristique de la rigidité de torsion

³Nous abandonnons la notation $\Omega = (\theta, \varphi)$ pour l'angle solide.

⁴Au cours d'une rotation $\frac{d}{ds}(\dots) = \vec{\Omega} \wedge (\dots)$.

⁵Les dimensions sont cohérentes. Si $I = A/\beta$, alors T_{rot} est une énergie par unité de longueur tandis que \vec{L} est une énergie (le temps s , est en fait une longueur). Rappelons que pour l'analogie quantique (annexe A) l'opérateur moment cinétique est sans dimension car $\hbar = 1$ (§ B.2).

par $I' = C/\beta$. Une valeur pour C peut être obtenue par ajustement de résultats numériques aux courbes de torsion/extension établies expérimentalement, à force d'étirement fixée [BM00]. La procédure est en revanche beaucoup moins directe que pour A : lorsqu'on applique une torsion additionnelle à une molécule d'ADN dont les extrémités sont contraintes⁶, elle présente des structures, des topologies complexes (revoir la figure 2 p. ex.) dont la modélisation n'a rien de trivial. Il est actuellement admis un rapport $1.2 \lesssim C/A \lesssim 2.2$, soit $180 \text{ bp} \lesssim C \lesssim 330 \text{ bp}$ [BM00, MN97, MN98]. La torsion intrinsèque produisant une période d'environ 10.5 bp, cette longueur de persistance (de torsion) comprend donc entre 17 et 32 tours de la double hélice.

L'analogie quantique peut être étendue en considérant dans l'énergie du WLC la composante $L_3(s)$, maintenant non nulle, du moment cinétique. De même que contraindre l'extension du WLC selon \vec{e}_z introduisait une force (complexe) d'alignement du rotateur suivant un champ électrique appliqué dans cette direction, imposer un surenroulement se traduit par l'apparition d'un monopôle magnétique. Comme expliqué par Bouchiat et Mézard [BM98, BM00, RM02, BM02], cette dernière contrainte rend de plus le hamiltonien pathologique, nécessitant l'introduction d'une coupure en $\theta(s)$ dans l'intégrale des chemins. Le hamiltonien obtenu par ces auteurs n'est diagonalisable analytiquement qu'en l'absence du champ électrique. Autrement, notre algorithme numérique pourrait être modifié, à condition d'inspecter prudemment les configurations que nous voulons décrire. La formation de plectonèmes peut par exemple nécessiter de prendre également en compte les effets d'interaction électrostatique et de volume exclu.

7.3 Écrantage et volume exclu

Les groupements phosphates de chacune des deux chaînes constituant la double hélice de la molécule d'ADN sont chargés négativement. Cette charge n'est pas négligeable : elle correspond à celle de deux électrons, soit $(-2e)$. Un tour de double hélice contient par exemple $(-21e)$.

En fait les propriétés électrostatiques de la molécule dépendent du solvant qui l'entoure, et plus particulièrement de sa concentration ionique en $[\text{Na}^+]$. La charge des groupements phosphate produit une condensation de Manning [Man69], des cations sur la double hélice. La charge effective de chaque groupement est alors $\chi \times (-2e)$ avec $\chi < 1$. L'énergie électrostatique [Jac98] emmagasinée dans la boucle de longueur L

$$W = \int_0^L \lambda(s) \Phi(s) ds$$

⁶Le dispositif schématisé dans le médaillon de la figure 4.1 suggère comment une telle torsion peut être appliquée.

se calcule à partir :

- de la densité linéique de charges de la molécule

$$\lambda(s) = N \times \chi(-2e) \times f_P(s)$$

où $f_P(s)$ est une fonction continue (sans dimension) localisant les N groupements phosphates⁷ contenus dans L ;

- et du potentiel électrostatique créée par la molécule cyclisée

$$\Phi(s) = \frac{1}{4\pi\epsilon} \int_0^L \frac{\lambda(s')}{|\vec{u}(s) - \vec{u}(s')|} \exp\left[-\frac{|\vec{u}(s) - \vec{u}(s')|}{\ell_D}\right] ds'$$

où $\vec{u}(s)$ est le vecteur position du WLC. Le potentiel coulombien est plus ou moins écranté selon le solvant. Cet écrantage correspond au facteur exponentiellement décroissant sur la longueur caractéristique ℓ_D , dite « de Debye ». ⁸ En pratique, il n'est pas judicieux de prendre en compte la boucle entière : les valeurs de A et C usuelles⁹ contiennent déjà une contribution électrostatique [BMS99, BMS03, LWS02, ZRG02]. L'intégrale en s' doit donc être tronquée, le choix de la coupure n'étant pas nécessairement évident. Cela peut aussi justifier *a posteriori* notre omission de la partie non cyclisée de la molécule dans le potentiel ci-dessus.

Cette contribution W doit alors être ajoutée à l'énergie élastique du WLC (avec ou sans torsion). Nous n'attendons d'effets que pour les solvants réalisant un faible écrantage, ou encore pour les configurations de boucle dans lesquelles deux sites de la molécule se retrouvent proches. Une interaction supplémentaire du type *volume exclu* peut alors être rajoutée [SSG05]. La répulsion électrostatique assure déjà une telle interaction, mais il peut être utile de rajouter une interaction prohibant tout chevauchement de la double hélice. Un phénomène spectaculaire typique est le collapse de la molécule, observé avec de très mauvais solvants.

Ces interactions s'expriment en fonction du vecteur position $\vec{u}(s)$ et non du vecteur tangent $\vec{t}(s)$, ce qui rend un traitement analytique exact impossible. Notre algorithme numérique doit pour la même raison être radicalement changé. Dans la littérature, les solutions les plus souvent retenues sont : numériquement, des simulations type dynamique brownienne [MRKL98, KML98] ou Monte Carlo [UJPS97, BS00, RHMM03] ; analytiquement, l'application du champ moyen [HT99, HT03] ou de méthode assimilées [GCM01, HSPP99, HP01]. Notons que Balaeff et coll. ont aussi adapté leur algorithme [BMS99, BMS03]. Nous trouvons également : des traitements perturbatifs utilisant la méthode variationnelle [MN04], une adaptation du développement de Wilhelm et Frey (complétée numériquement) [ZRG02, ZRG03a, ZRG03b], ainsi qu'une tentative de reproduction des effets électrostatiques

⁷En fermant proprement la boucle, nous avons autant de groupements que de bases.

⁸L'expression du potentiel constitue l'approximation de Debye/Hückel, issue d'une linéarisation de l'équation de Poisson/Boltzmann [STRJ02].

⁹En procédant par ajustements aux données expérimentales.

et de volume exclu à partir d'une superposition de WLC neutres [CP02, CP04]. Il a aussi été montré l'utilité d'une longueur de persistance effective dans les conditions de fort écrantage¹⁰ [ZRG02, ZRG03b]. Une telle approche doit toutefois être envisagée avec prudence [MK96, LW95].

Pour finir insistons sur la non pertinence des effets de volume exclu aux faibles longueurs $L \ll A$, la rigidité élastique rendant alors particulièrement difficile la courbure du polymère. Dans le cas d'une boucle, il se peut que la protéine impose des configurations présentant des chevauchements, ce même aux faibles longueurs. Dans ce cas, nous pouvons toutefois écarter empiriquement de telles boucles et nous concentrer sur celles qui apparaissent plus réalistes [BMS99, BMS03, BMS04].

7.4 Défauts dans la double hélice et effets de séquence

En ce qui concerne la modélisation de cassures provoquées, par l'accrochage de protéines ou l'apparition d'un quelconque autre défaut dans la structure idéale en double hélice de l'ADN, nous pouvons généraliser nos résultats à l'ensemble des cas mentionnés au chapitre 5 : multiples cassures, spécifiques en sites ou non, permanentes ou activées thermiquement, *a priori* semi-flexible et présentant une amplitude de cassure préférentielle. La plupart de ces études ne demandent que de faibles modifications de notre algorithme.

Se lancer dans un tel travail de nature quasi-encyclopédique, en l'absence de données expérimentales systématiques sur le complexe que forme la protéine étudiée avec la molécule d'ADN, apparaît peu pertinent. Soulignons de plus qu'une grande quantité d'algorithmes similaires au nôtre ont été proposés ces derniers mois dans la littérature. L'ensemble adresse déjà la majorité des cas envisageables, les exceptions ne tarderont certainement pas à être traitées. Peut être serait-il bon de prendre quelque recul sur ce sujet : les expériences (fin 2004) ayant relancé l'intérêt théorique se révèlent controversées. Le problème présente un intérêt fondamental indéniable, mais pratique faible (discutez en avec un expérimentateur !). Il pourrait être utile de préciser quantitativement le rôle de la relation de force/extension ou du facteur de cyclisation comme indicateur¹¹ sur les propriétés d'un accrochage spécifique : le taux d'occupation de la molécule par les protéines, la rigidité du complexe formé, etc. En attendant notons que le cas d'une unique cassure permanente, rigide et placée à mi-longueur apparaît comme une bonne première caractérisation du phénomène.

Les effets de séquences peuvent entraîner des défauts ponctuels dans la double

¹⁰Ou de faible interaction. Pour notre part, nous avons ci-dessus considéré la densité de charges comme donnée du problème.

¹¹Pouvoir procéder par simple ajustement, comme pour la longueur de persistance.

hélice. De tels défauts peuvent également être en phases (leur effet s'additionne) [CHN90]. Dans ces deux cas nous nous ramenons à la situation traitée ci-dessus. En revanche, si nous pouvons découper la molécule en plusieurs fragments de longueurs diverses, et de rigidités effectives prenant en compte leur contenu en paires de bases (dominante A-T ou C-G ?), nous n'avons pas à calculer le propagateur associé à chaque fragments au contraire de la cassure à mi-longueur. C'est une conséquence du principe de superposition et de la description déjà effective des effets de séquences proposée. Pour trois de ces fragments

$$\int \frac{d\Omega_0 d\Omega_3}{4\pi} d\Omega_1 d\Omega_2 Z_{3,2} Z_{2,1} Z_{1,0} = \sum_{\ell, m; \ell', m'} \left\{ \left[\int d\Omega_3 \frac{1}{\sqrt{4\pi}} Y_\ell^m(\Omega_3) \right] \langle \ell, m | \times \right. \\ \times e^{\frac{L_{2,3}}{A_{2,3}} \hat{H}} \left(\int d\Omega_2 |\Omega_2\rangle \langle \Omega_2| \right) e^{\frac{L_{1,2}}{A_{1,2}} \hat{H}} \left(\int d\Omega_1 |\Omega_1\rangle \langle \Omega_1| \right) e^{\frac{L_{0,1}}{A_{0,1}} \hat{H}} \times \\ \left. \times | \ell', m' \rangle \int d\Omega_0 \frac{1}{\sqrt{4\pi}} \tilde{Y}_{\ell'}^{m'}(\Omega_0) \right\} = \langle 0, 0 | \exp \left[- \left(\frac{L_{2,3}}{A_{2,3}} + \frac{L_{1,2}}{A_{1,2}} + \frac{L_{0,1}}{A_{0,1}} \right) \hat{H} \right] | 0, 0 \rangle$$

où nous avons utilisé la commutation de \hat{H} avec lui-même. La somme des longueurs réduites peut s'écrire L/A_{eff} . Si les trois fragments ont la même longueur $L/3$ et que la rigidité A' du milieu diffère de celle A des extrémités, alors $A_{\text{eff}} = \frac{3A}{2+A/A'}$. La rigidification maximale $A_{\text{eff}} = \frac{3}{2} A$ correspond à $A' \gg A$, tandis que pour obtenir $A_{\text{eff}} = A/2$ nous devons prendre $A' = A/4$ (pas de limite pour la flexibilité). Une étude quantitative doit inclure une rigidité ainsi qu'une courbure dépendant de la position s , dans le polymère. Dans sa formulation continue¹², le modèle ainsi obtenu

$$\beta E [\vec{t}; L] = \frac{1}{2} \int_0^L \left\{ A(s) \left[\frac{d\vec{t}(s)}{ds} - \vec{\tau}(s) \right]^2 + C(s) \left[\Omega_3(s) - \omega(s) \right]^2 \right\} ds$$

a été étudié en 1998 pour A et C constants par Nelson [Nel98], puis très récemment repris (2005, prépublication) dans le cas général par Popov et Tkachenko [PT05b]. Dans les deux travaux, les courbures $\vec{\tau}$ et ω sont des bruits blancs.

7.5 Étirement par une force

Une étude quantitative du facteur de cyclisation sous une force d'étirement repose sur une optimisation de notre algorithme numérique. Nous n'avons pas encore trouvé de formulation alternative du problème qui permettrait d'épargner significativement du temps de calcul. Si cette voie est poursuivie, une solution technique consiste peut-être à paralléliser le code.

Notons que nous n'avons pas imposé d'orientations particulières aux extrémités de la boucle. Nous pouvons proposer un choix (§ 7.1) s'accordant mieux aux

¹²Qui n'est peut être pas la plus pertinente pour juger des effets de séquences.

configurations suggérées, par exemple par les expériences de Lia et coll. [LBC⁺03]. Nous avons également souligné l'importance de considérer l'effet de la force sur le voisinage de la boucle, selon la configuration effectivement réalisée (figure 6.6).

7.6 Mais aussi. . .

Le facteur de cyclisation est une propriété statique du WLC. Les résultats obtenus dans ce manuscrit peuvent servir à étudier la dynamique de fermeture de la boucle. Nous pouvons considérer un système à deux états : ouvert (boucle) et fermé (pas de boucle). Le calcul des durées de vie moyennes de ces deux états permettrait une comparaison directe aux données de Finzi et coll. [FG95] sur le LacR, ou encore de Lia et coll. [LBC⁺03] sur le GalR. Le poids de Boltzmann associé à la différence d'énergie libre entre les deux états correspond au rapport de leur durée de vie. Comme cette énergie contient notamment le facteur de cyclisation, un traitement analytique exact n'est pas possible.

Nous sommes également intéressés par le temps moyen de cyclisation, c'est-à-dire de contact¹³ entre les extrémités de la future boucle. Le point de départ d'une étude analytique est l'équation de type Fokker/Planck du polymère. En 2002, Dua et Cherayii [DC02a, DC02b] ont étudié l'équation de Smoluchowski (grandes frictions) pour un WLC dont l'inextensibilité est relaxée (champ moyen). En 2003 Jun, Bechhoefer et Ha ont appliqué les résultats de Kramers [vdZ03] en utilisant l'extension r comme degré de liberté. Leur travail [JBH03, Jun04] utilise les résultats de Shimada et Yamakawa, du champ moyen ainsi que la formule interpolative de Rippe et Ringrose. Il met en évidence un comportement, en fonction de la longueur de la boucle, identique au facteur de cyclisation (minimum du temps moyen de cyclisation pour $L \simeq 500$ bp). Cette correspondance était déjà utilisée dans les simulations (numériques donc) de dynamiques browniennes.

Mentionnons pour finir le récent article de Sankararaman et Marko [SM05a] qui, à l'aide du résultat de Shimada et Yamakawa, ont envisagé la formation de plusieurs boucles (interagissantes et auto-évitantes) le long de la molécule d'ADN. Quand la densité est importante, l'arrangement adopté par les boucles peut alors significativement réduire la longueur de boucle la plus probable.

¹³*Modulo* un certain volume de réaction.

Annexe A

Analogie entre WLC et rotateur rigide quantique

Le WLC est une courbe différentiable, décrite par son vecteur tangent $\vec{t}(s)$. Nous assimilons l'abscisse curviligne s avec un temps et décidons de voir le vecteur tangent comme un bâtonnet rigide (de longueur unitaire). En balayant la courbe $0 \leq s \leq L$, l'évolution de $\vec{t}(s)$ décrit la courbure du WLC. Alternativement nous observons les rotations du bâtonnet \vec{t} dans l'espace, sur un intervalle de temps L c'est-à-dire l'évolution d'un *rotateur rigide*. La présence d'une force d'étirement du WLC se traduit par une *polarisation* du bâtonnet (rotations contraintes selon l'axe d'étirement), identique à celle provoquée par exemple par un champ électrique extérieur. Écrivant le hamiltonien du rotateur rigide sous champ, les règles usuelles de quantification permettent de passer au problème quantique correspondant.

Ce raisonnement intuitif n'éclaircit pas pourquoi nous pouvons utiliser le système quantique pour résoudre le problème initial classique. Certains lecteurs savent pertinemment qu'il existe une analogie entre physique statistique (classique) et mécanique quantique. Ces derniers peuvent omettre la prochaine section et se contenter d'une lecture rapide de la section A.2. Pour ceux moins habitués au concept d'*intégrale de chemins*, la section A.1, basée sur [FH65, Har66], permet de se faire une idée minimale de ce qui est raconté dans la suite.

A.1 Formulation lagrangienne de la mécanique quantique

En mécanique quantique, par « résoudre un problème » nous entendons le plus souvent diagonaliser l'opérateur hamiltonien du système considéré : $\hat{H}|n\rangle = E_n|n\rangle$. Nous disposons de règles de quantification pour construire \hat{H} à partir du hamiltonien du système classique. Est-il possible de proposer une formulation *lagrangienne* de la mécanique quantique ?

Nous en illustrons le principe pour une particule à un degré de liberté x soumise à un potentiel $V(x)$. Sachant qu'elle se trouve à l'instant t' en x' , quelle est la probabilité de la trouver en x'' à t'' ? Cette probabilité conditionnelle est le *propagateur* (ou encore noyau, kernel) de la particule

$$Z(x'', t'' | x', t') = \langle x'', t'' | x', t' \rangle = \langle x'' | \exp \left[-\frac{i}{\hbar} (t'' - t') \widehat{H} \right] | x' \rangle.$$

Nous utilisons alternativement la représentation $|x\rangle$ dans le point de vue de Heisenberg $|x, t\rangle$, ou l'opérateur d'évolution $\exp \left[-\frac{i}{\hbar} (t'' - t') \widehat{H} \right]$ de la particule. Nous discrétisons un chemin possible $x(t)$ menant de (x', t') à (x'', t'') , en N intervalles de pas en temps ϵ constant

$$\begin{aligned} t' &= t_0, t_1, \dots, t_n, \dots, t_{N-1}, t'' = t_N. \\ x' &= x_0, x_1, \dots, x_n, \dots, x_{N-1}, x'' = x_N. \end{aligned}$$

Nous introduisons $(N - 1)$ relations de fermeture $\mathbf{1} = \int dx_n |x_n, t_n\rangle \langle x_n, t_n|$

$$Z(x'', t'' | x', t') = \lim_{\epsilon \rightarrow 0} \int \left(\prod_{n=1}^{N-1} dx_n \right) \left(\prod_{n=0}^{N-1} \langle x_{n+1}, t_{n+1} | x_n, t_n \rangle \right).$$

Par « $\lim_{\epsilon \rightarrow 0}$ » nous signifions $\epsilon \rightarrow 0$, mais aussi $N \rightarrow \infty$ tout en gardant $N \times \epsilon = t'' - t'$ constant. Nous avons obtenu un principe de *superposition* : le propagateur est la somme sur tous les chemins possibles, chacun affecté d'un poids statistique. Feynman proposa le *postulat* suivant

$$\langle x_{n+1}, t_{n+1} | x_n, t_n \rangle = \alpha(\epsilon) \exp \left[\frac{i}{\hbar} S(x_n, t_n; x_{n+1}, t_{n+1}) \right]$$

utilisant le lagrangien de la particule à travers son action

$$S(x_n, t_n; x_{n+1}, t_{n+1}) = \int_{t_n}^{t_{n+1}} \left\{ \frac{m}{2} \left[\frac{dx(t)}{dt} \right]^2 - V[x(t)] \right\} dt.$$

Sous ce postulat le propagateur se réécrit comme l'*intégrale des chemins*

$$\begin{aligned} Z(x'', t'' | x', t') &= \lim_{\epsilon \rightarrow 0} \alpha^N(\epsilon) \int \left(\prod_{n=1}^{N-1} dx_n \right) \exp \left[\frac{i}{\hbar} \sum_{n=0}^{N-1} S(x_n, t_n; x_{n+1}, t_{n+1}) \right] \\ &= \int \mathcal{D}x(t) \exp \left\{ \frac{i}{\hbar} S[x(t)] \right\}. \end{aligned}$$

Quel crédit apporter à cette formulation du problème quantique à partir du système classique? Nous considérons la propagation de la fonction d'onde

$$\psi(x'', t'') = \int_{-\infty}^{\infty} Z(x'', t'' | x', t') \psi(x', t') dx'$$

sur l'intervalle de temps $(t'' - t') = \epsilon$, correspondant à $(x'' - x') = \Delta$. Pour un potentiel suffisamment régulier, l'action s'approxime

$$S(x'' - \Delta, t'; x'', t' + \epsilon) \simeq \frac{m \Delta^2}{2 \epsilon} - \epsilon V(x'').$$

Il suffit de développer les fonctions d'onde aux ordres

$$\begin{aligned} \psi(x'', t' + \epsilon) &= \psi(x'', t') + \epsilon \frac{\partial \psi}{\partial t}(x'', t') + O(\epsilon^2), \\ \psi(x'' - \Delta, t') &= \psi(x'', t') - \Delta \frac{\partial \psi}{\partial x}(x'', t') + \frac{\Delta^2}{2} \frac{\partial^2 \psi}{\partial x^2}(x'', t') + O(\Delta^3). \end{aligned}$$

Quelques intégrales gaussiennes plus tard nous identifions à l'« ordre 0 » en ϵ la constante de proportionnalité $\alpha(\epsilon) = \sqrt{2i\pi\hbar\epsilon/m}$. L'« ordre 1 » donne

$$i\hbar \frac{\partial \psi}{\partial t}(x'', t') = -\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2}(x'', t') - V(x'') \psi(x'', t'),$$

c'est-à-dire l'équation de Schrödinger usuelle de la particule libre sous potentiel.

A.2 Équation de Schrödinger du WLC

Nous discrétisons une configuration possible $\vec{t}(s)$ du WLC en une chaîne de N spins *unitaires*, espacés de ϵ le long de la courbe originelle

$$\begin{aligned} \vec{t}' &= \vec{t}_0, \vec{t}_1, \dots, \vec{t}_n, \dots, \vec{t}_{N-1}, \vec{t}'' = \vec{t}_N. \\ s' &= s_0, s_1, \dots, s_n, \dots, s_{N-1}, s'' = s_N. \end{aligned}$$

La fonction de partition du WLC s'écrit

$$Z(\Omega_L, L|\Omega_0, 0; f) = \lim_{\epsilon \rightarrow 0} \alpha^N(\epsilon) \int \left(\prod_{n=1}^{N-1} d\Omega_n \right) \left[\prod_{n=0}^{N-1} e^{-\beta E(\Omega_n, s_n; \Omega_{n+1}, s_{n+1}; f)} \right].$$

Le poids de Boltzmann à pris la place du postulat de Feynman utilisant l'action comme phase (§ A.1). La fonction de partition s'interprète comme la probabilité pour le WLC d'avoir une orientation finale \vec{t}_L sachant que l'initiale est \vec{t}_0 . L'énergie couplant deux monomères voisins est

$$\begin{aligned} \beta E(\Omega_n, s_n; \Omega_{n+1}, s_{n+1}; f) &= \frac{A}{2} \int_{s_n}^{s_{n+1}} \left[\frac{d\vec{t}(s)}{ds} \right]^2 ds - \beta f \int_{s_n}^{s_{n+1}} \cos \theta(s) ds \\ &\simeq \frac{A}{\epsilon} (1 - \vec{t}_{n+1} \cdot \vec{t}_n) - \beta \epsilon f \cos \theta_{n+1}. \end{aligned}$$

Notons au passage l'analogie du WLC discrétisé avec le modèle de spins classiques d'Heisenberg. Le produit scalaire entre monomères voisins s'exprime en coordonnées sphériques

$$1 - \vec{t}_{n+1} \cdot \vec{t}_n = 1 - \cos(\theta_{n+1} - \theta_n) + [1 - \cos(\varphi_{n+1} - \varphi_n)] \sin \theta_n \sin \theta_{n+1}.$$

Un calcul identique à celui présenté en section précédente nous permet de dériver une « équation de Schrödinger », pour la probabilité d'observer une orientation $\vec{t}(s)$ en un endroit quelconque s du WLC. Nous partons de la loi de composition

$$\psi(\Omega'', s''; f) = \int d\Omega' Z(\Omega'', s'' | \Omega', s'; f) \psi(\Omega', s'; f).$$

Nous l'appliquons sur une longueur $s'' - s' = \epsilon$, correspondant à $\varphi'' - \varphi' = \Delta\varphi$ et $\theta'' - \theta' = \Delta\theta$ (abrégé en $\Omega'' - \Omega' = \Delta\Omega$). Il suffit de développer les probabilités aux ordres

$$\begin{aligned} \psi(\Omega'', s' + \epsilon; f) &= \psi(\Omega'', s'; f) + \epsilon \frac{\partial \psi}{\partial s}(\Omega'', s'; f) + O(\epsilon^2) \\ \psi(\Omega'' - \Delta\Omega, s'; f) &= \psi(\Omega'', s'; f) - \Delta\varphi \frac{\partial \psi}{\partial \varphi}(\Omega'', s'; f) - \Delta\theta \frac{\partial \psi}{\partial \theta}(\Omega'', s'; f) \\ &\quad + \frac{\Delta\varphi^2}{2} \frac{\partial^2 \psi}{\partial \varphi^2}(\Omega'', s'; f) + \frac{\Delta\theta^2}{2} \frac{\partial^2 \psi}{\partial \theta^2}(\Omega'', s'; f) \\ &\quad + \frac{\Delta\theta \Delta\varphi}{2} \frac{\partial^2 \psi}{\partial \theta \partial \varphi}(\Omega'', s'; f) + O(\Delta\theta^{3-m} \Delta\varphi^m). \end{aligned}$$

Nous développons aussi dans la mesure d'intégration et l'énergie

$$\begin{aligned} \sin(\theta'' - \Delta\theta) &= \sin \theta'' - \Delta\theta \cos \theta'' - \frac{\Delta\theta^2}{2} \sin \theta'' + O(\Delta\theta^3) \\ 1 - \cos \Delta\varphi &= \frac{\Delta\varphi^2}{2} - \frac{\Delta\varphi^4}{4!} + O(\Delta\varphi^6) \quad (\text{idem } 1 - \cos \Delta\theta). \end{aligned}$$

En conservant les termes pertinents, nous nous ramenons au calcul d'intégrales gaussiennes. Une fois effectuées nous identifions à l'« ordre 0 » en ϵ la constante de proportionnalité $\alpha(\epsilon) = A/(2\pi\epsilon)$. L'« ordre 1 » donne pour $\psi(\Omega'', s'; f)$

$$\frac{\partial \psi}{\partial s} = \frac{1}{2A} \underbrace{\left(\frac{1}{\tan \theta''} \frac{\partial \psi}{\partial \theta} + \frac{\partial^2 \psi}{\partial \theta^2} + \frac{1}{\sin^2 \theta''} \frac{\partial^2 \psi}{\partial \varphi^2} \right)}_{\text{opérateur moment cinétique : } -\langle \Omega | \widehat{L}^2 | \psi \rangle} + \beta f \cos \theta'' \psi,$$

soit l'équation de Schrödinger d'un rotateur rigide de moment d'inertie A , soumis à un champ électrique. Nous pouvons alors écrire la fonction de partition du WLC comme le propagateur

$$Z(\Omega_L, L | \Omega_0, 0; f) = \langle \Omega_L, L | \Omega_0, 0 \rangle = \langle \Omega_L | \exp \left[-\frac{L}{A} \left(\frac{\widehat{L}^2}{2} - \beta A f \widehat{\cos \theta} \right) \right] | \Omega_0 \rangle.$$

présentant l'invariance par translation le long du polymère. Pour finir, signalons qu'une approximation harmonique en $\Delta\varphi$ et $\Delta\theta$ de l'énergie

$$\beta E(\Omega', s'; \Omega'', s''; f) \simeq \frac{A}{2\epsilon} (\Delta\theta^2 + \Delta\varphi^2 \sin \theta' \sin \theta'') - \beta \epsilon f \cos \theta''$$

est insuffisante car elle produit le terme supplémentaire $\psi/(4A \sin^2 \theta'')$ dans l'équation de Schrödinger. Le résultat correct est récupérable en considérant pour

le WLC discret, un potentiel effectif tel que

$$\beta E(\Omega', s'; \Omega'', s''; f) \simeq \frac{A}{2\epsilon} (\Delta\theta^2 + \Delta\varphi^2 \sin\theta' \sin\theta'') - \frac{\epsilon}{4A \sin^2\theta''} - \beta\epsilon f \cos\theta''.$$

Cette règle de discrétisation est préférée par certains auteurs [Yam97, SY84].¹

¹On peut également symétriser $\sin\theta' \sin\theta''$ en $\sin^2[\frac{1}{2}(\theta' + \theta'')]$.

Annexe B

Opérateurs $\widehat{\cos \theta}$ et $\widehat{\cos \varphi \sin \theta}$ développés sur $Y_\ell^m(\Omega)$

En section B.1 nous calculons les composantes, dans la bases des harmoniques sphériques, des opérateurs $\widehat{\cos \theta}$ et $\widehat{\cos \varphi \sin \theta}$ intervenant dans l'évaluation numérique des distributions $P(z; L, \kappa)$ (§§ 4.2, 5.3) et $\mathcal{P}(x, z; L, f)$ (§ 6.1). Une section supplémentaire (§ B.2) reproduit quelques résultats bien connus sur les harmoniques sphériques.

B.1 Composantes des opérateurs

Nous avons besoin des intégrales suivantes

$$\begin{aligned}\langle \ell, m | \widehat{\cos \theta} | \ell', m' \rangle &= \int d\Omega \tilde{Y}_\ell^m(\Omega) \cos \theta Y_{\ell'}^{m'}(\Omega) \\ \langle \ell, m | \widehat{\cos \varphi \sin \theta} | \ell', m' \rangle &= \int d\Omega \tilde{Y}_\ell^m(\Omega) \cos \varphi \sin \theta Y_{\ell'}^{m'}(\Omega)\end{aligned}$$

L'astuce consiste à reconnaître les harmoniques sphériques

$$\cos \theta = \sqrt{\frac{4\pi}{3}} Y_1^0(\Omega) \quad \cos \varphi \sin \theta = -\sqrt{\frac{8\pi}{3}} \operatorname{Re} [Y_1^1(\Omega)].$$

Nous utilisons alors la règle de sélection

$$\begin{aligned}\int d\Omega \tilde{Y}_j^m(\Omega) Y_{j_1}^{m_1}(\Omega) Y_{j_2}^{m_2}(\Omega) &= \sqrt{\frac{(2j_1+1)(2j_2+1)}{4\pi(2j+1)}} \times \\ &\times \langle j_1 j_2 0 0 | j_1 j_2 j 0 \rangle \langle j_1 j_2 m_1 m_2 | j_1 j_2 j m \rangle.\end{aligned}$$

Le coefficient de Clebsch-Gordan $\langle j_1 j_2 m_1 m_2 | j_1 j_2 j m \rangle$, abrégé en CG, est non nul si ses arguments obéissent aux règles d'addition de deux moments cinétiques quantiques¹, à savoir

$$m_1 + m_2 = m \quad \text{et} \quad |j_1 - j_2| \leq j \leq j_1 + j_2.$$

¹Rappel : $j = 0, 1/2, 1, 3/2, \dots$ (valeurs entières seulement pour ℓ) et $|m| \leq j$.

En pratique nous devons calculer que le deuxième CG : il se réduit au premier CG pour $m_1 = m_2 = m = 0$ qui impose donc la même sélection sur j et ne sert, en quelques sortes, que de normalisation. Prenons le premier opérateur

$$\langle \ell, m | \widehat{\cos \theta} | \ell', m' \rangle = \sqrt{\frac{2\ell' + 1}{2\ell + 1}} \langle 1\ell'00 | 1\ell'\ell 0 \rangle \langle 1\ell'0m' | 1\ell'\ell m \rangle.$$

L'expression est non nulle pour $m = m'$, et $\ell = (\ell' - 1), \ell', (\ell' + 1)$ si $\ell' \geq 1$, $\ell = 1$ si $\ell' = 0$. Les CG correspondant sont

$$\begin{aligned} \langle 1\ell'0m' | 1\ell'(\ell' - 1)m' \rangle &= -\sqrt{\frac{(\ell' - m')(\ell' + m')}{\ell'(2\ell' + 1)}} \quad (\text{nul si } \ell' = 0) \\ \langle 1\ell'0m' | 1\ell'\ell'm' \rangle &= -\sqrt{\frac{m'}{\ell'(\ell' + 1)}} \quad (\text{nul si } \ell' = 0) \\ \langle 1\ell'0m' | 1\ell'(\ell' + 1)m' \rangle &= \sqrt{\frac{(\ell' + 1 - m')(\ell' + 1 + m')}{(\ell' + 1)(2\ell' + 1)}}. \end{aligned}$$

La déduction du premier CG ($m' = 0$) montre en fait que les composantes $\ell = \ell'$ de l'opérateur sont toujours nulles. Nous pouvons résumer ces résultats en une unique expression grâce au symbole de Kronecker $\delta_{i,j}$. Nous l'écrivons de manière à rendre explicite sa symétrie

$$\langle \ell, m | \widehat{\cos \theta} | \ell', m' \rangle = \frac{\sqrt{(\ell' - m')(\ell' + m')} \delta_{\ell, \ell' - 1} + \sqrt{(\ell - m)(\ell + m)} \delta_{\ell, \ell' - 1}}{\sqrt{(2\ell' + 1)(2\ell + 1)}} \delta_{m, m'}.$$

Comme $\text{Re}[Y_1^1(\Omega)] = [Y_1^1(\Omega) - Y_1^{-1}(\Omega)]/2$, le calcul des composantes du deuxième opérateur demande *a priori* l'évaluation de deux intégrales. En fait, nous n'avons besoin que d'en calculer une car (notant $\text{c.c.}[\dots] = \overline{\dots}$)

$$\int d\Omega \tilde{Y}_\ell^m(\Omega) Y_1^{-1}(\Omega) Y_{\ell'}^{m'}(\Omega) = -\text{c.c.} \left[\int d\Omega \tilde{Y}_{\ell'}^{m'}(\Omega) Y_1^1(\Omega) Y_\ell^m(\Omega) \right].$$

Considérons alors l'intégrale

$$\sqrt{\frac{8\pi}{3}} \int d\Omega \tilde{Y}_\ell^m(\Omega) Y_1^1(\Omega) Y_{\ell'}^{m'}(\Omega) = \sqrt{2} \sqrt{\frac{2\ell' + 1}{2\ell + 1}} \langle 1\ell'00 | 1\ell'\ell 0 \rangle \langle 1\ell'1m' | 1\ell'\ell m \rangle.$$

Elle est non nulle pour $m = m' + 1$, et $\ell = (\ell' - 1), \ell', (\ell' + 1)$ si $\ell' \geq 1$, $\ell = 1$ si $\ell' = 0$. Le CG « de normalisation » n'a pas changé, et il annule donc l'intégrale de $\ell = \ell'$. Les CG restants sont

$$\begin{aligned} \langle 1\ell'1m' | 1\ell'(\ell' - 1)(m' + 1) \rangle &= \sqrt{\frac{(\ell' - m')[\ell' - (m' + 1)]}{2\ell'(2\ell' + 1)}} \quad (\text{nul si } \ell' = 0) \\ \langle 1\ell'1m' | 1\ell'(\ell' + 1)(m' + 1) \rangle &= \sqrt{\frac{(\ell' + 1 + m')(\ell' + 1 + m' + 1)}{2(\ell' + 1)(2\ell' + 1)}}. \end{aligned}$$

Les valeurs de l'intégrale se résume (ne pas oublier la normalisation) en

$$\begin{aligned} \sqrt{\frac{8\pi}{3}} \int d\Omega \tilde{Y}_\ell^m(\Omega) Y_1^1(\Omega) Y_{\ell'}^{m'}(\Omega) &= \\ &= \frac{-\sqrt{(\ell' - m')(\ell' - m)} \delta_{\ell, \ell' - 1} + \sqrt{(\ell + m)(\ell + m')} \delta_{\ell, \ell' - 1}}{\sqrt{(2\ell + 1)(2\ell' + 1)}} \delta_{m, m' + 1}. \end{aligned}$$

Elle est réelle : la deuxième intégrale est donc l'opposée de cette expression, dans laquelle les paires d'indices (ℓ, m) et (ℓ', m') ont été permutées. La somme ou la différence de ces deux intégrales fournissent les composantes des opérateurs

$$\begin{aligned} \langle \ell, m | \widehat{\cos \varphi \sin \theta} | \ell', m' \rangle &= \int d\Omega \tilde{Y}_\ell^m(\Omega) \sqrt{\frac{8\pi}{3}} \left[-\frac{Y_1^1(\Omega) - Y_1^{-1}(\Omega)}{2} \right] Y_{\ell'}^{m'}(\Omega) \\ \langle \ell, m | \widehat{\sin \varphi \sin \theta} | \ell', m' \rangle &= \int d\Omega \tilde{Y}_\ell^m(\Omega) \sqrt{\frac{8\pi}{3}} \left[-\frac{Y_1^1(\Omega) + Y_1^{-1}(\Omega)}{2i} \right] Y_{\ell'}^{m'}(\Omega). \end{aligned}$$

Nous avons (arbitrairement) choisit d'utiliser le premier.² En réarrangeant les termes, nous obtenons

$$\begin{aligned} \langle \ell, m | \widehat{\cos \varphi \sin \theta} | \ell', m' \rangle &= -\frac{1}{2} \frac{1}{\sqrt{(2\ell + 1)(2\ell' + 1)}} \times \\ &\times \left\{ \left[\sqrt{(\ell' + m')(\ell' + m)} \delta_{m, m' - 1} - \sqrt{(\ell' - m')(\ell' - m)} \delta_{m', m - 1} \right] \delta_{\ell, \ell' - 1} + \right. \\ &\left. + \left[\sqrt{(\ell + m)(\ell + m')} \delta_{m', m - 1} - \sqrt{(\ell - m)(\ell - m')} \delta_{m, m' - 1} \right] \delta_{\ell, \ell' - 1} \right\}. \end{aligned}$$

B.2 Rappels sur les harmoniques sphériques

Les harmoniques sphériques $Y_\ell^m(\Omega) = \langle \Omega | \ell, m \rangle$ sont les fonctions d'onde propres des opérateurs de moment cinétique quantique ($\hbar = 1$)

$$\begin{aligned} \langle \Omega | \widehat{L}^2 | \ell, m \rangle &= - \left(\frac{1}{\tan \theta} \frac{\partial}{\partial \theta} + \frac{\partial^2}{\partial \theta^2} + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \varphi^2} \right) Y_\ell^m(\Omega) = \ell(\ell + 1) Y_\ell^m(\Omega) \\ \langle \Omega | \widehat{L}_z | \ell, m \rangle &= \left(-i \frac{\partial}{\partial \varphi} \right) Y_\ell^m(\Omega) = m Y_\ell^m(\Omega). \end{aligned}$$

Les valeurs propres de \widehat{L}^2 sont indexées par l'entier positif ou nul ℓ , tandis que celles de \widehat{L}_z le sont par l'entier m variant entre $(-\ell)$ et ℓ . Les harmoniques sphériques forment une base orthonormée

$$\langle \ell, m | \ell', m' \rangle = \int d\Omega \tilde{Y}_\ell^m(\Omega) Y_{\ell'}^{m'}(\Omega) = \delta_{\ell, \ell'} \delta_{m, m'}.$$

²Le deuxième interviendrait dans le calcul de la distribution en extensions y et z qui est, par symétrie cylindrique, égale à celle en x et z .

Leurs dépendances en θ et φ se séparent

$$Y_\ell^m(\Omega) = \sqrt{\frac{2\ell+1}{4\pi} \frac{(\ell-m)!}{(\ell+m)!}} \mathbb{P}_\ell^m(\cos\theta) \exp(im\varphi),$$

où $\mathbb{P}_\ell^m(x)$ est le polynôme de Legendre généralisé (réel). Si $m = 0$ on ne parle plus que du polynôme de Legendre, simplement noté $\mathbb{P}_\ell(x)$.³ Les polynômes de Legendre sont orthogonaux mais ne sont pas normalisés, comme le montre l'orthonormalisation des harmoniques sphériques. Nous voyons également que l'harmonique sphérique de $m = 0$ est réelle et ne dépend que de $\cos\theta$ (et pas de φ). Les polynômes de Legendre généralisés sont tels que le complexe conjugué d'une harmonique sphérique se réduit à $\tilde{Y}_\ell^m(\Omega) = (-1)^m Y_\ell^{-m}(\Omega)$.

Explicitons pour finir quelques unes de ces harmoniques sphériques

$$\begin{aligned} Y_0^0(\Omega) &= \frac{1}{\sqrt{4\pi}} & Y_2^2(\Omega) &= \sqrt{\frac{15}{32\pi}} \sin^2\theta \exp(i2\varphi) \\ Y_1^1(\Omega) &= -\sqrt{\frac{3}{8\pi}} \sin\theta \exp(i\varphi) & Y_2^1(\Omega) &= -\sqrt{\frac{15}{8\pi}} \sin\theta \cos\theta \exp(i\varphi) \\ Y_1^0(\Omega) &= \sqrt{\frac{3}{4\pi}} \cos\theta & Y_2^0(\Omega) &= \sqrt{\frac{5}{16\pi}} (3\cos^2\theta - 1), \end{aligned}$$

dans lesquelles apparaissent les polynômes de Legendre

$$\mathbb{P}_0(x) = 1 \qquad \mathbb{P}_1(x) = x \qquad \mathbb{P}_2(x) = \frac{1}{2} (3x^2 - 1).$$

Cette liste illustre la parité des polynômes de Legendre $\mathbb{P}_\ell(-x) = (-1)^\ell \mathbb{P}_\ell(x)$.

³Qui est bien un polynôme en x (degré ℓ), contrairement à $\mathbb{P}_\ell^m(x)$.

Annexe C

Configuration de point col

Nous montrons ici comment le paramètre x (ou $1/k$) caractérise la configuration d'énergie minimale d'un WLC dont on fixe la distance bout à bout relative r/L et l'orientation initiale $\omega(0) = \alpha/2$ (§ 2.2). Nous dérivons pour cela l'équation de sa configuration. Nous considérons d'emblée la présence d'une cassure à mi-longueur – également fixée – correspondant à une orientation « finale » (de la moitié de polymère) : $\omega(L/2) = a/2$. Le lecteur n'ayant lu que le chapitre 4 peut suivre cette annexe en prenant $a/2 = \pi$ (ou $b = \Omega(L/2) = 0$), l'interprétation de a lui étant fournie au chapitre 5.

En sections 2.2 ou 5.2 l'intégration sur $0 \leq s \leq L/2$, de l'intégrale première du mouvement du rotateur rigide sous champ, a produit les équations intégrales (2.2) ou (5.1). Son intégration sur $s' \leq s \leq L/2$ produit la relation générale

$$\left(1 - \frac{s'}{L/2}\right) \xi = \int_b^{\Omega(s')} \frac{d\Omega}{\sqrt{1 - k^2 \sin^2 \Omega}} = -F(b, k) + F[\Omega(s'), k]$$

dans laquelle nous explicitons ξ puis nous effectuons la transformation de Jacobi réelle (§§ 2.2 ou 5.2) afin de couvrir $x = 1/k < 1$. Nous arrivons à

$$\frac{s'}{L/2} F[B_{L/2}(x), x] + \left(1 - \frac{s'}{L/2}\right) F[B_0(x), x] = F[B(x, s'), x],$$

avec les fonctions $B_0(x) = B(x, 0)$ et $B_{L/2}(x) = B(x, L/2)$ issues de $B(x, s') = \arcsin[\sin \Omega(s')/x]$. Ainsi $a/2 = \pi$ annule les fonctions $B_{L/2}(x)$ et $F[B_{L/2}(x), x]$, quelque soit le paramètre x . La fonction sinus elliptique de Jacobi¹ permet de réécrire cette relation

$$\sin \Omega(s') = x \operatorname{sn} \left\{ \frac{s'}{L/2} F[B_{L/2}(x), x] + \left(1 - \frac{s'}{L/2}\right) F[B_0(x), x], x \right\}.$$

Mais $\sin^2 \Omega(s') = \frac{1}{2}[1 + \cos \omega(s)]$ (§ 2.2), et cette relation se lit donc en $\omega(s)$

$$\cos \omega(s') = 2x^2 \operatorname{sn}^2 \left\{ \frac{s'}{L/2} F[B_{L/2}(x), x] + \left(1 - \frac{s'}{L/2}\right) F[B_0(x), x], x \right\} - 1$$

¹Si $F(B, x) = f$, la fonction sinus elliptique de Jacobi donne $\operatorname{sn}(f, x) = \sin B$ [AS70].

d'où nous pouvons déduire une relation analogue pour $\sin \omega(s')$.² L'équation de la configuration est alors la courbe paramétrée

$$\vec{X}(s) = \left[\frac{r}{2} + \int_0^s \cos \omega(s') ds' \right] \vec{e}_r + \left[\int_0^s \sin \omega(s') ds' \right] \vec{e}_\perp.$$

Une fois fixées la distance bout à bout relative, l'orientation initiale et une éventuelle cassure à mi-longueur du polymère, sa configuration d'énergie minimale est en effet caractérisée par le paramètre x . La relation entre l'orientation initiale α_* du point col et son paramètre x_* implique $B_0(x) = \frac{\pi}{2}$, et le F correspondant est alors l'intégrale elliptique de Legendre complète du premier type (§§ 2.2 ou 5.2).

²Il existe aussi des fonctions cosinus et tangente elliptiques de Jacobi [AS70]. Numériquement nous conseillons d'en choisir une seule des trois (ici le sinus).

Annexe D

Relation entre distributions

$S(r; L)$ et $P(z; L)$

Cette annexe démontre (§ D.1) la relation qui existe à force nulle, entre la distribution marginale en extension z et celle isotrope en extension r . Une section supplémentaire (§ D.2) contient des rappels sur la théorie des distributions (ou fonctions généralisées), se focalisant sur celles rencontrées dans ce manuscrit.

D.1 Dérivation de la relation

Comme la longueur du polymère n'intervient pas dans le calcul nous pouvons omettre la dépendance explicite des distributions en L . En l'absence de force la symétrie sphérique rend la distribution en distance bout à bout $Q(\vec{r})$ radiale, c'est-à-dire fonction uniquement de $r = |\vec{r}|$. Sa transformée de Fourier possède alors la même symétrie, et n'est donc fonction que de $k = |\vec{k}|$. Pour $\vec{k} = k \vec{e}_z$ elle se réécrit

$$\int d\vec{r} \exp[-ik(\vec{r} \cdot \vec{e}_z)] Q(\vec{r}) = \int_{-\infty}^{\infty} dz \exp(-ikz) \underbrace{\int \delta[(\vec{r} \cdot \vec{e}_z) - z] Q(\vec{r}) d\vec{r}}_{\text{distribution } P(z)}$$

où nous reconnaissons la transformée de Fourier de la distribution $P(z)$, à savoir la fonction de partition $Z(-ik/\beta)$, modulo la normalisation par $Z(0)$. La distribution $Q(\vec{r})$ est obtenue par transformée de Fourier inverse. Prenant $\vec{r} = r \vec{e}_z$ nous réalisons les deux intégrations angulaires sur \vec{k} , pour obtenir l'expression de la distribution isotrope

$$\begin{aligned} S(r) &= 4\pi r^2 \times \frac{1}{Z(0)} \int_0^{2\pi} \frac{dk_\phi}{2\pi} \int_0^\infty \frac{dk}{2\pi} k^2 Z(-ik/\beta) \int_{-1}^1 \frac{d(\cos k_\theta)}{2\pi} e^{ikr \cos k_\theta} \\ &= 4\pi r^2 \times \frac{1}{Z(0)} \frac{1}{\pi r} \int_0^\infty \frac{dk}{2\pi} k \sin(kr) Z(-ik/\beta). \end{aligned}$$

L'intégrale restante est une transformée de Fourier radiale en dimension 3. Nous pourrions donc calculer la distribution $S(r)$ directement à partir de l'amplitude

du vide $Z(-ik/\beta)$. Nous préférons établir une relation analytique avec la distribution $P(z)$, obtenue par transformée de Fourier « usuelle » (unidimensionnelle) de $Z(-ik/\beta)$. Si nous ne reconnaissons pas $Z(-ik/\beta)$ dans la première relation, la seconde se réécrit

$$\begin{aligned} S(r) &= \frac{r}{\pi} \int_{-\infty}^{\infty} dz P(z) \int_0^{\infty} dk (-ik) e^{-ikz} (e^{ikr} - e^{-ikr}) \\ &= \frac{r}{\pi} \int_{-\infty}^{\infty} dz P(z) \frac{d}{dz} \int_{-\infty}^{\infty} dk \Theta(k) [e^{-ik(z-r)} - e^{-ik(z+r)}], \end{aligned}$$

où la distribution de Heaviside $\Theta(k)$ a été introduite pour étendre le domaine d'intégration de $k \geq 0$ à tout l'axe réel. Sa transformée de Fourier (§ D.2) s'exprime en fonction des distributions de Dirac et de la *valeur principale*

$$\int_{-\infty}^{\infty} dp \Theta(p) e^{-ipq} = \pi \delta(q) - i \text{vp}(1/q).$$

En utilisant la notation de bra-kets, la distribution isotrope s'écrit

$$S(r) = r \langle \delta'(x); P(x+r) - P(x-r) \rangle - \frac{ir}{\pi} \langle \text{vp}'(1/x); P(x+r) - P(x-r) \rangle.$$

Elle est réelle, et le terme impliquant les dérivées de la valeur principale doit donc s'annuler. Nous reconnaissons la distribution *partie finie*

$$\begin{aligned} \langle \text{vp}'(1/x); P(x+r) \rangle &= -\langle \text{pf}(1/x^2); P(x+r) \rangle \\ &= -\int_0^{\infty} \frac{P(x+r) + P(-x+r) - 2P(r)}{x^2} dx. \end{aligned}$$

Comme la distribution $P(z)$ est paire, la partie imaginaire est en effet nulle

$$\langle \text{vp}'(1/x); P(x+r) \rangle = \langle \text{vp}'(1/x); P(x-r) \rangle.$$

Pour la partie réelle : $\langle \delta'(x); P(x-r) \rangle = -P'(r)$. Comme la dérivée de $P(z)$ est impaire, nous obtenons finalement

$$S(r) = -2rP'(r).$$

D.2 Quelques distributions

L'action d'une distribution régulière T_f sur une fonction φ indéfiniment dérivable et à support borné¹, s'écrit à l'aide de bra-kets

$$\langle T_f; \varphi \rangle = \int_{-\infty}^{\infty} f(x) \varphi(x) dx.$$

La distribution est régulière si elle est en fait associée à une fonction f localement sommable. Dans le cas contraire nous parlons de distribution singulière. L'action

¹Propriété essentielle. Rappelons que l'inextensibilité du WLC impose $P(|z| \geq L) = 0$.

d'une distribution T quelconque s'écrit donc $\langle T; \varphi \rangle$. En physique il est usuel d'associer également aux distributions singulières une fonction formelle. L'exemple typique est la *distribution de Dirac*

$$\langle \delta; \varphi \rangle = \int_{-\infty}^{\infty} \delta(x) \varphi(x) dx = \varphi(0) \quad \text{parfois notée} \quad \langle \delta(x); \varphi(x) \rangle.$$

Une distribution est par définition toujours dérivable $\langle T'; \varphi \rangle = -\langle T; \varphi' \rangle$, par généralisation de l'intégration par parties usuelle (φ est bornée). La dérivée d'un Dirac est donc $\langle \delta'; \varphi \rangle = -\varphi'(0)$. La distribution de Dirac est elle-même la dérivée de la *distribution de Heaviside*²

$$\langle \Theta; \varphi \rangle = \int_{-\infty}^{\infty} \Theta(x) \varphi(x) dx = \int_0^{\infty} \varphi(x) dx \quad \text{et} \quad \langle \Theta'; \varphi \rangle = -[\varphi(x)]_0^{\infty} = \langle \delta; \varphi \rangle.$$

car φ est bornée. Elle permet de construire la *fonction porte* $\frac{1}{\Lambda} [\Theta(x) - \Theta(x - \Lambda)]$ et plus généralement d'étendre une intégration sur une demi-droite, à l'axe réel entier. La *valeur principale de Cauchy* est définie comme

$$\langle \text{vp}; \varphi \rangle = \lim_{\epsilon \rightarrow 0^+} \int_{-\infty}^{-\epsilon} \varphi(x) dx + \int_{\epsilon}^{\infty} \frac{\varphi(x)}{x} dx = \lim_{\epsilon \rightarrow 0^+} \int_{\epsilon}^{\infty} \underbrace{\frac{\varphi(x) - \varphi(-x)}{x}}_{=\psi(x)} dx.$$

Le prolongement par continuité $\psi(0) = 2\varphi'(0)$ donne un sens à la distribution

$$\langle \text{vp}; \varphi \rangle = \int_0^{\infty} \frac{\varphi(x) - \varphi(-x)}{x} dx \quad \text{aussi notée} \quad \text{vp}(1/x).$$

Sa dérivée est au signe près la *partie finie de Hadamard*³

$$\langle \text{pf}; \varphi \rangle = \langle \text{vp}; \varphi' \rangle = \lim_{\epsilon \rightarrow 0^+} \left\{ \left[\frac{\varphi(x)}{x} \right]_{-\infty}^{-\epsilon} + \left[\frac{\varphi(x)}{x} \right]_{\epsilon}^{\infty} + \int_{-\infty}^{-\epsilon} \varphi(x) dx + \int_{\epsilon}^{\infty} \frac{\varphi(x)}{x^2} dx \right\}.$$

Comme φ est bornée, ne reste que $-[\varphi(\epsilon) + \varphi(-\epsilon)]/\epsilon \stackrel{(0)}{\sim} -2\varphi(0)/\epsilon$ des termes déjà intégrés. En condensant les deux intégrales, puis en introduisant dans son intégrant $0 = 2\varphi(0)/x^2 - 2\varphi(0)/x^2$

$$\langle \text{pf}; \varphi \rangle = \lim_{\epsilon \rightarrow 0^+} \int_{\epsilon}^{\infty} \underbrace{\frac{\varphi(x) + \varphi(-x) - 2\varphi(0)}{x^2}}_{=\psi(x)} dx - 2\varphi(0) \underbrace{\left(\frac{1}{\epsilon} - \int_{\epsilon}^{\infty} \frac{dx}{x^2} \right)}_{=0, \forall \epsilon}.$$

Le prolongement par continuité $\psi(0) = \varphi''(0)$ donne un sens à la distribution

$$\langle \text{pf}; \varphi \rangle = \int_0^{\infty} \frac{\varphi(x) + \varphi(-x) - 2\varphi(0)}{x^2} dx \quad \text{aussi notée} \quad \text{pf}(1/x^2).$$

Valeur principale et partie finie généralisent les fonctions $1/x$ et $1/x^2$. Elles ont des propriétés analogues, comme $\langle \text{vp}'; \varphi \rangle = -\langle \text{pf}; \varphi \rangle$.

²Aussi appelée *fonction marche*

³Aussi appelée *pseudo fonction* (le symbole est le même).

La transformée de Fourier d'une distribution nécessite quelques précautions. Les notations symboliques (association d'une fonction) peuvent parfois mener, sans contradiction apparente, à des résultats erronnés.⁴ Sans définition ni démonstration rigoureuses, mentionnons les deux transformées de Fourier de distributions que nous rencontrons dans ce travail

$$\delta(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} e^{ikx} \quad \text{et} \quad \frac{1}{2} \left[\delta(x) + \frac{i}{\pi} \text{vp}(1/x) \right] = \int_{-\infty}^{\infty} \frac{dk}{2\pi} e^{ikx} \Theta(k).$$

La première est fréquemment utilisée en électromagnétisme, optique ainsi qu'en mécanique quantique (développement en ondes planes, orthonormalisation au sens large de fonctions d'ondes, etc.). Nous la considérons comme faisant partie du bagage mathématique du lecteur. Bien que moins connue, la seconde se retrouve dans les mêmes disciplines (fonctions de Green quantiques, relations de Kramers-Kronig, etc.). Nous en donnons une démonstration. Une intégration naïve produit

$$\int_0^{\infty} \frac{dk}{2\pi} e^{ikx} = \frac{1}{2\pi} \frac{1}{ix} [e^{ikx}]_{k=0}^{\infty} \leftarrow \text{pas de sens!}.$$

L'idée est d'imposer une décroissance exponentielle à l'infini, en considérant la *régularisation* de la distribution de Heaviside⁵

$$\lim_{\epsilon \rightarrow 0^+} \Theta(k) \exp(-\epsilon k) = \Theta(k)$$

dont la transformée de Fourier a maintenant un sens, pour ϵ fini

$$\int_0^{\infty} \frac{dk}{2\pi} e^{-\epsilon k} e^{ikx} = \frac{1}{2\pi} \frac{1}{ix - \epsilon} [e^{-\epsilon k} e^{ikx}]_{k=0}^{\infty} = \frac{1}{2\pi} \frac{i}{x + i\epsilon}.$$

La réciprocity se vérifie moyennant une intégration complexe sur un contour fermé évitant ou incluant le pôle ($-i\epsilon$) selon que k est négatif ou positif respectivement. De quelle distribution ce résultat est-il la représentation ?

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{i}{x + i\epsilon} \varphi(x) dx &= i \int_{-\infty}^{\infty} \frac{x \varphi(x)}{x^2 + \epsilon^2} dx + \epsilon \int_{-\infty}^{\infty} \frac{\varphi(x)}{x^2 + \epsilon^2} dx \\ &= i \int_0^{\infty} \frac{x [\varphi(x) - \varphi(-x)]}{x^2 + \epsilon^2} dx + \epsilon \int_0^{\infty} \frac{\varphi(x) + \varphi(-x)}{x^2 + \epsilon^2} dx. \end{aligned}$$

Tant que $\epsilon > 0$ les intégrales sont bien définies⁶. Quand $\epsilon \rightarrow 0^+$ la première donne la distribution valeur principale, quand à la seconde

$$\epsilon \int_0^{\infty} \frac{\varphi(x) + \varphi(-x)}{x^2 + \epsilon^2} dx = \int_0^{\infty} \frac{\varphi(\epsilon x) + \varphi(-\epsilon x)}{x^2 + 1} dx \xrightarrow{(\epsilon \rightarrow 0^+)} 2\varphi(0) [\arctan x]_0^{\infty},$$

soit $\pi\varphi(0)$ que nous réécrivons $\pi\langle\delta; \varphi\rangle$. En regroupant les deux intégrales

$$\lim_{\epsilon \rightarrow 0^+} \frac{i}{x + i\epsilon} = i\text{vp}(1/x) - \pi\delta(x) \quad (\text{formule de Sochocki et Plemelj}),$$

ce qui démontre la transformée de Fourier annoncée.

⁴Les logiciels de calcul formels comme Mathematica ne sont alors à déconseiller.

⁵C'est-à-dire une suite de distributions convergeant, au sens des distributions, vers $\Theta(k)$.

⁶Nous intégrons sur \mathbb{R}^+ et les pôles ($\pm i\epsilon$) sont complexes.

Annexe E

Modèle du lignes & cercles : une alternative au point col

Cette annexe propose une alternative à l'approximation de point col du WLC, précédemment développée en présence (§ 5.2) ou non (§ 2.2) d'une cassure à mi-longueur. La caractérisation de la configuration de point col ne pose pas de problème analytique majeur, mais elle fait intervenir des fonctions spéciales *a priori* peu familières : les intégrales elliptiques de premier et deuxième types. Nous proposons ici un modèle géométrique simple dont la caractérisation ne nécessite que les fonctions trigonométriques classiques, réalisant qui plus est une bonne « première approximation » de la configuration de point col.

Nous avons constaté en section 2.2 que pour le WLC, la boucle ($r = 0$) d'énergie minimale n'était pas un cercle. Le point col possède en fait une énergie élastique de $\frac{\varepsilon_*(0)}{L/A} \simeq \frac{14.055}{L/A} k_B T$, inférieure d'environ 28.8% à celle du cercle. Dans ce cas l'énergie a pourtant une expression analytique simple

$$\beta E[\vec{t}_\circ; L] = \frac{2\pi^2}{L/A} \equiv \frac{A}{2} \frac{2\pi}{L/(2\pi)},$$

où l'on reconnaît le rayon (de courbure) $R_c = L/(2\pi)$ du cercle. Est-il possible de modifier la configuration de cercle de manière à mieux approcher le point col, tout en conservant son « caractère trigonométrique » sympathique ? En 2003, Kulić et Schiessel [KS03] proposèrent de ne considérer qu'un arc de cercle au lieu d'un cercle entier, la fermeture de la boucle étant alors assurée au moyen de deux segments linéaires. Ces derniers ne possèdent pas d'énergie élastique tandis que celle de l'arc de cercle peut être rendue inférieure à celle du cercle entier. C'est pourquoi nous pouvons espérer diminuer significativement l'énergie de la boucle par ce procédé. Nous proposons ici sa généralisation au cas d'un WLC *a priori* ouvert et présentant une cassure à mi-longueur¹. Le lecteur n'ayant lu que le chapitre 4 peut suivre cette annexe en prenant $a/2 = \pi$, l'interprétation de a

¹Le cas d'une cassure seule (polymère fermé $r = 0$) a été indépendamment traité par Sankararaman et Marko en 2005 [SM05b].

lui étant fournie au chapitre 5.

La configuration est planaire et symétrique : elle est en fait repérée et ca-

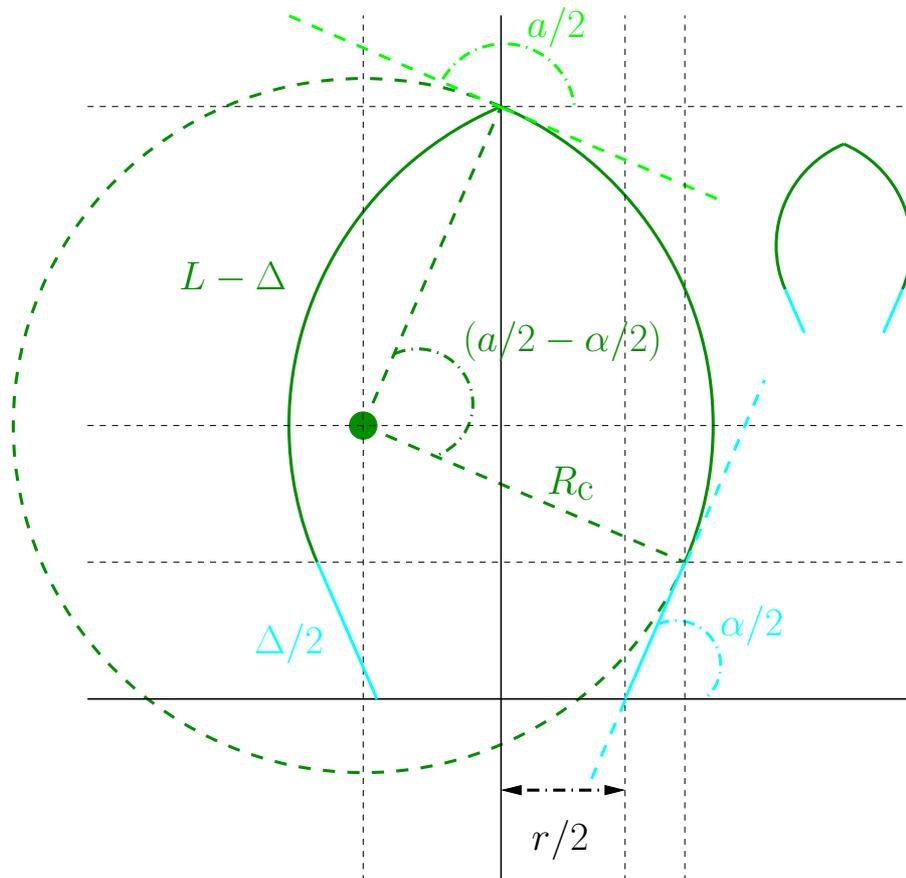


FIG. E.1 – Modèle géométrique simple permettant d’approcher la configuration de point col d’un WLC de longueur L , d’extension r et présentant une cassure a à mi-longueur. Comme pour le point col, le polymère est planaire et symétrique. Chacune des deux moitiés est constituée d’un arc de cercle (en vert, lignes pleines), prolongé d’un segment linéaire (en gris, lignes pleines) donc d’énergie élastique nulle. L’orientation initiale α_* minimisant l’énergie élastique du polymère est solution d’une équation transcendante ne faisant intervenir que les fonctions trigonométriques classiques.

ractérisée comme celle du point col (figures 2.2 et E.1). Les deux segments linéaires (en gris, lignes pleines) emportent une longueur Δ de la longueur totale L du polymère. Chacun des deux arcs de cercle (en vert, lignes pleines) possède alors une longueur $(L - \Delta)/2$, correspondant à un rayon de courbure tel que $L - \Delta = R_c(a - \alpha)$. Nous pouvons éliminer la variable Δ en constatant que $r + \Delta \cos(\alpha/2) = 2R_c [\sin(\alpha/2) - \sin(a/2)]$, d’où

$$\frac{1}{R_c} = \frac{2}{L} \frac{(a/2 - \alpha/2) \cos(\alpha/2) + \sin(\alpha/2) - \sin(a/2)}{\cos(\alpha/2) + r/L}.$$

L'énergie élastique d'un polymère de longueur L , d'extension r et présentant une cassure a à mi-longueur, se calcule alors en fonction de son orientation initiale α

$$\beta\mathcal{E}(r/L; \alpha, a) = 2 \times \frac{A a/2 - \alpha/2}{2 R_c}.$$

Comme pour le point col, nous sommes intéressés par l'orientation initiale α_* qui minimise cette énergie. L'annulation de sa dérivée première fournit l'équation (inconnue $x = \alpha_*/2$, paramètres $y = a/2$ et r/L fixés)

$$\begin{aligned} & \left\{ (y-x) \left[(y-x) \sin x + \cos x \right] + \sin x - \sin y \right\} \times r/L \\ & + \sin y \left[(y-x) \sin x - \cos x \right] + (y-x) \cos(2x) + \frac{1}{2} \sin(2x) = 0. \end{aligned}$$

Comme voulu seules les fonctions trigonométriques classiques sont impliquées, mais l'équation est transcendante et doit donc être résolue graphiquement ou numériquement. Cela demande toutefois peu d'effort : en guise d'illustration, pour la boucle ($r = 0$) sans cassure ($a/2 = \pi$) nous devons résoudre

$$\tan(2\pi - \alpha_*) = 2\pi - \alpha_*, \quad \text{de solution} \quad \alpha_*/2 \simeq 51.3^\circ.$$

L'énergie d'environ $\frac{15.7}{L/A} k_B T$ de ce polymère est par définition supérieure à celle du point col. Toutefois cette dernière ne lui est inférieure que de 10.5%, contre 28.8% pour le cercle.

L'énergie aux extensions $r \geq 0$ est montrée en figure 2.6, pour un polymère sans cassure à mi-longueur. Elle approche en effet mieux celle du point col que celle d'un simple arc de cercle (pas de segments linéaires ci-dessus, $\Delta = 0$). L'introduction de la cassure diminue significativement cette énergie, de manière très semblable à l'effet qu'elle produit sur le point col (figure 5.3).

Annexe F

Transformée de Fourier discrète 1D ou 2D

F.1 Dimension un : repliement, FFT

Nous considérons une fonction $f(x)$ définie¹ sur la droite $x \in \mathbb{R} =]-\infty; \infty[$. Nous définissons sa transformée de Fourier comme la fonction $g(k)$, également définie sur \mathbb{R} , obtenue par

$$g(k) = \int_{\mathbb{R}} dx \exp(-i k \times x) f(x).$$

Réciproquement nous retrouvons $f(x)$ par transformée inverse

$$f(x) = \int_{\mathbb{R}} \frac{dk}{2\pi} \exp(i k \times x) g(k).$$

Supposons $g(k)$ échantillonnée à pas constant Δk . Nous précisons plus bas un choix de Δk pertinent : nous exigeons pour le moment qu'il fournisse une définition convenable de $g(k)$. Le calcul de $f(x)$ par discrétisation de l'intégrale en k se fait donc à poids constant Δk (une somme de Riemann)²

$$f(x) = \frac{\Delta k}{2\pi} \sum_{n=-\infty}^{\infty} \exp(i x k_n) g(k_n).$$

En pratique nous devons tronquer cette somme. Nous supposons qu'en accord avec l'échantillonnage nous pouvons nous limiter à $n = 0, 1, \dots, (N-1)$ c'est-à-dire $k \in [0; k_{N-1}]$. Réciproquement, pour calculer $g(k_n)$ nous choisissons l'échantillonnage à pas constant $\Delta x = 2x_c/N$, sur $x \in]-x_c; x_c]$ avec $x_c = x_{N/2}$.³

¹C'est-à-dire pour nous : *bien* définie, possédant toutes les propriétés requises.

²La substitution de ce $f(x)$ dans un $g(k_m)$ fait apparaître la relation d'orthogonalité (au sens des distributions) des ondes planes $\int_{\mathbb{R}} \frac{dx}{2\pi} \exp[i(k_m - k_n)x] = \delta(k_m - k_n) = \delta_{m,n}/\Delta k$.

³Nous pouvons toujours choisir N pair afin que $N/2$ reste entier.

La réciprocity de la transformée de Fourier discrète implique une discrétisation à poids également constant

$$\begin{aligned} g(k_n) &= \sum_{\alpha=-N/2+1}^{N/2} p_\alpha \exp(-i x_\alpha k_n) f(x_\alpha) \\ &= \frac{\Delta k}{2\pi} \sum_{m=0}^{N-1} g(k_m) \sum_{\alpha=-N/2+1}^{N/2} p_\alpha \exp[i \alpha \Delta x (m-n) \Delta k] \end{aligned}$$

La somme sur α est une relation d'orthogonalité⁴ si $\Delta x \Delta k = 2\pi/N$, c'est-à-dire $\Delta k = \pi/x_c$ et $p_\alpha = \Delta x = 2x_c/N$ constant. Récapitulons : en pratique nous fixons la coupure x_c et le nombre de points N

$$\begin{aligned} g(n\pi/x_c) &= \frac{2x_c}{N} \sum_{\alpha=-N/2+1}^{N/2} \exp(-i 2\pi n \alpha/N) f(\alpha 2x_c/N) \\ f(\alpha 2x_c/N) &= \frac{1}{2x_c} \sum_{n=0}^{N-1} \exp(i 2\pi \alpha n/N) g(n\pi/x_c). \end{aligned}$$

La littérature consacrée à la transformée de Fourier discrète présente une écriture le plus souvent différente. Notons $G_n = g(n\pi/x_c)$, et F_α la somme discrète définissant $f(\alpha 2x_c/N)$, c'est-à-dire en omettant le préfacteur $1/(2x_c)$. Sous la discrétisation, le facteur exponentiel de la transformée est devenu N -périodique en α . La coupure x_c , aussi appelée *fréquence critique* ou *de Nyquist* (voir plus bas), de $f(x)$ nous permet de rendre sa discrétisation F_α de même période $F_{\alpha \pm N} = F_\alpha$, en laissant l'indice α libre. Avec cette convention, nous réécrivons la somme sur α telle que

$$F_\alpha = \sum_{n=0}^{N-1} \exp(i 2\pi \alpha n/N) G_n \quad \text{et} \quad G_n = \frac{1}{N} \sum_{\alpha=0}^{N-1} \exp(-i 2\pi n \alpha/N) F_\alpha.$$

Revenons sur les hypothèses de discrétisation. Nous avons notamment supposé $f(x)$ nulle au-delà de la fréquence critique x_c . Il est courant de dire que $g(k)$ est *limitée en largeur de bande* (à x_c), ou « bandwidth limited » en anglais. Si c'est bien le cas, alors le *théorème d'échantillonnage* ou *de Nyquist/Shannon* stipule que l'échantillonnage de $g(k)$ à pas constant $\Delta k = \pi/x_c$ suffit à reconstruire cette fonction en tout point. Autrement dit il s'agit de l'échantillonnage minimal, contenant toute l'information sur la fonction. Un pas plus petit (c.-à-d. une fréquence critique plus grande) ne constitue qu'un surplus *a priori* inutile. Plus précisément

$$g(k) = \sum_{n=-\infty}^{\infty} g(n\pi/x_c) \operatorname{sinc}[x_c(k - n\pi/x_c)].$$

⁴En effet $\sum_{\alpha=-N/2+1}^{N/2} \exp[i 2\pi \alpha (m-n)/N] = N \delta_{m,n}$.

En pratique l'échantillonnage n'est pas infini. Nous avons également supposé $g(k)$ nulle en dehors de l'intervalle $[0; (N - 1)\pi/x_c]$. Généralement la coupure n'est pas exacte, mais il est raisonnable de voir $g(k)$ comme une fonction s'annulant (plus ou moins vite). Le nombre de points N doit alors permettre de couvrir une fenêtre représentative du signal non nul.

En revanche, l'utilisation d'un pas plus grand correspond à une coupure x_c plus petite que la fréquence critique. Cette fréquence peut par ailleurs être infinie. La remise en cause de la limitation en bande de $g(k)$ conduit au *repliement* (« aliasing » en anglais) pour $f(x)$, des parties $|x| > x_c$ sur la bande $|x| \leq x_c$ (figure F.1). C'est une conséquence directe de la discrétisation de $g(k)$. Ce repliement correspond plus rigoureusement à l'introduction de la N -périodicité, c'est-à-dire de période $2x_c$, du facteur exponentiel de la transformée. Nous calculons en effet

$$F_\alpha = \sum_{n=0}^{N-1} \exp(i 2\pi \alpha n/N) g(n \pi/x_c).$$

En pratique, c'est $g(k)$ qui est échantillonnée et $f(x)$ qui est calculé. Examinons toutefois la réciproque : rigoureusement on doit calculer

$$\begin{aligned} g(n \pi/x_c) &= \frac{2x_c}{N} \sum_{\beta=-\infty}^{\infty} \exp(-i 2\pi n \beta/N) f(\beta 2x_c/N) \\ \sum_{\beta=-\infty}^{\infty} &= \sum_{\beta=-\infty}^{-N/2} + \sum_{\beta=-N/2+1}^{N/2} + \sum_{\beta=N/2+1}^{\infty} . \end{aligned}$$

Nous ne pouvons plus maintenant nous limiter à la somme centrale de N termes. Les deux séries peuvent toutefois être morcelées en sommes de N termes. Par exemple, celle suivant immédiatement la somme centrale est

$$\sum_{\beta=N/2+1}^{3N/2} \exp(-i 2\pi n \beta/N) f(\beta 2x_c/N) = \sum_{\gamma=-N/2+1}^{N/2} \exp(-i 2\pi n \gamma/N) f(\gamma 2x_c/N + 2x_c),$$

et peut donc y être incluse en posant $\gamma = \beta - N$. Le repliement des autres sommes se vérifie en utilisant, comme annoncé, la N -périodicité du facteur exponentiel de la transformée. Ainsi

$$g(n \pi/x_c) = \frac{2x_c}{N} \sum_{\beta=-N/2+1}^{N/2} \exp(-i 2\pi n \beta/N) \sum_{a=-\infty}^{\infty} f(\beta 2x_c/N + a \times 2x_c).$$

La substitution de cette expression dans la discrétisation F_n fait apparaître une relation d'orthogonalité (voir note de bas de page 127) ne laissant que le terme $\beta = \alpha$ de la somme. Par conséquent

$$\frac{1}{2x_c} F_\alpha = f(\alpha 2x_c/N) + f(\alpha 2x_c/N + 2x_c) + f(\alpha 2x_c/N - 2x_c) + \dots$$

Si $f(x)$ est nulle au-delà de x_c , F_α en est une estimation correcte. Sinon, la partie non nulle de $f(x)$ se retrouve accumulée dans F_α (figure F.1). Si ce repliement ne peut pas être supprimé en échantillonnant de nouveau $g(k)$, à un pas en accord avec la fréquence critique, sa minimisation n'a rien d'évident. Un test de base consiste à tester la réciprocity de la transformée de Fourier discrète. Nous pouvons également vérifier si $f(x)$ s'annule en x_c . Autrement nous pouvons

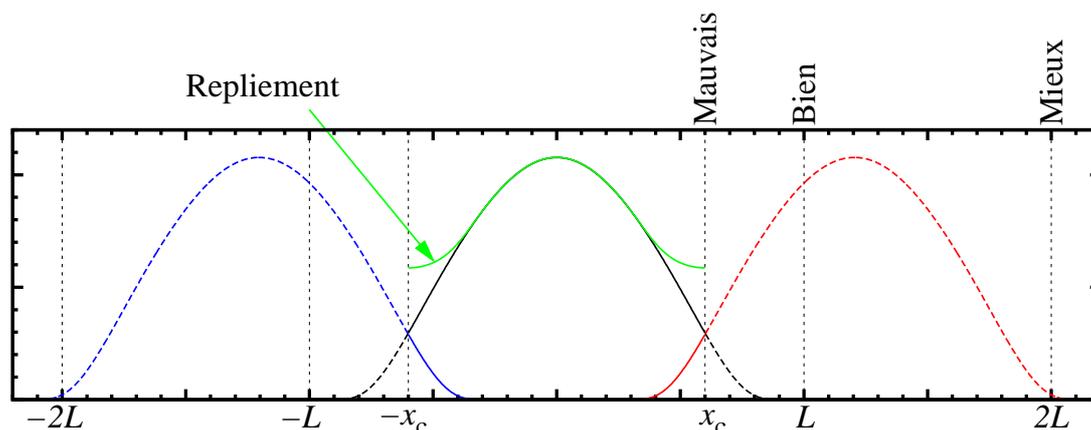


FIG. F.1 – Transformée de Fourier discrète d'une fonction limitée en bande à L . Son échantillonnage se fait à pas constant π/x_c . Si $x_c = L$, la transformée est correcte (ligne et pointillés noirs). En revanche, si $x_c < L$ la discrétisation introduit une période $2x_c$ (lignes et pointillés rouge et bleu), responsable du repliement des parties $|x| > x_c$ du résultat exact (pointillés noirs) sur la bande $|x| \leq x_c$. La transformée est alors fautive (ligne verte). En pratique il est d'usage de doubler la fréquence critique afin d'écartier tout phénomène de repliement.

imposer une fréquence critique artificielle, par exemple au moyen d'une fonction porte ($= 1/(2x_c)$ si $|x| \leq x_c$, nulle ailleurs : § D.2) dont la transformée de Fourier est un sinus cardinal. Nous nous ramenons alors à un problème de convolution, c'est-à-dire de filtrage, typique du traitement du signal.

Une transformée de Fourier discrète est la multiplication du vecteur G_n , par la matrice (symétrique) des puissances entières du nombre complexe $\exp(i2\pi/N)$. Nous pouvons faire beaucoup plus rapide que cet algorithme naïf, dont le temps de calcul varie comme $O(N^2)$. Pour les algorithmes de FFT, de l'anglais « Fast Fourier Transform », ce coût varie en $O(N \log_2 N)$. L'utilisation massive de la FFT depuis les années 1960 s'est accompagnée d'une grande quantité de bibliothèques, et de documents aussi bien pédagogiques que spécialisés. Le principe d'une FFT [PTVF02] consiste à réécrire la transformée de Fourier discrète sur N points en une somme de deux transformées de Fourier sur $N/2$ points. En itérant, on peut donc réduire ce calcul à des transformées élémentaires⁵. En établissant une correspondance entre chaque G_n , et leur transformée de Fou-

⁵C'est-à-dire des identités. La réduction peut aussi être stoppée aux transformées de Fourier discrètes sur 4 ou 8 points, pour lesquelles le facteur exponentiel prend des valeurs simples.

rier élémentaire à la fin du processus, le temps de calcul est alors essentiellement consacré au réarrangement des données. Notons que pour permettre cette réduction, le nombre de points N doit être une puissance de deux⁶. Remarquons pour finir que si $f(x)$ est réelle et paire, alors $g(k)$ l'est aussi. Au niveau discret cela nous permet, moyennant un arrangement particulier des données, de stocker le vecteur réel G_n de taille N dans un vecteur complexe de taille $N/2$.

F.2 Dimension deux : fonction radiale

Nous considérons une fonction $F(x_1, x_2)$, définie⁷ sur le plan $(x_1, x_2) \in \mathbb{R}^2 = [-\infty; \infty[\times [-\infty; \infty[$. La fonction est radiale si elle ne dépend que de $x = \sqrt{x_1^2 + x_2^2}$, et pas de $\phi : \tan \phi = x_1/x_2$. Alors $F(x_1, x_2) \equiv f(x)$ est définie sur la demi droite positive $x \in \mathbb{R}^+ = [0; \infty[$. Sa transformée de Fourier $G(k_1, k_2)$ est également définie sur \mathbb{R}^2 et partage la même symétrie de révolution. C'est-à-dire $G(k_1, k_2) \equiv g(k)$, avec $k = \sqrt{k_1^2 + k_2^2}$, est indépendante de $\theta : \tan \theta = k_1/k_2$. Par définition F et G se transforment selon

$$\begin{aligned} G(k_1, k_2) &= \int_{\mathbb{R}^2} dx_1 \times dx_2 \exp[-i(k_1 \times x_1 + k_2 \times x_2)] F(x_1, x_2) \\ F(x_1, x_2) &= \int_{\mathbb{R}^2} \frac{dk_1 \times dk_2}{(2\pi)^2} \exp[i(x_1 \times k_1 + x_2 \times k_2)] G(k_1, k_2). \end{aligned}$$

Alors $f(x) = F(x, 0)$ et $g(k) = G(k, 0)$ sont aussi reliés, suivant

$$\begin{aligned} g(k) &= \int_{\mathbb{R}^2} dx \times x d\phi \exp(-i k \times x \cos \phi) f(x) = 2\pi \int_{\mathbb{R}^+} dx x J_0(k x) f(x) \\ f(x) &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} dk \times k d\theta \exp(i x \times k \cos \theta) g(k) = \frac{1}{2\pi} \int_{\mathbb{R}^+} dk k J_0(x k) g(k) \end{aligned}$$

avec $J_0(u) = \int_0^{2\pi} e^{\pm iu \cos \alpha} \frac{d\alpha}{2\pi}$ la fonction (réelle et paire) de Bessel du premier type, d'ordre zéro. On parle de transformée : de Fourier radiale en dimension 2, de Hankel ou encore de Fourier/Bessel. Dans la fin de cette annexe nous exposons une discrétisation possible de cette transformation, due à Lado (1971) [Lad71], dans le but d'un calcul numérique.

Considérons dans un premier temps le calcul de $f(x)$. Pour cela, nous la supposons en fait nulle en dehors du cercle de rayon $R \geq 0$. Donc

$$g(k) = 2\pi \int_0^R dx x J_0(k x) f(x) \quad \text{et } \forall k \in \mathbb{R}^+, J_0(x k) = 0 \text{ si } x \geq R.$$

Nous voulons discrétiser l'intégrale sur k afin de calculer $f(x)$. Pour cela nous proposons un échantillonnage $\{k_{i \geq 1}\}$ utilisant les zéros $\{\mu_{i \geq 1}\}$ de la fonction de Bessel. Compte tenu de la coupure R , l'échantillonnage minimal est obtenu

⁶Le plus simple pour relaxer cette contrainte est de compléter G_n avec des 0.

⁷C'est-à-dire *bien* définie, possédant toutes les propriétés requises.

pour $J_0(R k_i) = 0$, soit $k_i = \mu_i/R$. Reste alors à déterminer les poids w_i associés à ce choix, permettant d'écrire

$$f(x) = \frac{1}{2\pi} \sum_{i=1}^{\infty} w_i k_i J_0(x k_i) g(k_i).$$

La substitution de $f(x)$ par cette somme dans l'expression d'un $g(k_j)$ quelconque ci-dessus, fait apparaître la relation d'orthogonalité des fonctions de Bessel

$$\int_0^R x J_0(x k_i) J_0(k_j x) dx = R^2 \int_0^1 r J_0(r \mu_i) J_0(\mu_j r) dr = \frac{R^2}{2} [J'_0(\mu_i)]^2 \delta_{i,j}.$$

Dans l'expression de $g(k_j)$, ne reste alors de la somme que le terme $i = j$, ce qui permet de trouver

$$w_j k_j = \frac{2/R^2}{[J'_0(\mu_j)]^2}.$$

Le problème semble résolu, mais en pratique nous devons tronquer la somme définissant $f(x)$. Cela revient à imposer une coupure également sur k , qui en accord avec l'échantillonnage choisit correspond à $\Lambda = k_N = \mu_N/R$. Réciproquement, pour calculer $g(k_i)$ l'échantillonnage minimal suivant les μ_α est alors $x_\alpha = \mu_\alpha/\Lambda$, auquel cas $\mu_N = \Lambda R$. Nous devons maintenant trouver les poids p_α associés à

$$\begin{aligned} g(k_i) &= 2\pi \sum_{\alpha=1}^N p_\alpha x_\alpha J_0(k_i x_\alpha) f(x_\alpha) \\ &= \sum_{j=1}^N w_j k_j g(k_j) \sum_{\alpha=1}^N p_\alpha x_\alpha J_0(k_i x_\alpha) J_0(x_\alpha k_j). \end{aligned}$$

Nous aimerions voir en la somme sur α une relation d'orthogonalité. Le choix

$$p_\alpha x_\alpha = \frac{2/\Lambda^2}{[J'_0(\mu_\alpha)]^2}$$

similaire au résultat pour w_j produit alors

$$g(k_i) = \sum_{j=1}^N g(k_j) \underbrace{\frac{4/\mu_N^2}{[J'_0(\mu_j)]^2} \sum_{\alpha=1}^N \frac{J_0(k_i x_\alpha) J_0(x_\alpha k_j)}{[J'_0(\mu_\alpha)]^2}}_{\text{orthogonalité approchée } \simeq \delta_{i,j}},$$

et la réciprocité de la transformée de Hankel discrète n'est donc qu'approchée⁸.

Récapitulons : en pratique nous fixons la coupure Λ et le nombre de points N

$$\begin{aligned} g(\Lambda \mu_i/\mu_N) &= \frac{4\pi}{\Lambda^2} \sum_{\alpha=1}^N \frac{J_0(\mu_i \mu_\alpha/\mu_N)}{[J'_0(\mu_\alpha)]^2} f(\mu_\alpha/\Lambda) \\ f(\mu_\alpha/\Lambda) &= \frac{\Lambda^2}{\pi \mu_N^2} \sum_{i=1}^N \frac{J_0(\mu_\alpha \mu_i/\mu_N)}{[J'_0(\mu_i)]^2} g(\Lambda \mu_i/\mu_N). \end{aligned}$$

⁸Numériquement, pour $N \simeq 20$ l'erreur commise est comparable à la précision sur les fonctions de Bessel. L'algorithme de Lado est en général utilisé pour $N \gtrsim 100$.

La matrice symétrique $J_0(\mu_i \mu_\alpha / \mu_N) / [J_0'(\mu_\alpha)]^2$ et les zéros μ_i de la fonction de Bessel ne sont calculés qu'une seule fois pour un N arbitrairement grand.⁹ Notons qu'à nombre de points N fixé, l'introduction d'une coupure Λ de plus en plus grande dans la « fonction source » $g(k)$ implique l'existence d'une coupure μ_N / Λ de plus en plus petite dans la « fonction résultat » $f(x)$. L'illustration typique est la transformée de Hankel du cercle (fonction porte, § D.2), donnant une fonction de Bessel modulée par la fonction inverse (analogue du sinus cardinal obtenu en 1D)

$$\frac{1}{2\pi} \int_0^\Lambda dk \, k \, J_0(xk) \frac{1}{\Lambda} = \frac{\Lambda}{2\pi} \frac{J_1(\Lambda x)}{\Lambda x}.$$

Elle tend vers zéro quand $\Lambda \rightarrow \infty$ pour x fini, non nul (elle diverge pour $x = 0$); idem quand $\Lambda \rightarrow 0$ pour x fini.¹⁰ Quand la coupure en x est une approximation importante, jauger l'erreur commise par l'algorithme n'est alors pas trivial : un test de base (§ F.1) est la vérification de la réciprocity de la transformée de Hankel discrète calculée.

Remarquons pour finir que l'échantillonnage obtenu n'est pas constant, contrairement à la transformée de Fourier discrète exposée en section F.1. L'algorithme de Lado est en fait plus proche des quadratures gaussiennes souvent utilisées lors d'intégrations numériques. Enfin, le lecteur intéressé pourra consulter la référence originelle [Lad71] qui contient un algorithme similaire pour la transformée de Fourier radiale en dimension 3, comme celle rencontrée en annexe D.1.

⁹Utiliser $J_0'(u) = -J_1(u)$, omis dans le texte afin d'éviter toute confusion entre indices.

¹⁰Même comportement pour le cercle non normalisé (= 1 sur $[0; \Lambda]$, nul ailleurs).

Annexe G

Publications

Ce travail a donné lieu à un article :

- N. Douarche and S. Cocco, *Protein-Mediated DNA Loops : Effects of Protein Bridge Size and Kinks*, Physical Review E **72** (2005), 061902-1 ;

dont le contenu a préalablement été déposé sur le serveur de prépublications (cond-mat/0508288). Nous avons également participé à une revue :

- J.-F. Allemand, S. Cocco, N. Douarche, and G. Lia, *Loops in DNA : an Overview of Experimental and Theoretical Approaches*, submitted to European Physical Journal E (2005).

Nous reproduisons ici ces deux références, prépublication incluse.

Protein-mediated DNA loops: Effects of protein bridge size and kinks

Nicolas Douarache and Simona Cocco

CNRS-Laboratoire de Physique Statistique de l'ENS, 24 rue Lhomond, 75005 Paris, France

(Received 11 August 2005; published 2 December 2005)

This paper focuses on the probability that a portion of DNA closes on itself through thermal fluctuations. We investigate the dependence of this probability upon the size r of a protein bridge and/or the presence of a kink at half DNA length. The DNA is modeled by the wormlike chain model, and the probability of loop formation is calculated in two ways: exact numerical evaluation of the constrained path integral and the extension of the Shimada and Yamakawa saddle point approximation. For example, we find that the looping free energy of a 100-base-pairs DNA decreases from $24 k_B T$ to $13 k_B T$ when the loop is closed by a protein of $r=10$ nm length. It further decreases to $5 k_B T$ when the loop has a kink of 120° at half-length.

DOI: [10.1103/PhysRevE.72.061902](https://doi.org/10.1103/PhysRevE.72.061902)

PACS number(s): 87.15.-v, 82.35.-x, 46.70.-p, 36.20.-r

I. INTRODUCTION: DNA LOOPS IN GENE TRANSCRIPTION REGULATION

Gene expression is regulated by a wide variety of mechanisms. These activities as well as repression phenomena may occur at all expression steps (translation, transcription, etc) and do involve interactions between several biological molecules (DNA, RNAs, proteins, etc). For instance, proteins bound on specific DNA sequences may turn on or off gene transcription by interacting with each other. By bringing those proteins closer, DNA looping can ease their interactions [1–4]. Such looping events may be observed over a wide range of lengths spreading from hundreds to thousands of base pairs (bp). The looping probability was measured by cycling DNA segments in solution with cohesive ends. Once the loop is formed, proteins called ligases stabilize it [5,6]. It is then possible to count the circular DNAs with respect to the linear ones. The loop formation mediated by proteins has also been experimentally studied. Two examples are the loops formed by the LacR or the GalR transcriptional repressors [7]. Two units of such proteins bind at two specific positions along the same DNA and associate to form a complex when the binding sites come in contact. The formation of such loops has been recently studied using micromanipulation experiments on a single DNA molecule [1,2]. The study of the GalR mediated loop has shed light on the role of a third protein called HU that sharply bends (i.e., kinks) the DNA at half-length.

The DNA loop probability depends mainly on its length and flexibility. Long DNAs (typically longer than 1500 bp) essentially behave as Gaussian polymers (GP) [8,9]: the cyclization cost is mainly of an entropic nature [10]. On the contrary, for small lengths DNA cyclization is difficult mainly because of the bending energy cost. The computation of the elastic energy for the wormlike chain (WLC) model [8,9,11] can be analytically performed [14]; numerical methods have also been employed when electrostatic properties are included [15]. At intermediate length scales (in the range 50–2500 bp) elastic rigidity and entropic loss are both important. Several approximations have been developed to study this length's range [3,12,13,16], among them the calculations of the fluctuations around the lowest bending energy configurations performed by Shimada and Yamakawa

[14]. Numerical approaches have also been developed: Monte Carlo [17] and Brownian dynamics-based simulations [4,18,19] as well as numerical calculations of the WLC path integral under the closed ends constraint [20,21]. This latter method allowed Yan, Kawamura, and Marko to study the elastic response of DNA subject to permanent or thermally excited bendings caused by binding proteins (such as HU) or inhomogeneities along the DNA double helix [21,22]. All these studies do provide a better understanding of the underlying regulation phenomena despite their overall complexity.

In this paper we study two processes that turned out to be important in DNA looping, namely, the size of the protein complex clamping the loop [4,18], acting as a bridge between the two DNA ends, and mechanisms implying DNA stiffness loss, which are taken into account in an effective way by kinking the WLC at half-contour length [4,18,19,23]. In Sec. II we define the model and the methods: we describe the numerical approach (Sec. II A) and the analytical saddle point approximation (SPA, Sec. II B) to calculate the r -dependent closure factor and the looping free energy. In Sec. III we compare the numerical and SPA results with previous experimental and theoretical results. In Sec. IV we extend the numerical and SPA approaches to a kinked loop; we discuss our results and we propose a simple formula that accounts for both the protein bridge and kink effects (Sec. IV C). We conclude (Sec. V) by sketching how to include omitted DNA properties, which may also play an important role in its closure such as twist rigidity or electrostatic interactions.

II. DEFINITIONS AND METHODS

We use the well-known wormlike chain (WLC) model [8,9,11]. The DNA polymer is described as an *inextensible* continuous differentiable curve of contour length L , with unit tangent vector $\vec{t}(s)$ ($0 \leq s \leq L$). The polymer is characterized by the persistence length A beyond which tangent vectors lose their alignment: $\langle \vec{t}(s) \cdot \vec{t}(s') \rangle = \exp(-|s-s'|/A)$. The energy of a configuration of the polymer stretched under an external force $f \vec{e}_z$ reads

$$E[\vec{t}; L, f] = \frac{1}{2} \frac{A}{\beta} \int_0^L \left[\frac{d\vec{t}(s)}{ds} \right]^2 ds - f \int_0^L \vec{e}_z \cdot \vec{t}(s) ds, \quad (1)$$

where we use $\beta = 1/k_B T$. No twist elasticity or extensibility will be considered. The partition function is

$$Z(L, f) = \int \mathcal{D}\vec{t} \exp\{-\beta E[\vec{t}; L, f]\}. \quad (2)$$

Notice that summation over all initial and final tangent vectors orientations, $\vec{t}(0)$ and $\vec{t}(L)$, is implicitly understood in this path integral.

In this paper we are interested in the formation of a loop in a DNA molecule, and the probability density functions (PDFs) of end-to-end distances play an important role. The quantities under study are denoted by Q , S , P , and J respectively and defined as follows: (i) the end-to-end extension $\vec{r} = (x, y, z)$ PDF at zero force,

$$Q(\vec{r}, L) = \frac{1}{Z(L, f=0)} \int \mathcal{D}\vec{t} \delta\left[\int_0^L \vec{t}(s) ds - \vec{r}\right] \times \exp\{-\beta E[\vec{t}; L, f=0]\}. \quad (3)$$

In the absence of force, Q depends on its argument \vec{r} through its modulus $r = |\vec{r}|$ only, and we may introduce the radial PDF

$$S(r, L) = 4\pi r^2 Q(r, L). \quad (4)$$

(ii) The z extension PDF reads

$$P(z, L) = \int_{-L}^L dx \int_{-L}^L dy Q((x, y, z), L). \quad (5)$$

In the absence of external force, notice the x or y extensions PDFs are given by P too. Interestingly, the radial and z extensions PDF are related to each other through the useful identity [24],

$$S(r, L) = -2r \frac{dP}{dz}(r, L). \quad (6)$$

(iii) The cyclization factor

$$J(L) = Q(0, L), \quad (7)$$

defined as the density of probability that the two ends of the DNA are in contact with one another. (iv) The r -dependent closure factor is

$$J(r, L) = \left[\int_0^r S(r', L) dr' \right] / \left(\frac{4}{3} \pi r^3 \right). \quad (8)$$

It gives the density probability for the two ends of the chain to stay within a sphere of radius r . It is easy to check that $J(r, L) \rightarrow J(L)$ when $r \rightarrow 0$. For experimental convenience, units used for $J(L)$ and $J(r, L)$ are moles per liter: $1 \text{ M} \equiv 1 \text{ mol/L} \approx 0,6 \text{ nm}^{-3}$. In these units $J(r, L)$ gives directly the concentration of one binding site in proximity of the other. (v) Finally we consider the looping free energy cost

$$\beta \Delta G(r, L) = -\ln \left[J(r, L) \times \frac{4}{3} \pi r^3 \right]. \quad (9)$$

Note that this definition does not include the details of the geometry or the affinities of the DNA-protein and protein-protein interactions. We actually assume all the sphere of radius r to be the reacting volume, i.e., that the loop will form if the DNA ends happen to be in this sphere.

Despite intense studies of the WLC model, no exact analytical expression is known for Q and the quantities of interest here, namely J and S . However, approximations expanding from the two limiting regimes (entropic [12,13] and elastic [14,16]) along with exact numerical computations are available. Hereafter, we have resorted to numerical as well as approximate analytical techniques (SPA) for calculating the cyclization factor J and the probability of almost closed DNA configurations.

A. Numerical calculation of the probabilities P , S , and J

Our starting point for the calculation of the z extension PDF is the Fourier representation of the Dirac δ -function in (3) and (5),

$$P(z, L) = \int_{-\infty}^{+\infty} \frac{dk}{2\pi} e^{+ikz} Z(L, f = -ik/\beta). \quad (10)$$

At fixed momentum k we are left with the calculation of the partition function $Z(L, f)$ at (imaginary) force $f = -ik/\beta$. The path integral (2) defining Z is interpreted as the evolution operator of a quantum system, the rigid rotator under an external imaginary field

$$Z(L, f) = \langle \text{final} | \exp[-L/A \times \hat{H}(f)] | \text{initial} \rangle. \quad (11)$$

The entries of its Hamiltonian \hat{H} are easily expressed in the spherical harmonics $|\ell, m\rangle$ basis: $\langle \ell, m | \hat{H}(f) | \ell', m' \rangle = H_{\ell, \ell'}(f) \delta_{m, m'}$ with

$$H_{\ell, \ell'}(f) = \frac{\ell(\ell+1)}{2} \delta_{\ell, \ell'} - \beta A f \frac{\ell \delta_{\ell', \ell-1} + \ell' \delta_{\ell, \ell'+1}}{\sqrt{(2\ell+1)(2\ell'+1)}}. \quad (12)$$

The entries of \hat{H} do not depend on the azimuthal number m due to cylindrical symmetry around the force axis. Finally, integration over all initial and final orientations for the tangent vectors at the ends of the polymer chain selects $|\text{initial}\rangle = |\text{final}\rangle = |0, 0\rangle$.

A recent paper [24] used MATHEMATICA to compute the vacuum amplitude (11) through a direct matrix exponentiation. We have instead used the EXPOKIT library [25] since it proves to be more accurate and faster for intensive numerical calculations. We truncated Hamiltonian (12) in such a way that the outcome is insensitive to the cutoff on the harmonics order. The (inverse) Fourier transform (10) is then handled by a fast Fourier transform (FFT) algorithm [26]. This task is in particular facilitated thanks to the inextensibility constraint which makes the distribution bandwidth limited. Our results for the z extension PDF are shown in Fig. 1. The crossover from the rigid elastic regime ($L/A = 0.1, 0.5, 1$) to

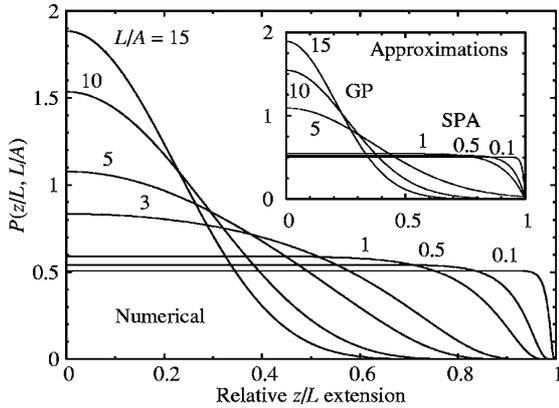


FIG. 1. Numerical computation of the z extension PDF over a wide range of contour lengths L . As expected, the agreement with the SPA prediction (see Sec. II B) improves as L decreases (upper bound $L/A \leq 1$). The long WLC behavior is caught by the GP approximation [8,9] as soon as, say, $L/A \geq 5$.

the flexible entropic regime ($L/A=5, 10, 15$) is clearly visible. Using (6) then gives us access to S , the radial extension PDF (see Fig. 2). The most probable value for the distance r_* switches from full extension $r_* \leq L$ for contour lengths $L \leq A$ (elastic dominated regime) to the GP most probable extension $\sqrt{4LA}/3$ for longer ones $L \gg A$ (entropic dominated regime). Also note that S always goes continuously to zero in the $r \rightarrow L$ limit due to WLC inextensibility. We have finally calculated the r -dependent closure factor J according to (8) by numerical integration of S , and the looping free energy cost ΔG defined in (9).

Let us now discuss the numerical errors that could be important when calculating probabilities of rare events. Main sources of error are the Hamiltonian (12) truncation and the integration step to compute the r dependent closure factor (8). As mentioned above, the cutoff on the harmonic order was systematically chosen to observe convergence. We have

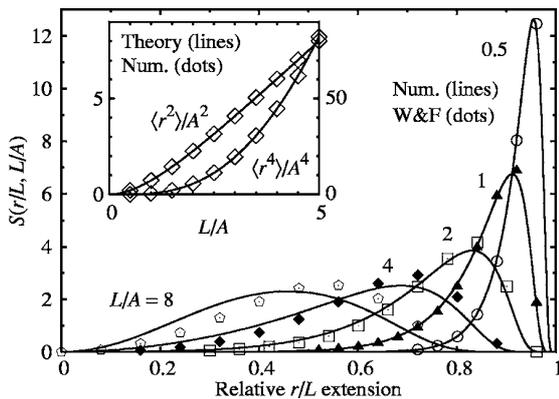


FIG. 2. Numerical computation of S , the radial r extension PDF. The outcomes of the numerical calculations are tested against exactly known values for the first even moments $\langle r^{2n} \rangle$ (inset) [8,9]. The shape of S compares very well to the widely used Wilhelm and Frey (W&F) expansion [16], valid up to $L/A \leq 2$. Similar tests were achieved with other popular approximation schemes [3] (data not shown).

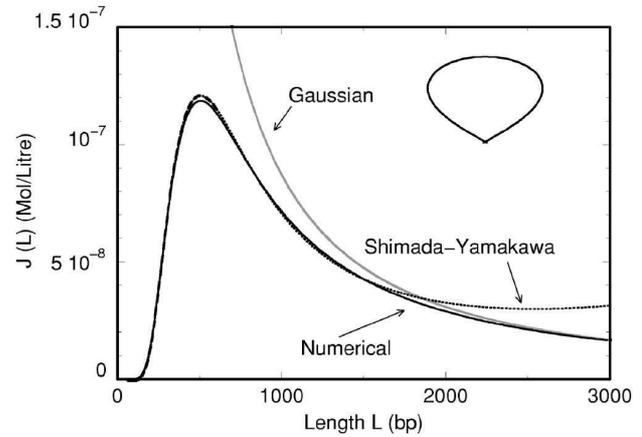


FIG. 3. Cyclization factor as a function of the DNA length with the Gaussian model (gray line); the WLC model with the Shimada and Yamakawa formula (dotted black line); the WLC model with the numerical calculation (full black line). The most probable length is 500 bp. Inset: the lowest bending energy configuration of a closed loop.

used $\ell=50$ after having verified that the result is unchanged for $\ell=100$. Concerning integration, limitation comes from the number of available data in the range $0 \leq r' \leq r$ which is directly related to the k sampling of the partition function Z . For $r=1$ nm the numerical integration is still reliable, but decreasing further r turns out to be critical. Other potential sources of error are negligible. Indeed, bandwidth limitation of $P(z;L)$ prevents any aliasing [26] during FFT (10) and the derivative of P (6) can actually be skipped by an integration by part of (8) to compute J . Further hypothetical errors would then come from the EXPOKIT library [25] itself but its routines were coded to accurately compute matrix exponentials over a broad range of matrices [27]. We have checked the numerical precision of our method by the comparison with the exact values for the first even moments of $S(r)$ (Fig. 2 inset); moreover, as shown in Fig. 2, $S(r)$ agrees with the Wilhelm and Frey expansion for small L/A values; finally, we will see in Sec. III that the numerically calculated cyclization factor $J(L)$ (8) is in agreement with the Shimada-Yamakawa and Gaussian approximation results for respectively small and large L (Fig. 3).

B. Saddle point approximation for J

In addition to the exact numerical calculations detailed above, we have carried out approximate calculations based on a saddle point estimate of the partition function (2). We follow the Shimada and Yamakawa calculation for the saddle point configuration [14], extending it for an opened DNA. The saddle point configuration for a closed loop is shown in Fig. 3 inset: it is a *planar loop*. The tangent vector at position s along the chain is characterized by its angle $\theta(s)$ with respect to the end-to-end extension \vec{r} . The optimal configuration is symmetric with respect to the perpendicular \vec{e}_\perp -axis while the half-length angle is $\theta(L/2)=180^\circ$. The initial angle $\theta(0)$ is chosen by the minimization of the bending energy of

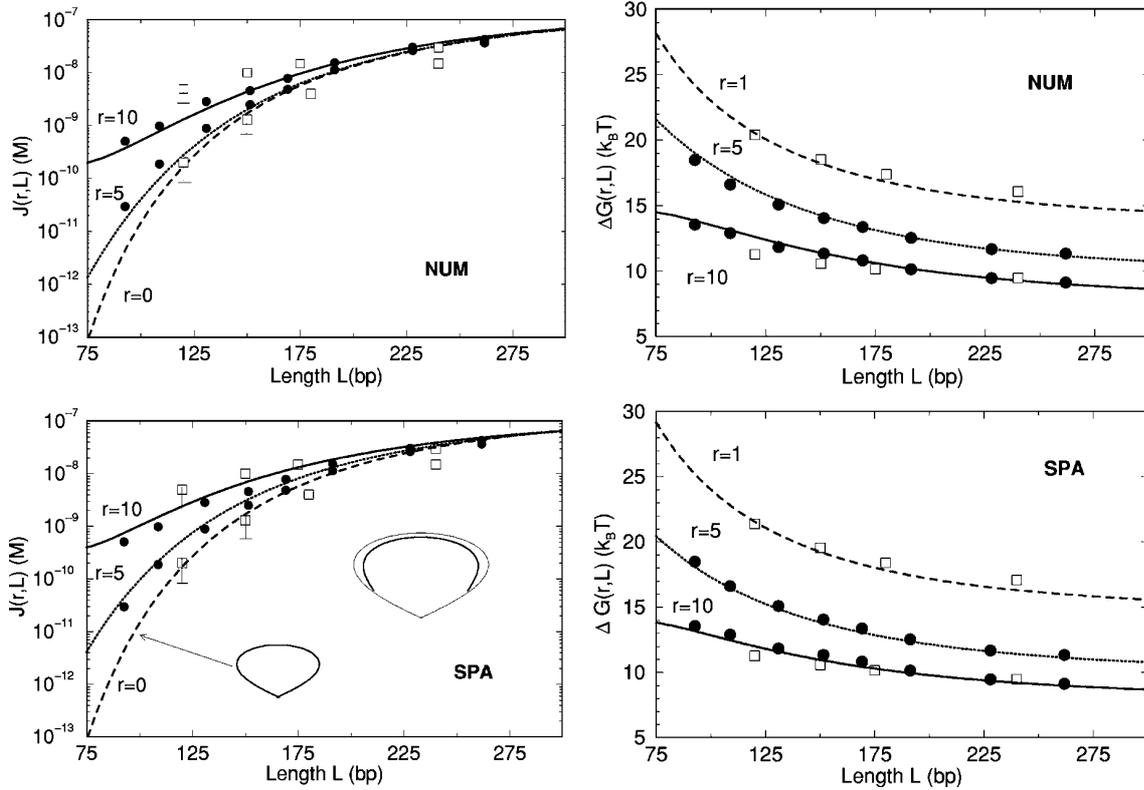


FIG. 4. Closure factor (left panel) and free energy (right panel) with a protein bridge of sizes: $r=1$ nm (dashed line); 5 nm (dotted line); 10 nm (full line). The error bars are shown when they are larger than the symbol sizes. Top: numerical calculation of the constrained path integral. The $r=1$ nm curve coincides with the $r=0$ cyclization factor. Bottom: closure factor obtained by the extension of the Shimada and Yamakawa calculation, which includes r . Theoretical results are in very good agreement with Monte Carlo simulations obtained by Podtelezhnikov and Vologodskii (filled circle, ●) [17], and in fair agreement with Brownian dynamics simulations obtained by Langowski *et al.* (empty square, □) [4,18]. Inset of the bottom left panel: lowest bending energy configurations for 100 bp and $r=0$ (bottom) or $r=10$ nm (top), the closure of the $r=10$ nm configuration is shown by a thin line.

the chain. The optimization gives rise to the following condition on the parameter $x = \cos[\theta(0)/4]$

$$(1 + r/L)\hat{K}(x^2) = 2\hat{E}(x^2), \quad (13)$$

where $\hat{K}(x^2) = K(\pi/2, x^2)$ and $\hat{E}(x^2) = E(\pi/2, x^2)$ are the complete elliptic integrals of the first and second kinds respectively [28]. The corresponding elastic energy is

$$\beta\Delta E(r, L) = \frac{4}{L/A} \hat{K}^2(x^2) \times (2x^2 - 1 + r/L). \quad (14)$$

The end-to-end extension PDF is then approximated as

$$Q(r, L) = C(r, L) \exp[-\beta\Delta E(r, L)]. \quad (15)$$

The prefactor $C(r, L)$ should be calculated by taking into account quadratic fluctuations to the saddle point configuration. Since such calculation is quite involved, we will actually only extend the Shimada and Yamakawa results, which were computed considering fluctuations to a closed loop. In $M = \text{mol/L}$ units this reads

$$C_{SY}(L) = \frac{1.66}{A^3} \times \frac{112.04}{(L/A)^5} \exp(0.246 \times L/A). \quad (16)$$

For this factor to fit the correct fluctuations (of the opened loop) we have to consider the fluctuations of a fake closed loop of similar bending energy. Such a loop may be obtained by considering the optimal configuration of a loop of contour length $(L+2r)$ instead of L , as shown in Fig. 4 (bottom, inset). The choice of the factor $2r$ derives from the following geometrical considerations: the closed saddle point configuration has an initial angle $\theta(0) = 49.2^\circ$, the ΔL closing the loop could be calculated for each value of r by requiring

$$\int_0^{\Delta L} \cos \theta(s) ds = r. \quad (17)$$

The angle $\theta(s)$ increases slightly on the first part of the trajectory $\theta(s) > \theta(0)$ and $\Delta L > 1.53r$. We have chosen $\Delta L = 2r$ as an approximate value; this approximation has the advantage that it can be directly put in the fluctuation expression $C(r, L) \approx C_{SY}(L+2r)$ in Eq. (15) to obtain:

$$Q(r,L) = C_{SY}(L + 2r)\exp[-\beta\Delta E(r,L)], \quad (18)$$

where ΔE is given in formula (14) and C_{SY} is given in formula (16). The validity of this approximation for the fluctuations prefactor was checked out by comparing $J(r,L)$ obtained from $Q(r,L)$ through formula (8), with the numerical results. The good agreement shown in Fig. 4 allows us to obtain a semianalytical formula for the loop probability with a finite interacting volume, which is valid for molecules of up to 2 kb (kilo base pairs).

III. RESULTS: THE EFFECT OF A PROTEIN BRIDGE

In Fig. 3 we show the cyclization factor $J(L)$ for lengths L up to 3 kb. The most probable loop length is of about $L_* = 500$ bp that is $L_*/A \approx 3.5$ [4,14]. Both shorter and longer cyclized lengths are less probable: stiffness makes difficult the bending of smaller polymers while entropy makes longer polymers ends unlikely to meet. The numerical cyclization factor is compared with the saddle point calculation of Shimada and Yamakawa [14] and with the Gaussian polymer (GP) model. The former is in good agreement with the numerical results for lengths smaller than about 1.5 kb while the latter works well for loops larger than 2.5 kb.

The effects of the finite size r of the protein bridge are displayed in Fig. 4, which shows $J(r,L)$ (on the left) and $\Delta G(r,L)$ (on the right) for lengths L ranging from 75 bp (i.e., 25 nm) to 300 bp (i.e., 100 nm) and r respectively ranging from 1 nm to 10 nm. The numerical results (on the top) are in good agreement with the SPA results (on the bottom). Note that for small lengths L , $J(r,L)$ has a peak for the $L \approx r$ event corresponding to rigid rodlike configurations. Figure 4 does not show this peak, occurring for $r=10$ nm at about $L=30$ bp (or $L=10$ nm) because we focus on the cyclization events. As shown in this figure, r values ranging from 1 nm to 10 nm make no difference for the r -dependent closure factor, considering contour lengths L larger than 300 bp (or 100 nm). The cyclization factor $J(L)$ is evaluated as the limit $r \rightarrow 0$ of $J(r,L)$. In practice, convergence is reached as soon as r is about one order of magnitude smaller than L . In the range $L > 75$ bp (25 nm), $J(5$ nm, 150 bp) converges to $J(1$ nm, 150 bp). On the other hand, $J(r,L)$ is considerably different from $J(L)$ when r is of the same order of magnitude as L . For example, for loops of $L=100$ bp (34 nm) an end-to-end extension of 10 nm increases by two orders of magnitude the closure factor. Therefore proteins of size ≈ 10 nm are expected to produce drastic enhancements in looping short DNA sequences.

In terms of energetics (see Fig. 4, right) cycling a 100 bp DNA sequence costs $25 k_B T$ when the loop ends are required to stay within a sphere of radius $r=1$ nm. This cost decreases to $13 k_B T$ if the sphere has the typical protein size of $r=10$ nm. For loop lengths larger than 300 bp the only difference in the three curves of Fig. 4 (right) is a free energy shift due to the difference in the reacting volume. For instance, a reaction radius of 10 nm decreases the looping free energy of $3 \times \ln(10) \approx 7 k_B T$ with respect to a 1 nm reaction radius.

Our results for the closure factor and the looping free energy are compared in Fig. 4 with the Monte Carlo (MC) and Brownian dynamics (BD) simulations results, obtained respectively by Podtelezhnikov and Vologodskii [17] (shown in the figure with filled circles, ●) and Langowski *et al.* [4,18] (displayed in the figure with empty squares, □). Numerical and SPA data (that for $r=1$ nm converge to the Shimada and Yamakawa curve) are in better agreement with the MC data than with the BD ones; indeed numerical, SPA and MC data are obtained with a simpler model than BD ones, which does not include twist rigidity and electrostatic effects.

Considerations about Lac operon repression energetics will help us illustrating our results and compare them with other previous results. Expression of proteins enabling bacteria *E. Coli* to perform the lactose metabolism can be prevented at the transcriptional level by cycling two different sequences [1], the smallest one including the operon promoter. Let $O_1O_3=76$ bp and $O_1O_2=385$ bp denote these two resulting DNA loops where the so-called operators $O_{1,2,3}$ actually are the small specific DNA sequences (10 bp or so) at which the tetrameric repressor protein LacR can bind thus clamping the desired loop. Notice both processes are needed for efficient repression: despite O_1O_3 containing the operon promoter, its cyclization is much less probable to occur than the O_1O_2 one (see cyclization factor $J(L)$ in Fig. 3). The LacR size is estimated from its crystallized structure to $r \approx 13$ nm [7]. In the *in vitro* experiments many parameters are under control, among which are the operators' sequence and location. The distance between the two operators, which defines the length of the DNA loop, has been fixed in [2] to about 100 bp. Our results are in good agreement with the experimental measured stability of a 114 bp DNA loop mediated by a LacR protein, obtained by Brenowitz *et al.* in 1991 [7]. By measuring the proportion of looped complexes present in a solution with respect to the unlooped molecules they obtained a looping free energy of $20.3 \pm 0.3 k_B T$ to which they associated a closure factor of $8 \times 10^{-10} M$. From Fig. 4, the closure factor of a loop of 114 bp with a protein bridge of $r=10$ nm is $J(10$ nm, 114 bp) = $10^{-9} M$ to which we associate, from formula (9) a cyclization free energy $\Delta G(10$ nm, 114 bp) = $12 k_B T$. Note that the very good agreement between the closure factor contrasts with the bad agreement for the cyclization free energy. The latter could have been calculated considering a different reaction volume or it could also include the competition with configurations that do not allow the formation of a loop (see Fig. 2 of [7]). To explain the high value found for the closure factor, Brenowitz *et al.* already included the size of the protein in the analysis of their results by comparing their J result with the value expected for the cyclization probability of a free DNA when the length of the protein is included in the size of the loop.

Another result on a DNA loop mediated by LacR protein has been obtained by Balaeff *et al.* [15] who have numerically calculated the elastic energy of the O_1O_3 loop from a WLC model also including: the twist rigidity, the short range electrostatic repulsion and the details of the LacR/DNA complex crystal structure. The elastic energy is estimated to be $23 k_B T$ in [15], of which 81% (that is $18 k_B T$) is due to the bending and 19% is due to the unwinding. The bending energy of $18 k_B T$ is to be compared to the saddle point energy

(14) of $15 k_B T$ for a 75 bp loop with an end-to-end extension of $r=10$ nm. The corresponding free energy of loop formation obtained by (18) and after integration over the reacting volume (9) is $\Delta G(10 \text{ nm}, 75 \text{ bp})=14 k_B T$ (see Fig. 4, bottom/right). For a 400 bp loop since LacR is only $\approx 10\%$ of the loop length, its size is expected to play a less important role. Indeed the cyclization probability does not depend on $r < 10$ nm and the free energy of forming such a loop decreases from $15 k_B T$ for $r=1$ nm to $8 k_B T$ for $r=10$ nm only because the reaction volume increases by a factor $3 \times \ln(10) \approx 7 k_B T$.

IV. EFFECT OF THE PRESENCE OF A KINK

The previous protein mediated DNA looping modelization (Sec. III) assumes that the only intervening proteins are the ones clamping the loop ends (e.g., the Lac operon repressor LacR). Actually regulation phenomena involve several proteins which may bind along the whole DNA. Indeed naked DNA situations barely exist *in vivo*. For instance, single molecule manipulations [2] have shown that efficient Gal operon repression needs a stiffness loss of the 113 bp DNA portion to be looped. The HU protein produces such loss by kinking the sequence.

A. Numerical calculation of J for a bridged and kinked loop

Such stiffness loss may be taken into account in an effective way by kinking the WLC at half-length $L/2$. Let us call θ_- and θ_+ the angles of the DNA just before and after the kink respectively. We assume that the kink plane is vertical and choose it to define the origin of the azimuthal angle ϕ : $\phi_- = \phi_+ = 0$. Using the quantum language of Sec. II A, we replace the calculation of the evolution operator $Z(L, f)$ (11) with its kinked counterpart

$$\begin{aligned} Z_{\text{kinked}}(L, f) &= \langle \text{final} | \exp \left[-\frac{L}{2A} \hat{H}(f) \right] | \theta_+, \phi_+ \rangle \\ &\times \langle \theta_-, \phi_- | \exp \left[-\frac{L}{2A} \hat{H}(f) \right] | \text{initial} \rangle \\ &= \sum_{\ell, \ell'} \langle 0, 0 | \exp \left[-\frac{L}{2A} \hat{H}(f) \right] | \ell, 0 \rangle \langle \ell', 0 | \\ &\times \exp \left[-\frac{L}{2A} \hat{H}(f) \right] | 0, 0 \rangle \times Y_{\ell}^0(\theta_+, 0) Y_{\ell'}^0(\theta_-, 0), \end{aligned} \quad (19)$$

where we have used the change of basis from angles to spherical harmonics

$$| \theta_{\pm}, \phi_{\pm} = 0 \rangle = \sum_{\ell \geq 0} Y_{\ell}^0(\theta_{\pm}, 0) | \ell, 0 \rangle. \quad (20)$$

Although calculations are a little bit more involved, the evaluation scheme for the cyclization factor J remains unchanged in its principle (Sec. II A). We now have “ $|\ell \neq 0, 0\rangle$ elements” corresponding to particular orientations arriving at (–) and leaving from (+) $s=L/2$. The kink angle κ is intuitively defined from these WLC tangent vectors at

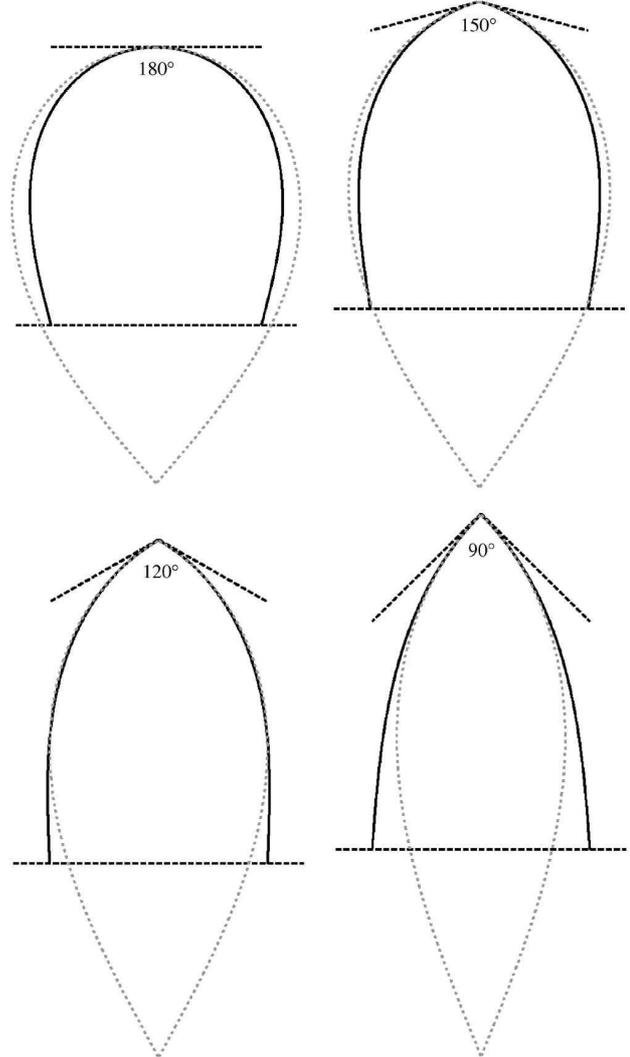


FIG. 5. Lowest bending energy configurations for a 100 bp DNA loop with an end-to-end extension $r=10$ nm and kinks $\kappa=90^\circ, 120^\circ, 150^\circ, 180^\circ$, respectively. The gray configurations are the closed loops used in the calculation of the fluctuation.

half-length (in spherical coordinates) through

$$\kappa = \theta_+ - \theta_-. \quad (21)$$

B. Saddle point approximation of J for a bridged and kinked loop

The saddle point calculation of Sec. II B can be straightforwardly extended to the case of a kinked loop. In Fig. 5 we show the configurations with the lowest bending energy for a 100 bp DNA loop with an end-to-end extension $r=10$ nm. The kink is accounted for by a bending angle in the middle of the chain $\kappa=2 \times \theta(L/2) - \pi$ a priori different from the previous trivial value $\kappa=\pi$, ranging from 150° to 90° . We introduce the phase $\psi = \arcsin[\sin(\pi - \kappa/4)/x]$. The parameter x is now obtained from equation

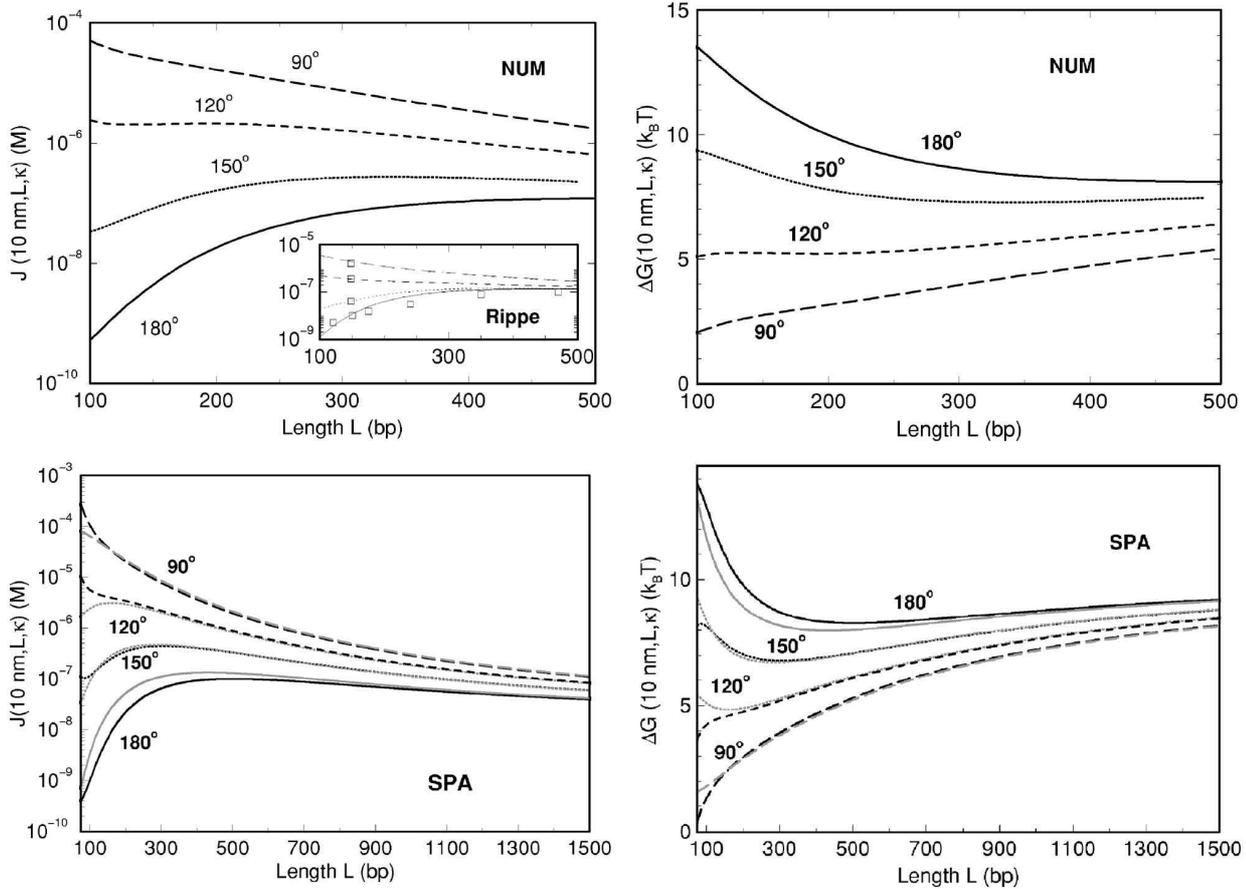


FIG. 6. Closure factor (left) and free energy for a loop with $r=10$ nm and a kink κ in the middle of the chain: $\kappa=180^\circ$ (full lines); $\kappa=150^\circ$ (dotted lines); $\kappa=120^\circ$ (dashed line); $\kappa=90^\circ$ (long dashed lines). Top: numerical calculation of the constrained path integral. Bottom: extension of the Shimada and Yamakawa calculation (black lines) and approximate formula (gray line) given in the text (25). Numerical results are in very good agreement with the SPA approximation and the approximate formula (25). Inset: Brownian dynamics simulations point obtained by the Langowski and collaborators (empty square, \square) [4,18], fitted by a simple formula by Rippe [19].

$$(1 + r/L)[\hat{K}(x^2) - K(\psi, x^2)] = 2[\hat{E}(x^2) - E(\psi, x^2)] \quad (22)$$

and the total bending energy is

$$\beta \Delta E(r, L, \kappa) = \frac{4}{L/A} [\hat{K}(x^2) - K(\psi, x^2)]^2 (2x^2 - 1 + r/L). \quad (23)$$

In analogy with (15) we obtain the end-to-end extension \vec{r} PDF

$$Q(r, L, \kappa) = C_{SY}(L + 2r) \exp[-\beta \Delta E(r, L, \kappa)], \quad (24)$$

where C_{SY} is given in (16), from which we calculate $J(r, L, \kappa)$ through formula (8). Note that (24) reduces to the loop probability given in [23] for a closed and kinked DNA. Again the good agreement obtained with numerical results allows us to establish a semianalytical formula for the loop probability with a finite interacting volume and kink.

C. Results for a kinked and bridged loop

In Fig. 6 results for the looping probability density (left)

and free energy (right) are shown for a typical end-to-end extension $r=10$ nm and kinks $\kappa=150^\circ$, 120° , and 90° . The curve with no kink ($\kappa=180^\circ$) is also shown for comparison. The numerical results on the top of the figure are in very good agreement with the extension of the Shimada and Yamakawa saddle point approach on the bottom of the figure (black lines). In Fig. 6 (inset) we show the results obtained by BD simulations by Langowski *et al.* [4,18] for $J(10 \text{ nm}, L, \kappa)$ (empty squares, \square) fitted by Rippe [19] with a simple formula containing one fitting parameter for each curve. The curves in the inset of Fig. 6 reproduce the same behavior with L and κ of our numerical (top of Fig. 6) or SPA curves (bottom of Fig. 6). The numerical gap with BD results increases for large kinks: at $\kappa=90^\circ$ our closure factor is 10 times larger than the closure factor obtained by BD simulations. The electrostatic and twist rigidity effect could indeed play a more important role when the chain is kinked. Note that the lengths' range of the numerical results are from 100 bp to 500 bp, while the lengths' range of the saddle point results is from 75 bp to 1500 bp. The lengths' range of the SPA is limited by the validity of the approximation ≈ 1500 bp shown in Fig. 3. As an example, the closure factor J

(10 nm, 113 bp, κ) of a 113 bp fragment, obtained by the numerical calculations, increases from the value of $10^{-9} M$ for a nonkinked loop ($\kappa=180^\circ$) to $4 \times 10^{-8} M$, $2 \times 10^{-6} M$, and $4 \times 10^{-5} M$ for, respectively, $\kappa=150^\circ$, 120° , and 90° . The corresponding looping free energy is ΔG (10 nm, 113 bp, κ) = $13 k_B T$, $9 k_B T$, $5 k_B T$ and $2.5 k_B T$ for $\kappa=180^\circ$, 150° , 120° , and 90° respectively. With the saddle point approach we find similar results: J (10 nm, 113 bp, κ) = $2 \times 10^{-9} M$, $10^{-7} M$, $5 \times 10^{-6} M$, and $8 \times 10^{-5} M$ while ΔG (10 nm, 113 bp, κ) = $11 k_B T$, $8 k_B T$, $4.4 k_B T$, and $1.7 k_B T$. As it is shown in Fig. 6, the presence of the kink becomes irrelevant for DNA segments larger than about 1500 bp. The example of a 113 bp DNA segment has been chosen to compare the results with the single-molecule experiments on the GalR mediated loop of an ≈ 113 bp DNA portion between the two operators. From the kinetics of the loop formation mediated by the GalR and HU proteins, Lia *et al.* [2] have deduced a looping free energy of $12 k_B T$, that with respect to our values should correspond to a kink angle of more than 150° or to an end-to-end extension smaller than $r=10$ nm. Another significant change in the cyclization probability is that the stiffness loss induced by κ reduces the most probable loop length from 500 bp for a nonkinked DNA to 340 bp and 190 bp (from the numerical calculation) or to 300 bp and 150 bp (from SPA) for kinks of respectively 150° and 120° . Note that for a kinked loop with an end-to-end extension r the minimal length L_0 corresponding to the rigid rodlike configuration fulfills the relation $L_0 \sin(\kappa/2)=r$ and therefore it is of ≈ 42 bp for $\kappa=90^\circ$ instead of ≈ 30 bp for $\kappa=180^\circ$. For $\kappa=90^\circ$ the most probable loop length is the rigid *kinked* rodlike configuration of the two half-DNA portions. To catch both kink and protein bridge effects in a simple formula, we have calculated the cyclization factor with the extension of Shimada and Yamakawa formula for a kinked closed loop of length $(L+2r)$. This approach is similar to what was suggested in 1991 by Brenowitz *et al.* [7] to interpret their experimental data, i.e., to directly consider the protein as part of the length of the loop. A linear fit [26] of the bending energy (23) for the optimal closed configuration ($r=0$) in the presence of a kink (expressed in degrees), $\beta \Delta E(r=0, L, \kappa) \approx (-7.1 + 0.1155\kappa)/(L/A)$, is shown in Fig. 7. It gives the following approximated formula for the closure factor as a function of the protein size r , the length L of the DNA, and the kink angle κ

$$J_{\text{approx}}(L, r, \kappa) = C_{\text{SY}}(L + 2r) \exp \left[\frac{7.1 - 0.1155\kappa}{(L + 2r)/A} \right], \quad (25)$$

where C_{SY} is given in (16). Notice the integration step (8) has been skipped since it does not make any significant difference. Formula (25) allows us to obtain a simple prediction for the loop probability in the presence of a kink in the middle of the sequence and a finite separation between the extremities. As shown in Fig. 6 this formula (gray lines) is in good agreement with the loop probability obtained with the exact calculation of the saddle point energy of the open configuration (full line). In particular, Fig. 6 shows that for kink angles in the range $90^\circ < \kappa < 150^\circ$ this simple formula works remarkably well for lengths L larger than about $5r$,

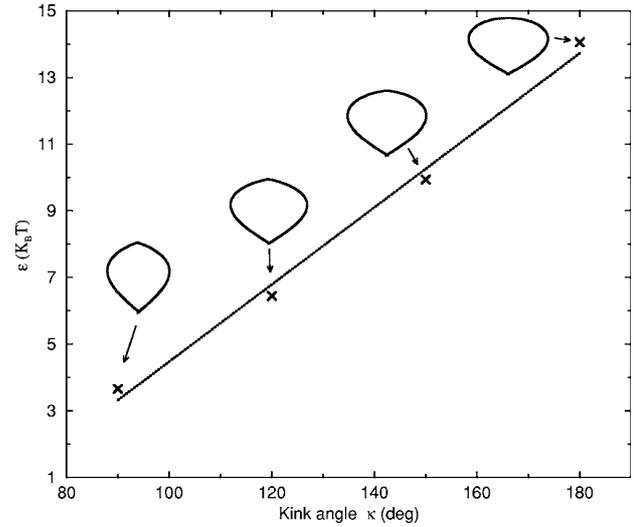


FIG. 7. The “x” points: saddle point energy in units of L/A (that is, $\varepsilon = \Delta E \times L/A$) for the saddle point configurations (also displayed in the figure) with kink angles of $\kappa=90^\circ, 120^\circ, 150^\circ, 180^\circ$. Dotted line: linear interpolation used in formula (25).

that is 150 bp for $r=10$ nm. For smaller lengths the optimal configuration is more a rigid rodlike configuration and cannot be approximated by a closed loop. Similar simple formulas that includes a kink angle κ and a finite end-to-end distance r in an effective way have also been written down by Rippe or Ringrose to fit their Brownian dynamics simulation [19] or experimental data [29], but these formulas contain a parameter that must be fitted for each values of r and κ from the data points (Fig. 6).

V. CONCLUSION

We performed both numerical and analytical calculations of the closure factor J , even in the presence of a protein bridge and of a protein-mediated kink. More precisely we have numerically calculated the path integral of the WLC polymer model under the constraints of a fixed end-to-end distance r and a kink κ in the middle of the DNA portion. Analytically we have extended the Shimada and Yamakawa saddle point approximation [14] to the case of a bridged and kinked loop. We have seen that the formation of DNA loops is significantly sensitive to the size of the protein bridge when this size r is more than 10% of the loop length L , that is, 300 bp (or 100 nm) for a typical protein bridge size of $r=10$ nm. To give an example, the closure factor for a 100 bp DNA segment increases from $J(100 \text{ bp}, 0) \approx 10^{-11} M$ to $J(100 \text{ bp}, 10 \text{ nm}) \approx 10^{-9} M$. Correspondingly, looping free energy decreases from $\Delta G(100 \text{ bp}, 0) = 24 k_B T$ to $\Delta G(100 \text{ bp}, 10 \text{ nm}) = 13 k_B T$. A kink ranging from 150° to 90° produces a significant increase of J for DNA fragments of lengths up to about 2500 bp. For instance, the closure factor for a 100 bp DNA segment with a protein bridge $r=10$ nm and a kink of 90° is $J(100 \text{ bp}, 10 \text{ nm}, 90^\circ) \approx 10^{-4} M$, and the corresponding looping free energy $\Delta G(100 \text{ bp}, 10 \text{ nm}, 90^\circ) \approx 2 k_B T$. A kink also changes the most probable

loop length from 500 bp (no kink) to about 175 bp for a kink of 120° , going to the rigid *kinked* rodlike configuration for smaller κ values. This is an interesting mechanism because the loop lengths implied in *in vivo* DNA processing by proteins spread on a large lengths range. Our results were compared to previous analytical approximations (in particular the results of the Gaussian model, the Wilhelm and Frey expansion [16] and the Shimada and Yamakawa formula [14]) and numerical calculations (in particular the Monte Carlo simulations data obtained by Podtelezchnikov and Vologodskii [17], and the Brownian dynamics simulations data obtained by Langowski *et al.* [4,18,19]) as well as experimental results [1,2,7] (e.g., the ones obtained by Lia *et al.* on the looping dynamics mediated by the Gal and HU proteins). Finally a simple formula (25) including both the protein bridge and kink effects has been proposed. This formula has the advantage of not containing adjustable parameters with respect to the existing formulas that include both these effects [19].

Still, many effects omitted in this work can be included without significant changes in the numerical algorithm. The kink we considered is actually permanent (that is, not thermally excited), site specific (at half-length), and rigid (κ fixed). Although this rigidity seems relevant to most protein bindings to DNA at first glance [30], it was pointed out in the works of Yan *et al.* [20,22] that kinks may also be semiflexible (exhibiting higher or lower rigidities than the bare DNA) or even fully flexible [21,31]. For instance, the HU/DNA complex was recently observed to be very flexible under specific experimental conditions [32]. Flexible hinges were also

stated to occur along the DNA due to the opening of small denaturation bubbles [21,33], such as the one needed by HU to fit in the double helix [2]. Such flexibility could be taken into account [30] in our model. These kinds of defects could also be thermally activated, occurring at multiple nonspecific sites along the DNA [20,22,30,31,34]. Both effects may be included in our model. Note that using effective persistence lengths could turn out convenient, despite little information about the kink properties (number, location, rigidity, etc.). Actually this would be equivalent to study DNA stiffness loss due to sequence effects [35] by cutting WLC in different stiff fragments, depending on the CG or AT content of the whole sequence to cycle. The same approach may allow an approximative study of the DNA polyelectrolyte nature too [36]; otherwise, electrostatic potential has to be included in WLC energy. Twist elasticity leads to slight modifications of the quantum analog we used although requiring some care [37]. This is expected to play an important role in looping, especially when specific alignment of the loop extremities are required. Finally cyclization dynamics could be modeled using a simple two-states model where DNA is “closed” or “opened,” that is, cycled or not. Such study relies on the (statics) cyclization factor we computed in this article [38]. Direct comparison to experimental lifetimes measures would be possible [1,2].

ACKNOWLEDGMENTS

The authors would like to thank D. Chatenay, R. Monas-son, and F. Thalmann for useful discussions and a critical reading of the manuscript.

[1] L. Finzi and J. Gelles, *Science* **267**, 378 (1995).
 [2] G. Lia, D. Bensimon, V. Croquette, J.-F. Allemand, D. Dunlap, D. E. A. Lewis, S. Adhya, and L. Finzi, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11373 (2003).
 [3] J.-F. Allemand, S. Cocco, N. Douarche, and G. Lia, *Eur. Phys. J. B* (to be published).
 [4] K. Rippe, P. H. von Hippel, and J. Langowski, *Trends Biochem. Sci.* **20**, 500 (1995).
 [5] D. Shore, J. Langowski, and R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4833 (1981).
 [6] Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskaya, and A. Vologodskii, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5397 (2005).
 [7] M. Brenowitz, A. Pickar, and E. Jamison, *Biochemistry* **30**, 5986 (1991).
 [8] H. Yamakawa, *Helical Wormlike Chains in Polymer Solutions* (Springer-Verlag, Berlin, 1997).
 [9] H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets* (World Scientific Publishing Co., Singapore, 2004).
 [10] A. Hanke and R. Metzler, *Biophys. J.* **85**, 167 (2003).
 [11] O. Kratky and G. Porod, *Recl. Trav. Chim. Pays-Bas* **68**, 1106 (1949).
 [12] H. E. Daniels, *Proc. - R. Soc. Edinburgh, Sect. A: Math.* **63**, 290 (1952).
 [13] W. Gobush, H. Yamakawa, W. H. Stockmayer, and W. S. Ma-gee, *J. Chem. Phys.* **57**, 2839 (1972).
 [14] J. Shimada and H. Yamakawa, *Macromolecules* **17**, 689 (1984).
 [15] A. Balaeff, L. Mahadevan, and K. Schulten, *Phys. Rev. Lett.* **83**, 4900 (1999).
 [16] J. Wilhelm and E. Frey, *Phys. Rev. Lett.* **77**, 2581 (1996).
 [17] A. A. Podtelezchnikov and A. V. Vologodskii, *Macromolecules* **33**, 2767 (2000).
 [18] H. Merlitz, K. Rippe, K. V. Klenin, and J. Langowski, *Bio-phys. J.* **74**, 773 (1998).
 [19] K. Rippe, *Trends Biochem. Sci.* **26**, 733 (2001).
 [20] J. Yan and J. F. Marko, *Phys. Rev. E* **68**, 011905 (2003).
 [21] J. Yan and J. F. Marko, *Phys. Rev. Lett.* **93**, 108108 (2004).
 [22] J. Yan, R. Kawamura, and J. F. Marko, *Phys. Rev. E* **71**, 061905 (2005).
 [23] S. Sankararaman and J. F. Marko, *Phys. Rev. E* **71**, 021911 (2005).
 [24] J. Samuel and S. Sinha, *Phys. Rev. E* **66**, 050801(R) (2002).
 [25] R. B. Sidje, *ACM Trans. Math. Softw.* **24**, 130 (1998). The EXPOKIT library is available for download at URL <http://www.maths.uq.edu.au/expokit>
 [26] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: the Art of Scientific Computing* (Cambridge University Press, Cambridge, England, 1992).
 [27] C. Moler and C. van Loan, *SIAM Rev.* **45**, 3 (2003).

- [28] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover Publications Inc., New York, 1970). The elliptic integrals of first and second kinds respectively read in Legendre definition $K(\psi, x^2) = \int_0^\psi dy / \sqrt{1-x^2 \sin^2 y}$ and $E(\psi, x^2) = \int_0^\psi \sqrt{1-x^2 \sin^2 y} dy$.
- [29] L. Ringrose, S. Chabanis, P.-O. Angrand, C. Woodroffe, and A. F. Stewart, *EMBO J.* **18**, 6630 (1999).
- [30] Y. O. Popov and A. V. Tkachenko, *Phys. Rev. E* **71**, 051905 (2005).
- [31] P. A. Wiggins, R. Phillips, and P. C. Nelson, *Phys. Rev. E* **71**, 021909 (2005).
- [32] J. V. Noort, S. Verbrugge, N. Goosen, C. Dekker, and R. T. Dame, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6969 (2004).
- [33] P. Ranjith, P. B. Sunil Kumar, and G. I. Menon, *Phys. Rev. Lett.* **94**, 138102 (2005).
- [34] B. Chakrabarti and A. J. Levine, *Phys. Rev. E* **71**, 031905 (2005).
- [35] P. Nelson, *Phys. Rev. Lett.* **80**, 5810 (1998).
- [36] R. Zandi, J. Rudnick, and R. Golestanian, *Phys. Rev. E* **67**, 021803 (2003).
- [37] C. Bouchiat and M. Mézard *Phys. Rev. Lett.* **80**, 1556 (1998); V. Rossetto and A. C. Maggs, *ibid.* **88**, 089801 (2002) (comment); C. Bouchiat and M. Mézard, *ibid.* **88**, 089802 (2002) (reply).
- [38] S. Jun, J. Bechhoefer, and B.-Y. Ha, *Europhys. Lett.* **64**, 420 (2003).

Protein-Mediated DNA Loops: Effects of Protein Bridge Size and Kinks

Nicolas DOUARCHE and Simona COCCO

CNRS – Laboratoire de Physique Statistique de l'ENS, 24 rue Lhomond, 75005 Paris, France

This paper focuses on the probability that a portion of DNA closes on itself through thermal fluctuations. We investigate the dependence of this probability upon the size r of a protein bridge and/or the presence of a kink at half DNA length. The DNA is modeled by the Worm-Like Chain model, and the probability of loop formation is calculated in two ways: exact numerical evaluation of the constrained path integral and the extension of the SHIMADA and YAMAKAWA saddle point approximation. For example, we find that the looping free energy of a 100 base pairs DNA decreases from $24 k_B T$ to $13 k_B T$ when the loop is closed by a protein of $r = 10$ nm length. It further decreases to $5 k_B T$ when the loop has a kink of 120° at half-length.

PACS numbers: 36.20.Hb, 46.70.Hg, 82.35.Pq, 87.14.Gg, 87.15.Aa, 87.15.La

I. INTRODUCTION: DNA LOOPS IN GENE TRANSCRIPTION REGULATION

Gene expression is regulated by a wide variety of mechanisms. These activation as well as repression phenomena may occur at every expression steps (translation, transcription, etc) and do involve interactions between several biological molecules (DNA, RNAs, proteins, etc). For instance proteins bound on specific DNA sequences may turn on/off genes transcription by interacting with each other. By bringing closer those proteins DNA looping can ease their interactions [4, 5, 6, 17]. Such looping events may be observed over a wide range of lengths spreading from hundreds to thousands base pairs (bp). The looping probability has been firstly measured by the cyclization of DNA segments in solution with cohesive ends. Once the loop is formed proteins called ligases stabilize it [1, 2]. It is then possible to count the circular DNAs with respect to the linear ones. The loop formation mediated by proteins have also been experimentally studied. Two examples are the loops formed by the LacR or the GalR transcriptional repressors [3]. Two units of such proteins bind at two specific positions along a same DNA and associate to form a complex when the binding sites come in contact. The formation of such loops has been recently studied using micromanipulation experiments on a single DNA molecule [4, 5]. The study of the GalR mediated loop has shed light on the role of a third protein called HU that sharply bends (*i.e.* kinks) the DNA at half-length.

The DNA loop probability depends mainly on its length and flexibility. Long DNAs (typically longer than 1500 bp) essentially behave as Gaussian Polymers (GP) [8, 9]: the cyclization cost is mainly of an entropic nature [12]. On the contrary, for small lengths DNA cyclization is difficult mainly because of the bending energy cost. The computation of the elastic energy for the Worm-Like Chain (WLC) model [7, 8, 9] can be analytically performed [15]; numerical methods have also been employed when electrostatic properties are included [13]. At intermediate length scales (from about 500 bp) elastic rigidity and entropic loss are both important. Several approximations have been developed to study this lengths range [6, 10, 11, 14], among them the calculations of the fluctuations around the lowest bending energy configurations performed by SHIMADA and YAMAKAWA [15]. Numerical approaches have also been developed: Monte Carlo [16] and brownian dynamics based simulations [17, 18, 19] as well as numerical calculations of the WLC path integral under the closed ends constraint [20, 21]. This last method allowed YAN, KAWAMURA and MARKO to study the elastic response of DNA subject to permanent or thermally excited bendings caused by binding proteins (such as HU) or inhomogeneities along the DNA double helix [21, 22]. All these studies do provide a better understanding of the underlying regulation phenomena despite their overall complexity.

In this paper we study two processes that turned out to be important in DNA looping. Namely the size of the protein complex clamping the loop [17, 18], acting as a bridge between the two DNA ends, and mechanisms implying DNA stiffness loss which are taken into account in an effective way by kinking the WLC at half-contour length [17, 18, 19, 23]. In section II we define the model and the methods: we describe the numerical approach (§ II A) and the analytical Saddle Point Approximation (SPA, § II B) to calculate the r -dependent closure factor and the looping free energy. In section III we compare the numerical and SPA results with previous experimental and theoretical results. In section IV we extend the numerical and SPA approaches to a kinked loop; we discuss our results and we propose a simple formula that accounts for both the protein bridge and kink effects (§ IV C). We conclude (section V) by sketching how to include omitted DNA properties which may also play an important role in its closure such as twist rigidity or electrostatic interactions.

II. DEFINITIONS AND METHODS

We use the well known Worm-Like Chain (WLC) model [7, 8, 9]. The DNA polymer is described as an *inextensible* continuous differentiable curve of contour length L , with unit tangent vector $\vec{t}(s)$ ($0 \leq s \leq L$). The polymer is characterized by the persistence length A beyond which tangent vectors lose their alignment: $\langle \vec{t}(s) \cdot \vec{t}(s') \rangle = \exp(-|s - s'|/A)$. The energy of a configuration of the polymer stretched under an external force $f \vec{e}_z$ reads

$$E[\vec{t}; L, f] = \frac{1}{2} \frac{A}{\beta} \int_0^L \left[\frac{d\vec{t}(s)}{ds} \right]^2 ds - f \int_0^L \vec{e}_z \cdot \vec{t}(s) ds \quad (1)$$

where we use $\beta = 1/k_B T$. No twist elasticity nor extensibility will be considered. The partition function is

$$Z(L, f) = \int \mathcal{D}\vec{t} \exp\{-\beta E[\vec{t}; L, f]\}. \quad (2)$$

Notice that summation over all initial and final tangent vectors orientations, $\vec{t}(0)$ and $\vec{t}(L)$, is implicitly understood in this path integral.

In this paper we are interested in the formation of a loop in a DNA molecule, and the Probability Density Functions (PDFs) of end-to-end distances play an important role. The quantities under study are denoted by Q , S , P and J respectively and defined as follows:

- the end-to-end extension $\vec{r} = (x, y, z)$ PDF at zero force,

$$Q(\vec{r}; L) = \frac{1}{Z(L, f=0)} \int \mathcal{D}\vec{t} \delta \left[\int_0^L \vec{t}(s) ds - \vec{r} \right] \exp\{-\beta E[\vec{t}; L, f=0]\}. \quad (3)$$

In the absence of force Q depends on its argument \vec{r} through its modulus $r = |\vec{r}|$ only, and we may introduce the radial PDF

$$S(r, L) = 4\pi r^2 Q(r, L). \quad (4)$$

- The z extension PDF reads

$$P(z, L) = \int_{-L}^L dx \int_{-L}^L dy Q[(x, y, z), L]. \quad (5)$$

In the absence of external force, notice the x or y extensions PDFs are given by P too. Interestingly, the radial and z extensions PDF are related to each other through the useful identity [24]

$$S(r, L) = -2r \frac{dP}{dz}(r, L). \quad (6)$$

- The cyclization factor

$$J(L) = Q(0, L) \quad (7)$$

defined as the density of probability that the two ends of the DNA are in contact with one another.

- The r -dependent closure factor is

$$J(r, L) = \left[\int_0^r S(r', L) dr' \right] / \left(\frac{4}{3} \pi r^3 \right). \quad (8)$$

It gives the density probability for the two ends of the chain to stay within a sphere of radius r . It is easy to check that $J(r, L) \rightarrow J(L)$ when $r \rightarrow 0$. For experimental convenience, units used for $J(L)$ and $J(r, L)$ are moles per liter: $1 \text{ nm}^{-3} \approx 1.66 \text{ mol} \cdot \text{L}^{-1} \equiv 1.66 \text{ M}$. In these units $J(r, L)$ gives directly the concentration of one binding site in proximity of the other.

- Finally we consider the looping free energy cost

$$\beta\Delta G(r, L) = -\ln\left[J(r, L) \times \frac{4}{3}\pi r^3\right]. \quad (9)$$

Note that this definition does not include the details of the geometry nor the affinities of the DNA/protein and protein/protein interactions. We actually assume all the sphere of radius r to be the reacting volume, *i.e.* that the loop will form if the DNA ends happen to be in this sphere.

Despite intense studies of the WLC model no exact analytical expression is known for Q and the quantities of interest here, namely J and S . However approximations expanding from the two limiting regimes (entropic [10, 11] and elastic [14, 15]) along with exact numerical computations are available. Hereafter, we have resorted to numerical as well as approximate analytical techniques (SPA) for calculating the cyclization factor J and the probability of almost closed DNA configurations.

A. Numerical Calculation of the Probabilities P , S and J

Our starting point for the calculation of the z extension PDF is the FOURIER representation of the DIRAC δ -function in (3) and (5),

$$P(z, L) = \int_{-\infty}^{+\infty} \frac{dk}{2\pi} e^{+ikz} Z(L, f = -ik/\beta). \quad (10)$$

At fixed momentum k we are left with the calculation of the partition function $Z(L, f)$ at (imaginary) force $f = -ik/\beta$. The path integral (2) defining Z is interpreted as the evolution operator of a quantum system, the rigid rotator under an external imaginary field

$$Z(L, f) = \langle \text{final} | \exp\left[-L/A \times \widehat{H}(f)\right] | \text{initial} \rangle. \quad (11)$$

The entries of its hamiltonian \widehat{H} are easily expressed in the spherical harmonics $|\ell, m\rangle$ basis: $\langle \ell, m | \widehat{H}(f) | \ell', m' \rangle = H_{\ell, \ell'}(f) \delta_{m, m'}$ with

$$H_{\ell, \ell'}(f) = \frac{\ell(\ell+1)}{2} \delta_{\ell, \ell'} - \beta A f \frac{\ell \delta_{\ell, \ell-1} + \ell' \delta_{\ell, \ell'-1}}{\sqrt{(2\ell+1)(2\ell'+1)}}. \quad (12)$$

The entries of \widehat{H} do not depend on the azimuthal number m due to cylindrical symmetry around the force axis. Finally, integration over all initial and final orientations for the tangent vectors at the ends of the polymer chain selects $|\text{initial}\rangle = |\text{final}\rangle = |0, 0\rangle$.

A recent paper [24] used Mathematica to compute the vacuum amplitude (11) through a direct matrix exponentiation. We have instead used the Expokit library [25] since it proves to be more accurate and faster for intensive numerical calculations. We truncated hamiltonian (12) in a way the outcome is insensitive to the cut-off on the harmonics order. The (inverse) FOURIER transform (10) is then handled by a Fast FOURIER Transform (FFT) algorithm [26]. This task is in particular facilitated thanks to the inextensibility constraint which makes the distribution bandwidth limited. Our results for the z extension PDF are shown in Fig. 1. The cross-over from the rigid elastic regime ($L/A = 0.1, 0.5, 1$) to the flexible entropic regime ($L/A = 5, 10, 15$) is clearly visible. Using (6) then gives us access to S , the radial extension PDF (see Fig. 2). The most probable value for the distance r_* , switches from full extension $r_* \lesssim L$ for contour lengths $L \lesssim A$ (elastic dominated regime) to the GP most probable extension $\sqrt{4LA/3}$ for longer ones $L \gg A$ (entropic dominated regime). Also note that S always goes continuously to zero in the $r \rightarrow L$ limit due to WLC inextensibility. We have finally calculated the r -dependent closure factor J according to (8) by numerical integration of S , and the looping free energy cost ΔG defined in (9).

Let us now discuss the numerical errors that could be important when calculating probabilities of rare events. Main sources of error are the hamiltonian (12) truncation and the integration step to compute the r dependent closure factor (8). As mentioned above the cut-off on the harmonic order was systematically chosen in order convergence is observed. We have used $\ell = 50$ after having verified that the result is unchanged for $\ell = 100$. Concerning integration, limitation comes from the number of available data in the range $0 \leq r' \leq r$ which is directly related to the k sampling of the partition function Z . For $r = 1$ nm the numerical integration is still reliable, but decreasing further r turns out to be critical. Other potential sources of error are negligible. Indeed bandwidth limitation of $P(z; L)$ prevents

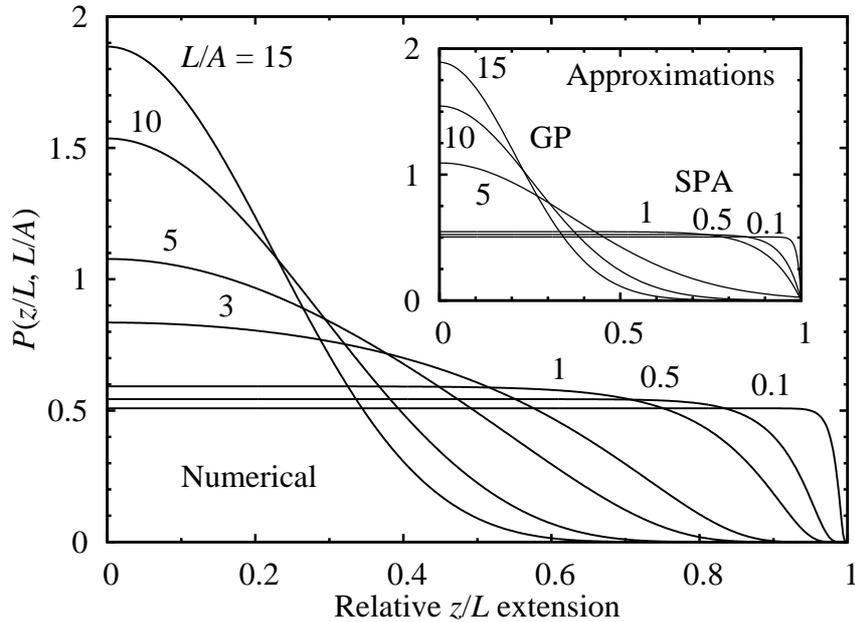


FIG. 1: Numerical computation of the z extension PDF over a wide range of contour lengths L . As expected the agreement with the SPA prediction (see section II B) improves as L decreases (upper bound $L/A \lesssim 1$). The long WLC behavior is caught by the GP approximation [8, 9] as soon as, say $L/A \gtrsim 5$.

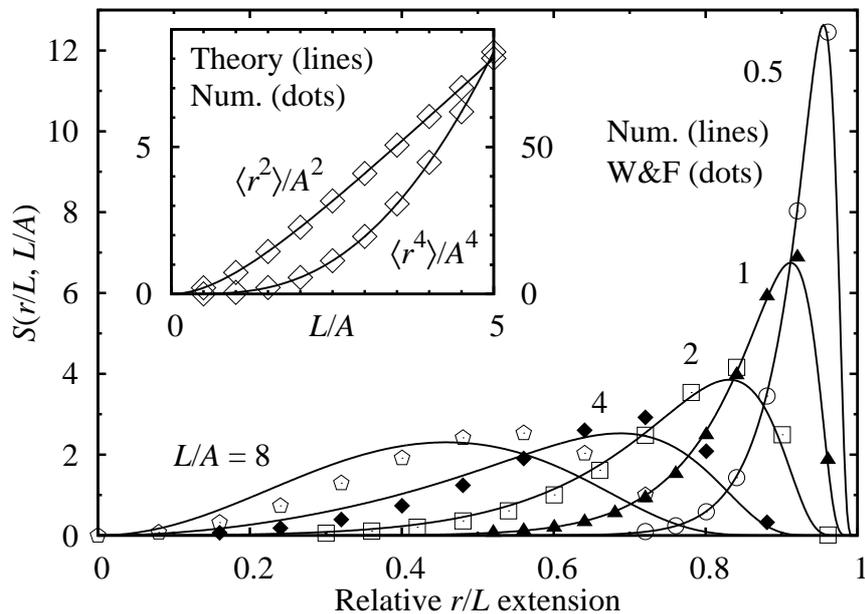


FIG. 2: Numerical computation of S , the radial r extension PDF. The outcomes of the numerical calculations are tested against exactly known values for the first even moments $\langle r^{2n} \rangle$ (inset) [8, 9]. The shape of S compares very well to the widely used WILHELM and FREY (W&F) expansion [14], valid up to $L/A \lesssim 2$. Similar tests were achieved with other popular approximation schemes [6] (data not shown).

any aliasing [26] during FFT (10) and the derivative of P (6) can actually be skipped by an integration by part of (8) to compute J . Further hypothetical errors would then come from the Expokit library [25] itself but its routines were

coded to compute accurately matrix exponentials over a broad range of matrices [27]. We have checked the numerical precision of our method by the comparison with the exact values for the first even moments of $S(r)$ (Fig. 2 Inset); moreover as shown in Fig.2 $S(r)$ agrees with Wilhelm and Frey expansion for small L/A values; finally we will see in section III that the numerically calculated cyclization factor $J(L)$ (8) is in agreement with the SHIMADA-YAMAKAWA and Gaussian approximation results for respectively small and large L (Fig. 3).

B. Saddle Point Approximation for J

In addition to the exact numerical calculations detailed above, we have carried out approximate calculations based on a saddle point estimate of the partition function (2). We follow the SHIMADA and YAMAKAWA calculation for the saddle point configuration [15], extending it for an opened DNA. The saddle point configuration for a closed loop is shown in Fig. 3 inset: it is a *planar loop*. The tangent vector at position s along the chain is characterized by its angle $\theta(s)$ with respect to the end-to-end extension \vec{r} . The optimal configuration is symmetric with respect to the perpendicular \vec{e}_\perp -axis while the half-length angle is $\theta(L/2) = 180^\circ$. The initial angle $\theta(0)$ is chosen by the minimization of the bending energy of the chain. The optimization gives rise to the following condition on the parameter $x = \cos[\theta(0)/4]$

$$(1 + r/L) \widehat{K}(x^2) = 2 \widehat{E}(x^2), \quad (13)$$

where $\widehat{K}(x^2) = K(\frac{\pi}{2}, x^2)$ and $\widehat{E}(x^2) = E(\frac{\pi}{2}, x^2)$ are the complete elliptical integrals of the first and second kinds respectively [28]. The corresponding elastic energy is

$$\beta \Delta E(r, L) = \frac{4}{L/A} \widehat{K}^2(x^2) \times (2x^2 - 1 + r/L). \quad (14)$$

The end-to-end extension PDF is then approximated as

$$Q(r, L) = C(r, L) \exp[-\beta \Delta E(r, L)]. \quad (15)$$

The prefactor $C(r, L)$ should be calculated by taking into account quadratic fluctuations to the saddle point configuration. Since such calculation is quite involved we will actually only extend the SHIMADA and YAMAKAWA results which was computed considering fluctuations to a closed loop. In $M = \text{mol} \cdot L^{-1}$ units this reads

$$C_{\text{SY}}(L) = \frac{1.66}{A^3} \times \frac{112.04}{(L/A)^5} \exp(0.246 \times L/A). \quad (16)$$

For this factor to fit the correct fluctuations (to the opened loop) we have to consider the fluctuations to a fake closed loop of similar bending energy. Such a loop may be obtained by considering the optimal configuration of a loop of contour length $(L + 2r)$ instead of L , as shown in Fig. 4 (bottom, inset). The choice of the factor $2r$ derives from the following geometrical considerations: the closed saddle point configuration has an initial angle $\theta(0) = 49.2^\circ$, the ΔL closing the loop could be calculated for each value of r by requiring

$$\int_0^{\Delta L} \cos \theta(s) ds = r. \quad (17)$$

The angle $\theta(s)$ increases slightly on the first part of the trajectory $\theta(s) > \theta(0)$ and $\Delta L > 1.53 r$. We have chosen $\Delta L = 2r$ as an approximate value, this approximation has the advantage that it can be directly put in the fluctuation expression $C(r, L) \approx C_{\text{SY}}(L + 2r)$ in equation (15) to obtain:

$$Q(r, L) = C_{\text{SY}}(L + 2r) \exp[-\beta \Delta E(r, L)] \quad (18)$$

where ΔE is given in formula (14) and C_{SY} is given in formula (16). The validity of this approximation for the fluctuations prefactor was checked out by comparing $J(r, L)$ obtained from $Q(r, L)$ through formula (8), with the numerical results. The good agreement shown in Fig. 4 allows to obtain a semi-analytical formula for the loop probability with a finite interacting volume, which is valid for molecules of up to 2 kb (kilo base pairs).

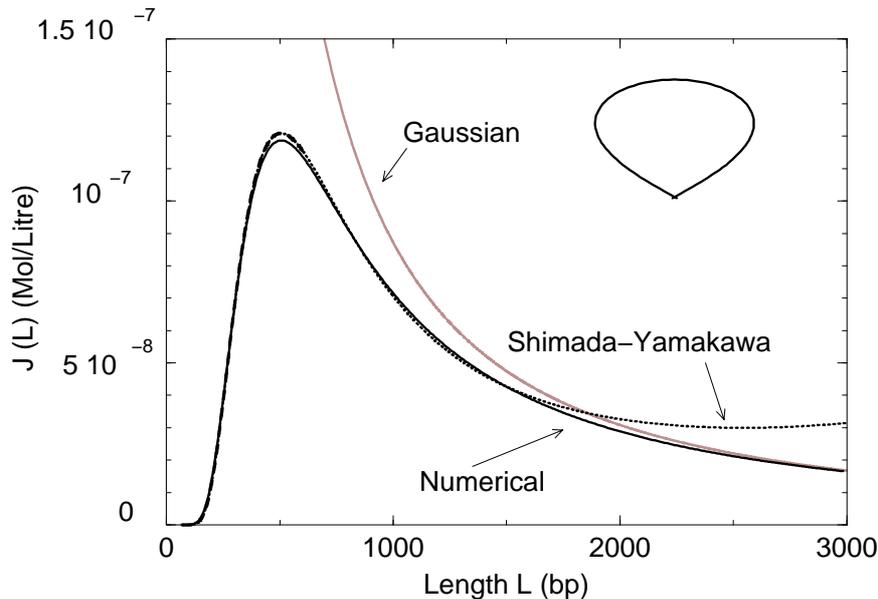


FIG. 3: Cyclization factor as a function of the DNA length with: the gaussian model (gray line); the WLC model with the SHIMADA and YAMAKAWA formula (dotted black line); the WLC model with the numerical calculation (full black line). The most probable length is 500 bp. Inset: the lowest bending energy configuration of a closed loop.

III. RESULTS: THE EFFECT OF A PROTEIN BRIDGE

In Fig. 3 we show the cyclization factor $J(L)$ for lengths L up to 3 kb. The most probable loop length is of about $L_* = 500$ bp that is $L_*/A \approx 3.5$ [15, 17]. Both shorter and longer cyclized lengths are less probable: stiffness makes difficult the bending of smaller polymers while entropy makes longer polymers ends unlikely to meet. The numerical cyclization factor is compared with the saddle point calculation of SHIMADA and YAMAKAWA [15] and with the gaussian polymer (GP) model. The former is in good agreement with the numerical results for lengths smaller than about 1.5 kb while the latter works well for loops larger than 2.5 kb.

The effects of the finite size r of the protein bridge are displayed in Fig. 4 which shows $J(r, L)$ (on the left) and $\Delta G(r, L)$ (on the right) for lengths L ranging from 75 bp (*i.e.* 25 nm) to 300 bp (*i.e.* 100 nm) and r respectively ranging from 1 nm to 10 nm. The numerical results (on the top) are in good agreement with the SPA results (on the bottom). Note that for small lengths L , $J(r, L)$ has a peak for the $L \approx r$ event corresponding to rigid rod-like configurations. Fig. 4 does not show this peak, occurring for $r = 10$ nm at about $L = 30$ bp (or $L = 10$ nm) because we focus on the cyclisation events. As shown in this figure r values ranging from 1 nm to 10 nm make no difference for the r -dependent closure factor, considering contour lengths L larger than 300 bp (or 100 nm). The cyclization factor $J(L)$ is evaluated as the limit $r \rightarrow 0$ of $J(r, L)$. In practice convergence is reached as soon as r is about one order of magnitude smaller than L . In the range $L > 75$ bp (25 nm) $J(5$ nm, 150 bp) converges to $J(1$ nm, 150 bp). On the other hand $J(r, L)$ is considerably different from $J(L)$ when r is of the same order of magnitude than L . For example for loops of $L = 100$ bp (34 nm) an end-to-end extension of 10 nm increases by two orders of magnitude the closure factor. Therefore proteins of size ≈ 10 nm are expected to produce drastic enhancements in looping short DNA sequences.

In terms of energetics (see Fig. 4, right) cycling a 100 bp DNA sequence costs $25 k_B T$ when the loop ends are required to stay within a sphere of radius $r = 1$ nm. This cost decreases to $13 k_B T$ if the sphere has the typical protein size of $r = 10$ nm. For loop lengths larger than 300 bp the only difference in the three curves of Fig. 4 (right), is a free energy shift due to the difference in the reacting volume. For instance, a reaction radius of 10 nm decreases the looping free energy of $3 \times \ln(10) \approx 7 k_B T$ with respect to a 1 nm reaction radius.

Our results for the closure factor and the looping free energy are compared in Fig. 4 with the Monte Carlo (MC) and Brownian Dynamics (BD) simulations results, obtained respectively by PODTELEZHNIKOV and VOLOGODSKII [16] (shown in the figure with filled circles, \bullet) and LANGOWSKI and *al.* [17, 18] (displayed in the figure with empty squares, \square). Numerical and SPA data (that for $r = 1$ nm converge to the SHIMADA and YAMAKAWA curve) are in better agreement with the MC data than with the BD ones; indeed numerical, SPA and MC data are obtained

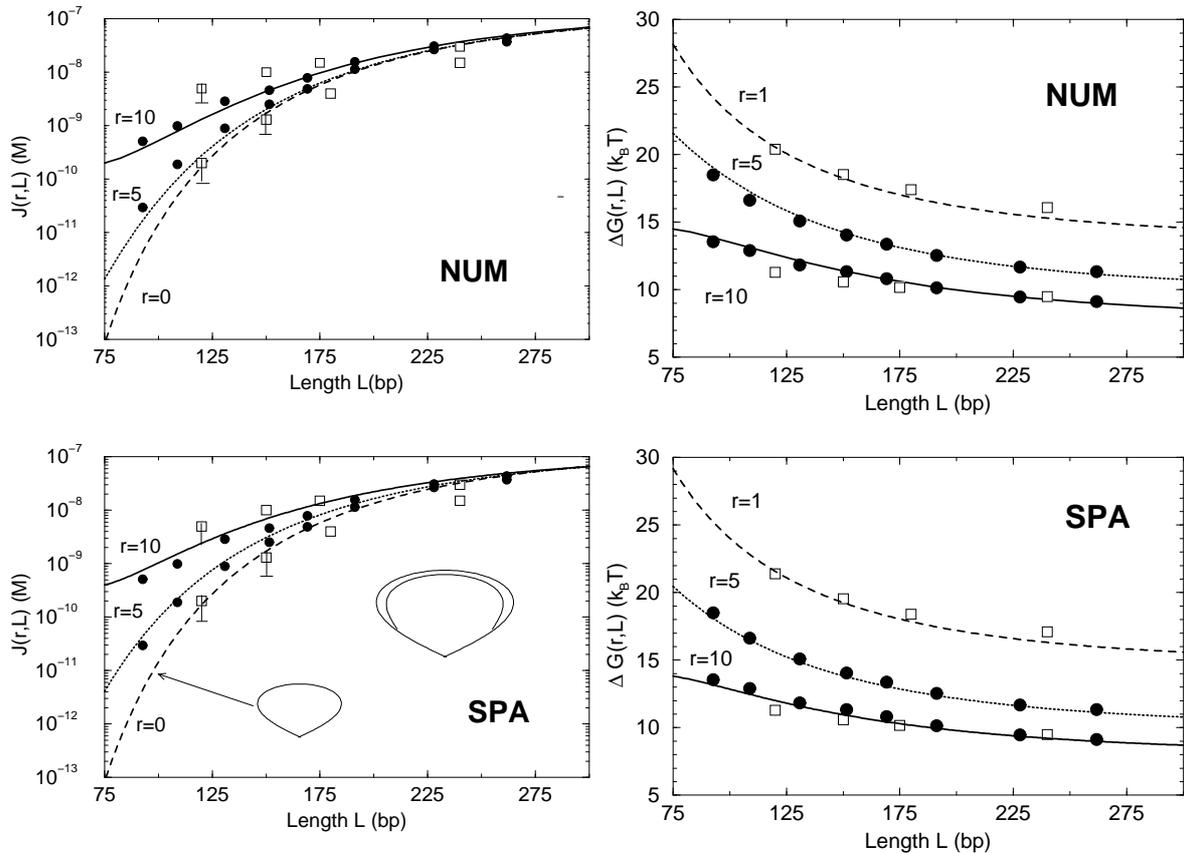


FIG. 4: Closure factor (left panel) and free energy (right panel) with a protein bridge of sizes: $r = 1$ nm (dashed line); 5 nm (dotted line); 10 nm (full line). The error bars are shown when they are larger than the symbol sizes. Top: numerical calculation of the constrained path integral. The $r = 1$ nm curve coincides with the $r = 0$ cyclization factor. Bottom: closure factor obtained by the extension of the SHIMADA and YAMAKAWA calculation, that includes r . Theoretical results are in very good agreement with Monte Carlo simulations obtained by PODTELEZHNIKOV and VOLOGODSKII (filled circle, \bullet) [16] and in fair agreement with brownian dynamics simulations obtained by LANGOWSKI and *al.* (empty square, \square) [17, 18]. Inset of the bottom left panel: lowest bending energy configurations for 100 bp and $r = 0$ (bottom) or $r = 10$ nm (top), the closure of the $r = 10$ nm configuration is shown by a thin line.

with a simpler model than BD ones, which does not include twist rigidity and electrostatic effects.

Considerations about Lac operon repression energetics will help us illustrating our results and compare them with other previous results. Expression of proteins enabling bacteria *E. Coli* to perform the lactose metabolism can be prevented at the transcriptional level by cycling two different sequences [4], the smallest one including the operon promoter. Let $O_1O_3 = 76$ bp and $O_1O_2 = 385$ bp denote these two resulting DNA loops where the so-called operators $O_{1,2,3}$ actually are the small specific DNA sequences (10 bp or so) at which the tetrameric repressor protein LacR can bind thus clamping the desired loop. Notice both processes are needed for efficient repression: despite O_1O_3 contains the operon promoter its cyclization is much less probable to occur than the O_1O_2 one (see cyclization factor $J(L)$ in Fig. 3). The LacR size is estimated from its crystallized structure to $r \approx 13$ nm [3]. In the *in vitro* experiments many parameters are under control among which the operators sequence and location. The distance between the two operators, that defines the length of the DNA loop has been fixed in [4, 5] to about 100 bp. Our results are in good agreement with the experimental measured stability of a 114 bp DNA loop mediated by a LacR protein, obtained by BRENOWITZ and *al.* in 1991 [3]. By measuring the proportion of looped complexes present in a solution with respect to the unlooped molecules they obtained a looping free energy of $20.3 \pm 0.3 k_B T$ to which they associated a closure factor of $8 \cdot 10^{-10}$ M. From Fig. 4, the closure factor of a loop of 114 bp with a protein bridge of $r = 10$ nm is $J(10 \text{ nm}, 114 \text{ bp}) = 10^{-9}$ M to which we associate, from formula (9) a cyclization free energy $\Delta G(10 \text{ nm}, 114 \text{ bp}) = 12 k_B T$. Note that the very good agreement between the closure factor contrasts with the bad agreement for the cyclization free energy. The latter could have been calculated considering a different reaction volume or it could also include the competition with

configurations that do not allow the formation of a loop (see Fig. 2 of [3]). To explain the high value found for the closure factor BRENOWITZ and *al.* already included the size of the protein in the analysis of their results by comparing their J result with the value expected for the cyclization probability of a free DNA when the length of the protein is included in the size of the loop.

Another result on DNA loop mediated by LacR protein has been obtained by BALAEFF and *al.* [13] who have numerically calculated the elastic energy of the O_1O_3 loop from a WLC model also including: the twist rigidity, the short range electrostatic repulsion and the details of the LacR/DNA complex crystal structure. The elastic energy is estimated to $23 k_B T$ in [13], of which 81% (that is $18 k_B T$) due to the bending and 19% due to the unwinding. The bending energy of $18 k_B T$ is to be compared to the saddle point energy (14) of $15 k_B T$ for a 75 bp loop with an end-to-end extension of $r = 10$ nm. The corresponding free energy of loop formation obtained by (18) and after integration over the reacting volume (9) is $\Delta G(10 \text{ nm}, 75 \text{ bp}) = 14 k_B T$ (see Fig. 4, bottom/right). For a 400 bp loop since LacR is only $\approx 10\%$ of the loop length, its size is expected to play a less important role. Indeed the cyclization probability does not depend on $r < 10$ nm and the free energy of forming such a loop decreases from $15 k_B T$ for $r = 1$ nm to $8 k_B T$ for $r = 10$ nm only because the reaction volume increases by a factor $3 \times \ln(10) \approx 7 k_B T$.

IV. EFFECT OF THE PRESENCE OF A KINK

The previous protein mediated DNA looping modelization (section III) assumes that the only intervening proteins are the ones clamping the loop ends (*e.g.*, the Lac operon repressor LacR). Actually regulation phenomena involve several proteins which may bind along the whole DNA. Indeed naked DNA situations barely exist *in vivo*. For instance single molecule manipulations [5] have shown that efficient Gal operon repression needs a stiffness loss of the 113 bp DNA portion to be looped. The HU protein produces such loss by kinking the sequence.

A. Numerical Calculation of J for a Bridged and Kinked Loop

Such stiffness loss may be taken into account in an effective way by kinking the WLC at half-length $L/2$. Let us call θ_- and θ_+ the angles of the DNA just before and after the kink respectively. We assume that the kink plane is vertical and choose it to define the origin of the azimuthal angle ϕ : $\phi_- = \phi_+ = 0$. Using the quantum language of section II A, we replace the calculation of the evolution operator $Z(L, f)$ (11) with its kinked counterpart

$$\begin{aligned} Z_{\text{kinked}}(L, f) &= \langle \text{final} | \exp \left[-\frac{L}{2A} \hat{H}(f) \right] | \theta_+, \phi_+ \rangle \times \langle \theta_-, \phi_- | \exp \left[-\frac{L}{2A} \hat{H}(f) \right] | \text{initial} \rangle \\ &= \sum_{\ell, \ell'} \langle 0, 0 | \exp \left[-\frac{L}{2A} \hat{H}(f) \right] | \ell, 0 \rangle \langle \ell', 0 | \exp \left[-\frac{L}{2A} \hat{H}(f) \right] | 0, 0 \rangle \times Y_{\ell}^0(\theta_+, 0) Y_{\ell'}^0(\theta_-, 0) \end{aligned} \quad (19)$$

where we have used the change of basis from angles to spherical harmonics

$$|\theta_{\pm}, \phi_{\pm} = 0\rangle = \sum_{\ell \geq 0} Y_{\ell}^0(\theta_{\pm}, 0) |\ell, 0\rangle. \quad (20)$$

Although calculations are a little bit more involved the evaluation scheme for the cyclization factor J remains unchanged in its principle (section II A). We now have “ $|\ell \neq 0, 0\rangle$ elements” corresponding to particular orientations arriving at ($-$) and leaving from ($+$) $s = L/2$. The kink angle κ is intuitively defined from these WLC tangent vectors at half-length (in spherical coordinates) through

$$\kappa = \theta_+ - \theta_-. \quad (21)$$

B. Saddle Point Approximation of J for a Bridged and Kinked Loop

The saddle-point calculation of section II B can be straightforwardly extended to the case of a kinked loop. In Fig. 5 we show the configurations with the lowest bending energy for a 100 bp DNA loop with an end-to-end extension $r = 10$ nm. The kink is accounted for by a bending angle in the middle of the chain $\kappa = 2 \times \theta(L/2) - \pi$ *a priori* different

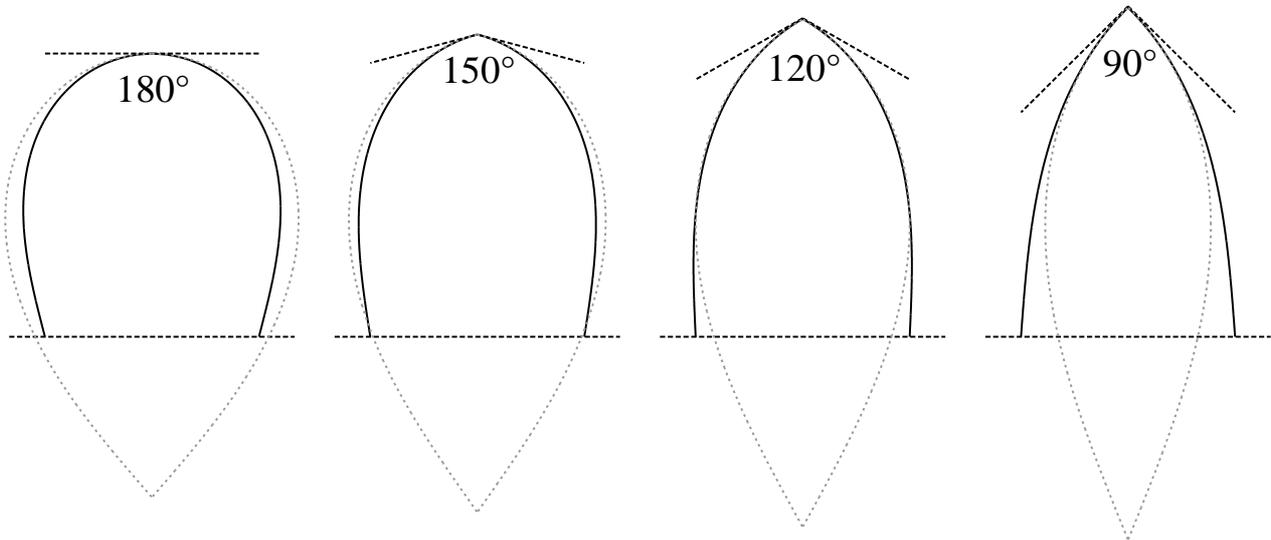


FIG. 5: Lowest bending energy configurations for a 100 bp DNA loop with an end-to-end extension $r = 10$ nm and kinks $\kappa = 90^\circ$, 120° , 150° , 180° respectively. The gray configurations are the closed loops used in the calculation of the fluctuation.

from the previous trivial value $\kappa = \pi$, ranging from 150° to 90° . We introduce the phase $\psi = \arcsin[\sin(\frac{\pi-\kappa}{4})/x]$. The parameter x is now obtained from equation

$$(1 + r/L) \left[\widehat{K}(x^2) - K(\psi, x^2) \right] = 2 \left[\widehat{E}(x^2) - E(\psi, x^2) \right] \quad (22)$$

and the total bending energy is

$$\beta\Delta E(r, L, \kappa) = \frac{4}{L/A} \left[\widehat{K}(x^2) - K(\psi, x^2) \right]^2 (2x^2 - 1 + r/L). \quad (23)$$

In analogy with (15) we obtain the end-to-end extension \vec{r} PDF

$$Q(r, L, \kappa) = C_{\text{SY}}(L + 2r) \exp[-\beta\Delta E(r, L, \kappa)], \quad (24)$$

where C_{SY} is given in (16), from which we calculate $J(r, L, \kappa)$ through formula (8). Note that (24) reduces to the loop probability given in [23] for a closed and kinked DNA. Again the good agreement obtained with numerical results allows us to establish a semi-analytical formula for the loop probability with a finite interacting volume and kink.

C. Results for a Kinked and Bridged Loop

In Fig. 6 results for the looping probability density (left) and free energy (right) are shown for a typical end-to-end extension $r = 10$ nm and kinks $\kappa = 150^\circ$, 120° , 90° . The curve with no kink ($\kappa = 180^\circ$) is also shown for comparison. The numerical results on the top of the figure are in very good agreement with the extension of the SHIMADA and YAMAKAWA saddle point approach on the bottom of the figure (black lines). In Fig. 6 (inset) we show the results obtained by BD simulations by LANGOWSKI and *al.* [17, 18] for $J(10 \text{ nm}, L, \kappa)$ (empty squares, \square) fitted by RIPPE [19] with a simple formula containing one fitting parameter for each curve. The curves in the inset of Fig. 6 reproduce the same behavior with L and κ of our numerical (top of Fig. 6) or SPA curves (bottom of Fig. 6). The numerical gap with BD results increases for large kinks: at $\kappa = 90^\circ$ our closure factor is ten times larger than the closure factor obtained by BD simulations. The electrostatic and twist rigidity effect could indeed play a more important role when the chain is kinked. Note that the lengths range of the numerical results are from 100 bp to 500 bp, while the lengths range of the saddle point results is from 75 bp to 1500 bp. The lengths range of the SPA is limited by the validity

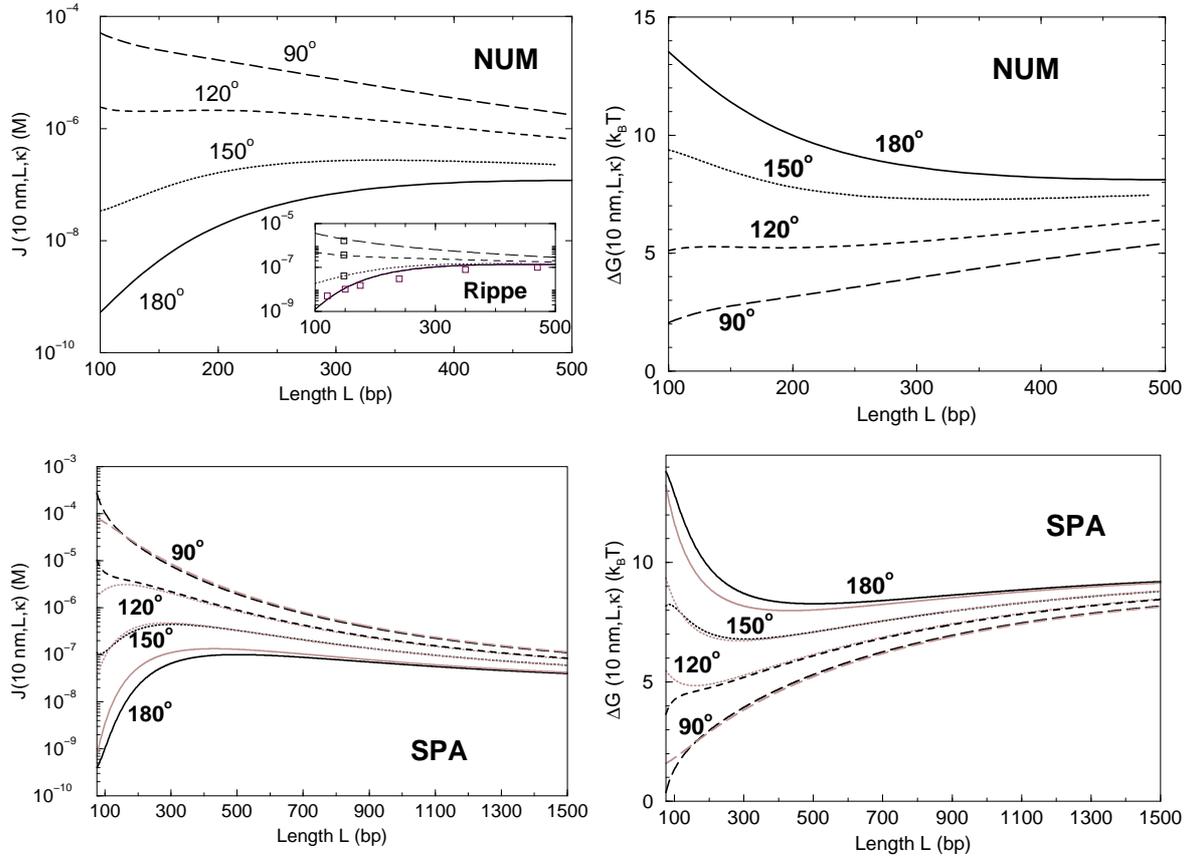


FIG. 6: Closure factor (left) and free energy for a loop with $r = 10$ nm and a kink κ in the middle of the chain: $\kappa = 180^\circ$ (full lines); $\kappa = 150^\circ$ (dotted lines); $\kappa = 120^\circ$ (dashed line); $\kappa = 90^\circ$ (long dashed lines). Top: numerical calculation of the constrained path integral. Bottom: extension of the SHIMADA and YAMAKAWA calculation (black lines) and approximate formula (gray line) given in the text (25). Numerical results are in very good agreement with the SPA approximation and the approximate formula (25). Inset: brownian dynamics simulations point obtained by the LANGOWSKI and collaborators (empty square, \square) [17, 18], fitted by a simple formula by RIPPE [19].

of the approximation ≈ 1500 bp shown in Fig. 3. As an example the closure factor $J(10 \text{ nm}, 113 \text{ bp}, \kappa)$ of a 113 bp fragment, obtained by the numerical calculations, increases from the value of 10^{-9} M for a non kinked loop ($\kappa = 180^\circ$) to $4 \cdot 10^{-8}$ M, $2 \cdot 10^{-6}$ M and $4 \cdot 10^{-5}$ M for respectively $\kappa = 150^\circ$, 120° and 90° . The corresponding looping free energy is $\Delta G(10 \text{ nm}, 113 \text{ bp}, \kappa) = 13 k_B T$, $9 k_B T$, $5 k_B T$ and $2.5 k_B T$ for $\kappa = 180^\circ$, 150° , 120° and 90° respectively. With the saddle point approach we find similar results: $J(10 \text{ nm}, 113 \text{ bp}, \kappa) = 2 \cdot 10^{-9}$ M, 10^{-7} M, $5 \cdot 10^{-6}$ M and $8 \cdot 10^{-5}$ M while $\Delta G(10 \text{ nm}, 113 \text{ bp}, \kappa) = 11 k_B T$, $8 k_B T$, $4.4 k_B T$ and $1.7 k_B T$. As it is shown in Fig. 6 the presence of the kink become irrelevant for DNA segment larger than about 1500 bp. The example of a 113 bp DNA segment has been chosen to compare the results with the single molecule experiments on the GalR mediated loop of an ≈ 113 bp DNA portion between the two operators. From the kinetics of the loop formation mediated by the GalR and HU proteins, LIA and *al.* [5] have deduced a looping free energy of $12 k_B T$, that with respect to our values should correspond to a kink angle of more than 150° or to an end-to-end extension smaller than $r = 10$ nm. Another significant change in the cyclization probability is that the stiffness loss induced by κ reduces the most probable loop length from 500 bp for a non kinked DNA to 340 bp and 190 bp (from the numerical calculation) or to 300 bp and 150 bp (from SPA) for kinks of respectively 150° and 120° . Note that for a kinked loop with an end-to-end extension r the minimal length L_0 corresponding to the rigid rod-like configuration fulfills the relation $L_0 \sin(\kappa/2) = r$ and therefore it is of ≈ 42 bp for $\kappa = 90^\circ$ instead of ≈ 30 bp for $\kappa = 180^\circ$. for $\kappa = 90^\circ$ the most probable loop length is the rigid *kinked* rod-like configuration of the two half-DNA portions. To catch both kink and protein bridge effects in a simple formula, we have calculated the cyclization factor with the extension of SHIMADA and YAMAKAWA formula for a kinked closed loop of length $(L + 2r)$. This approach is similar to what was suggested in 1991 by BRENOWITZ and *al.* [3] to interpret their experimental data, *i.e.* to directly consider the protein as part of the length of the loop. A linear

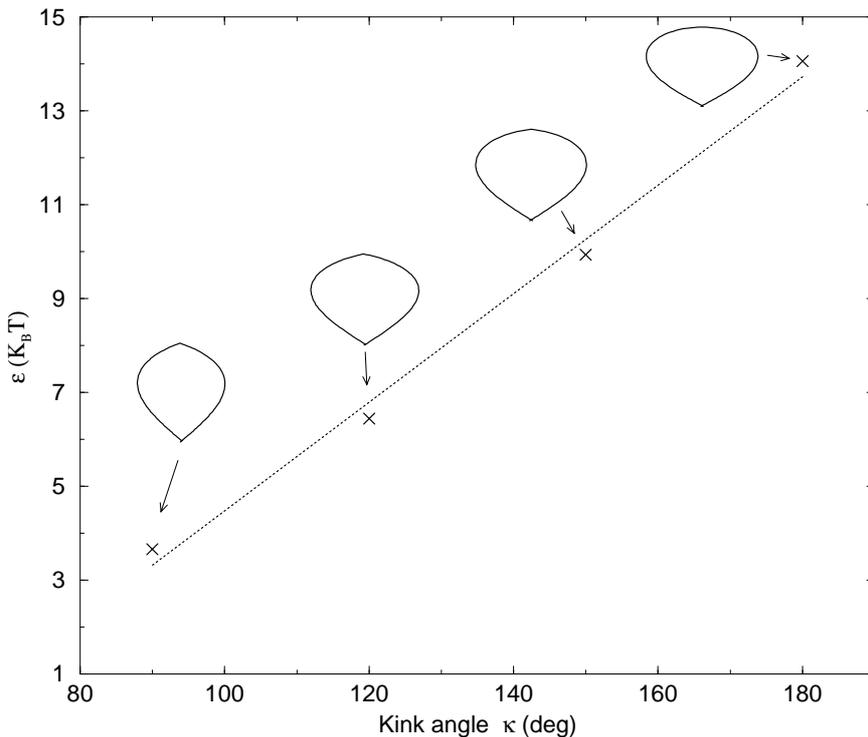


FIG. 7: The “x” points: saddle point energy in units of L/A (that is $\varepsilon = \Delta E \times L/A$) for the saddle point configurations (also displayed in the figure) with kink angles of $\kappa = 90^\circ, 120^\circ, 150^\circ, 180^\circ$. Dotted line: linear interpolation used in formula (25).

fit [26] of the bending energy (23) for the optimal closed configuration ($r = 0$) in the presence of a kink (expressed in degrees): $\beta\Delta E(r = 0, L, \kappa) \approx (-7.1 + 0.1155 \kappa)/(L/A)$, is shown in Fig. 7. It gives the following approximated formula for the closure factor as a function of the protein size r , the length L of the DNA, and the kink angle κ

$$J_{\text{approx}}(L, r, \kappa) = C_{\text{SY}}(L + 2r) \exp\left[\frac{7.1 - 0.1155 \kappa}{(L + 2r)/A}\right] \quad (25)$$

where C_{SY} is given in (16). Notice the integration step (8) has been skipped since it does not make any significant difference. Formula (25) allows us to obtain a simple prediction for the loop probability in presence of a kink in the middle of the sequence and a finite separation between the extremities. As shown in Fig. 6 this formula (gray lines) is in good agreement with the loop probability obtained with the exact calculation of the saddle point energy of the open configuration (full line). In particular Fig. 6 shows that for kink angles in the range $90^\circ < \kappa < 150^\circ$ this simple formula works remarkably well for lengths L larger than about $5r$, that is 150 bp for $r = 10$ nm. For smaller lengths the optimal configuration is more a rigid rod-like and cannot be approximated by a closed loop. Similar simple formulas that includes a kink angle κ and a finite end-to-end distance r in an effective way have also been written down by RIPPE or RINGROSE to fit their brownian dynamics simulation [19] or experimental data [29], but these formulas contain a parameter that must be fitted for each values of r and κ from the data points (Fig. 6).

V. CONCLUSION

We performed both numerical and analytical calculations of the closure factor J , even in the presence of a protein bridge and of a protein-mediated kink. More precisely we have numerically calculated the path integral of the WLC polymer model under the constraints of a fixed end-to-end distance r and a kink κ in the middle of the DNA portion. Analytically we have extended the SHIMADA and YAMAKAWA saddle point approximation [15] to the case of a bridged and kinked loop. We have seen that the formation of DNA loops is significantly sensitive to the size of the protein bridge when this size r is more than 10% of the loop length L , that is 300 bp (or 100 nm) for a typical protein bridge size of $r = 10$ nm. To give an example, the closure factor for a 100 bp DNA segment increases from $J(100 \text{ bp}, 0) \approx 10^{-11}$ M to $J(100 \text{ bp}, 10 \text{ nm}) \approx 10^{-9}$ M. Correspondingly, looping free energy decreases

from $\Delta G(100 \text{ bp}, 0) = 24 k_B T$ to $\Delta G(100 \text{ bp}, 10 \text{ nm}) = 13 k_B T$. A kink ranging from 150° to 90° produces a significant increase of J for DNA fragments of lengths up to about 2500 bp. For instance the closure factor for a 100 bp DNA segment with a protein bridge $r = 10 \text{ nm}$ and a kink of 90° is $J(100 \text{ bp}, 10 \text{ nm}, 90^\circ) \approx 10^{-4} \text{ M}$, and the corresponding looping free energy $\Delta G(100 \text{ bp}, 10 \text{ nm}, 90^\circ) \approx 2 k_B T$. A kink also changes the most probable loop length from 500 bp (no kink) to about 175 bp for a kink of 120° , going to the rigid *kinked* rod-like configuration for smaller κ values. This is an interesting mechanism because the loop lengths implied in *in vivo* DNA processing by proteins spread on a large lengths range. Our results were compared to previous analytical approximations (in particular the results of the gaussian model, the WILHELM and FREY expansion [14] and the SHIMADA and YAMAKAWA formula [15]) and numerical calculations (in particular the Monte Carlo simulations data obtained by PODTELEZHNIKOV and VOLOGODSKII [16], and the brownian dynamics simulations data obtained by LANGOWSKI and *al.* [17, 18, 19]) as well as experimental results [3, 4, 5] (*e.g.*, the ones obtained by LIA and *al.* on the looping dynamics mediated by the Gal and HU proteins). Finally a simple formula (25) including both the protein bridge and kink effects has been proposed. This formula has the advantage of not containing adjustable parameters with respect to the existing formulas that include both these effects [19].

Still many effects omitted in this work can be included without significant changes in the numerical algorithm. The kink we considered is actually permanent (that is not thermally excited), site specific (at half-length) and rigid (κ fixed). Although this rigidity seems relevant to most protein bindings to DNA at first glance [30], it was pointed out in YAN and *al.* works [20, 22], kinks may also be semi-flexible (exhibiting higher or lower rigidities than the bare DNA) or even fully flexible [21, 31]. For instance, the HU/DNA complex was recently observed to be very flexible under specific experimental conditions [32]. Flexible hinges were also stated to occur along the DNA due to the opening of small denaturation bubbles [21, 33], such as the one needed by HU to fit in the double helix [5]. Such flexibility could be taken into account [30] in our model. This kind of defects could also be thermally activated, occurring at multiple non-specific sites along the DNA [20, 22, 30, 31, 34]. Both effects may be included in our model. Note that using effective persistence lengths could turn out convenient, despite these inform little about the kink properties (number, location, rigidity, etc). Actually this would be equivalent to study DNA stiffness loss due to sequence effects [35] by cutting WLC in different stiff fragments, depending on the CG or AT content of the whole sequence to cyclize. The same approach may allow an approximative study of the DNA polyelectrolyte nature too [36]. Otherwise electrostatic potential has to be included in WLC energy. Twist elasticity leads to slight modifications of the quantum analog we used although requiring some care [37]. This is expected to play an important role in looping, especially when specific alignment of the loop extremities are required. Finally cyclization dynamics could be modeled using a simple two states model where DNA is “closed” or “opened”, that is cyclized or not. Such study relies on the (statics) cyclization factor we computed in this article [38]. Direct comparison to experimental lifetimes measures would be possible [4, 5].

Acknowledgments

The authors would like to thank D. CHATENAY, R. MONASSON and F. THALMANN for useful discussions and a critical reading of the manuscript.

-
- [1] D. SHORE, J. LANGOWSKI, and R. L. BALDWIN, PNAS **78**, 4833 (1981).
 - [2] Q. DU, C. SMITH, N. SHIFFELDRIM, M. VOLOGODSKAIA, and A. VOLOGODSKII, PNAS **102**, 5397 (2005).
 - [3] M. BRENOWITZ, A. PICKAR, and E. JAMISON, Biochemistry **30**, 5986 (1991).
 - [4] L. FINZI, and J. GELLES, Science **267**, 378 (1995).
 - [5] G. LIA, D. BENSIMON, V. CROQUETTE, J.-F. ALLEMAND, D. DUNLAP, D. E. A. LEWIS, S. ADHYA, and L. FINZI, PNAS **100**, 11373 (2003).
 - [6] J.-F. ALLEMAND, S. COCCO, N. DOUARCHE, and G. LIA, Eur. Phys. J. B (to be published).
 - [7] O. KRATKY, and G. POROD, Rec. Trav. Chim. **68**, 1106 (1949).
 - [8] H. YAMAKAWA, *Helical Wormlike Chains in Polymer Solutions* (Springer-Verlag, Berlin, 1997).
 - [9] H. KLEINERT, *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets* (World Scientific Publishing Co., Singapore, 2004).
 - [10] H. E. DANIELS, Proc. Roy. Soc. Edinburgh A **63**, 290 (1952).
 - [11] W. GOBUSH, H. YAMAKAWA, W. H. STOCKMAYER, and W. S. MAGEE, J. Chem. Phys. **57**, 2839 (1972).
 - [12] A. HANKE, and R. METZLER, Biophys. J. **85**, 167 (2003).
 - [13] A. BALAEFF, L. MAHADEVAN, and K. SCHULTEN, Phys. Rev. Lett. **83**, 4900 (1999).
 - [14] J. WILHELM and E. FREY, Phys. Rev. Lett. **77**, 2581 (1996).
 - [15] J. SHIMADA, and H. YAMAKAWA, Macromolecules **17**, 689 (1984).
 - [16] A. A. PODTELEZHNIKOV, and A. V. VOLOGODSKII, Macromolecules **33**, 2767 (2000).

- [17] K. RIPPE, P. H. VON HIPPEL, and J. LANGOWSKI, Trends Biochem. Sci. **20**, 500 (1995).
- [18] H. MERLITZ, K. RIPPE, K. V. KLENIN, and J. LANGOWSKI, Biophys. J. **74**, 773 (1998).
- [19] K. RIPPE, Trends Biochem. Sci. **26**, 733 (2001).
- [20] J. YAN, and J. F. MARKO, Phys. Rev. E **68**, 011905 (2003).
- [21] J. YAN, and J. F. MARKO, Phys. Rev. Lett. **93**, 108108 (2004).
- [22] J. YAN, R. KAWAMURA, and J. F. MARKO, Phys. Rev. E **71**, 061905 (2005).
- [23] S. SANKARARAMAN, and J. F. MARKO, Phys. Rev. E **71**, 021911 (2005).
- [24] J. SAMUEL, and S. SINHA, Phys. Rev. E **66**, 050801(R) (2002).
- [25] R. B. SIDJE, ACM Trans. Math. Softw. **24**, 130 (1998). The Expokit library is available for download at URL <http://www.maths.uq.edu.au/expokit>.
- [26] W. H. PRESS, S. A. TEUKOLSKY, W. T. VETTERLING, and B. P. FLANNERY, *Numerical Recipes in C: the Art of Scientific Computing* (Cambridge University Press, 1992).
- [27] C. MOLER, and C. van LOAN, SIAM Review **45**, 3 (2003).
- [28] M. ABRAMOWITZ, and I. A. STEGUN, *Handbook of mathematical functions* (Dover Publications Inc., New York, 1970).
The elliptic integrals of first and second kinds respectively read in LEGENDRE definition $K(\psi, x^2) = \int_0^\psi \frac{dy}{\sqrt{1-x^2 \sin^2 y}}$
and $E(\psi, x^2) = \int_0^\psi \sqrt{1-x^2 \sin^2 y} dy$.
- [29] L. RINGROSE, S. CHABANIS, P.-O. ANGRAND, C. WOODROOFE, and A. F. STEWART, EMBO J. **18**, 6630 (1999).
- [30] Y. O. POPOV, and A. V. TKACHENKO, Phys. Rev. E **71**, 051905 (2005).
- [31] P. A. WIGGINS, R. PHILLIPS, and P. C. NELSON, Phys. Rev. E **71**, 021909 (2005).
- [32] J. V. NOORT, S. VERBRUGGE, N. GOOSEN, C. DEKKER, and R. T. DAME, PNAS **101**, 6969 (2004).
- [33] P. RANJITH, P. B. SUNIL KUMAR, and G. I. MENON, Phys. Rev. Lett. **94**, 138102 (2005).
- [34] B. CHAKRABARTI, and A. J. LEVINE, Phys. Rev. E **71**, 031905 (2005).
- [35] P. NELSON, Phys. Rev. Lett. **80**, 5810 (1998).
- [36] R. ZANDI, J. RUDNICK, and R. GOLESTANIAN, Phys. Rev. E **67**, 021803 (2003).
- [37] C. BOUCHIAT, and M. MÉZARD Phys. Rev. Lett. **80**, 1556 (1998); V. ROSSETTO, and A. C. MAGGS, *ibid.* **88**, 089801 (2002) [comment]; C. BOUCHIAT, and M. MÉZARD, *ibid.* **88**, 089802 (2002) [reply].
- [38] S. JUN, J. BECHHOEFER, and B.-Y. HA, Europhys. Lett. **64**, 420 (2003).

Loops in DNA: an Overview of Experimental and Theoretical Approaches

J.-F. Allemand, S. Cocco, N. Douarche, G. Lia

CNRS – Laboratoire de Physique Statistique de l'ENS, 24 rue Lhomond, 75005 Paris, France

DNA loop formation plays a central role in many cellular processes. The aim of this paper is to present the state of the art and open problems regarding the experimental and theoretical approaches to DNA looping. A particular attention is devoted to the effects of the protein bridge size and of protein induced sharp DNA bending on DNA loop formation enhancement.

I. IMPORTANCE OF LOOP FORMATION IN BIOLOGICAL PROCESSES

DNA processing is not always the result of a single protein going to a single target sequence. It can also be the result of a single or multimeric proteins connecting two distant sites on DNA thus forming a loop (see Fig. 1). Indeed protein mediated DNA loops play a central role in many biological processes of DNA life [1, 2].

Looping is widely used in gene regulation, in particular various genetic events of transcriptional regulation (RNA synthesis), including activation but also repression, can involve formation of such DNA structure. One of the most known example of transcription regulation which is accomplished through DNA looping is the regulation of the *gal* operon (see Fig. 1) [3]. In this genetic system transcription can be blocked by loop formation between two distinct sites of defined sequences distant of 113 base pairs (bp) that encompass promoter region. When a sugar, the galactose, is present in the surrounding medium, looping is inhibited, and transcription can occur. Other typical examples of gene regulation by looping are the *ara*, *lac* and *deo* operon in *E Coli*, the lysogenic-to-lytic switch in phage λ , the protein/DNA complex between c-Myb and C/EBP in the *mim-1* promoter, and the protein p53, to name just a few [4–9].

Nucleoprotein interaction resulting in a loop formation can also transiently occur between proteins and the RNA polymerase. In this case, looping regulates the quantity of RNA produced by RNA polymerase. Example are the NtrC system (nitrogen regulatory protein C) in *E Coli*, the Jun activator with the pre-initiation complex and the eukaryotic silencers and enhancers, like the human β -*globin* locus [10–13].

DNA looping in these processes may involve loop sizes from about fifty to several thousands of base pairs. The proteins that activate transcriptional factors are located 50-200 bp upstream the promoter while enhancers and silencers sequences are several thousands base pairs upstream or downstream of their promoters; the lambda repressor switch can act looping a sequence of roughly 2300 bp; the lactose repressing act by looping a 305 bp sequence between the two DNA binding sites called O_1 and O_2 or a sequence of 76 bp between O_1 and a third site O_3 ; the two binding site of the Gal repressor are at a distance of 113 bp.

Such a loop between two sequence specific sites is a common event in site specific recombination [14, 15], a process by which the DNA fragment between the two sites can be excised (or inserted in an inverse process used to transpose DNA fragments see Fig. 2). The most popular example is the Cre protein that recognizes two 24 bp sites, named loxP, makes loop between them and delete the DNA fragments between those sites. When the surrounding DNA is correctly designed this reaction allows switching on a gene. This is routinely used for artificially switch on or off *in vivo* a gene [16, 33].

Even the proteins used in DNA cloning, the type II restriction enzymes, that recognize a sequence and cut the DNA at a specific position in the same region, can be involved in DNA looping (see Fig. 2). Some of them, like SfiI or NgoMIV [17, 18], have an enhanced cutting activity when bound to two target sites (in the same or in two different DNA molecules) instead of just one. In these cases looping is likely to happen because of random collisioning.

DNA looping is then a very common phenomenon, that generally is generated passively. Nevertheless, loop can also be formed by more active processes.

In fact some motors use ATP to form loop (see Fig. 1). Type I restriction enzymes are a good example [19, 20]. Those enzymes bind to a single specific sequence and then start extruding one or two loops from this sequence until for still an unknown reason it cuts the DNA at some blocking point. As a consequence the loop size before the cutting event displays a large distribution.

At a higher level of organization chromosomes are also organised in topologically isolated loops (see Fig. 3). In *E Coli* the average size of these loops has been measured at about 10 kb (that is 1000 bp) [21]. In eukaryotes some structural proteins interact with chromatin (DNA compacted by histone proteins) organizing it in very large loop (between 30 and 140 kb) [36].

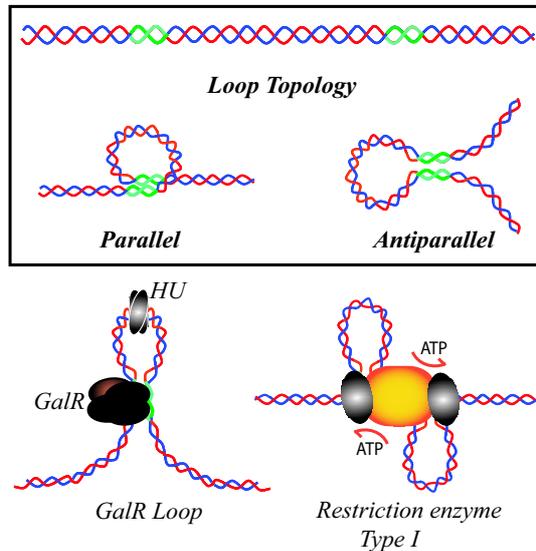


FIG. 1: Upper panel: looping between two distant sites can lead to different forms. Two major type of conformations are the parallel and antiparallel ones. Energy cost is different for every conformation. Angles may vary, local bends might be present inside the loop, phasing of the two helical pitch may also be a important parameter. Lower panel: In *E Coli* a tetramer of galR, in the presence of the structural protein HU, forms a loop between two sites distant of 113 bp. This loop inhibits the transcription of the gene whose origin of replication lies inside the loop. Type I restriction enzymes bind to a specific site and, using ATP as an energy source, extrude loops. Loop size before cutting depends on conditions. Initiation of the loop, that at first glance seems energy consumable, is still an unknown phenomenon.

II. EXPERIMENTAL APPROACHES

Experimentally various approaches has been developed to study DNA looping. Since loops will affect DNA shape it is possible to detect loop formation in gel electrophoresis (see for instance [23]). As it brings two sites in close distance to each other, loop formation can also be detected by fluorescence energy transfer (FRET) [24–26] between two fluorophores located close to the two sites involved in looping. Other bulk methods imply the use of additional proteins [27, 28, 32]. The most known method uses the ligase [27], an enzyme that is able to ligate two DNA extremities. To perform this operation it requires that the two extremities are placed at the same position, *i.e.* a loop is formed. In fact it also requires an angular phasing between the two extremities because of the requirement of having a number of helical pitches in agreement with the ~ 10.5 bp turn for the circular DNA. Recombination efficiency [28], detected by the efficiency of the excision of the DNA fragment between the two sites, can also be used in that purpose. These techniques, can be very sensitive but are subjected to protein concentration and activity [22]. Of course high resolution microscopy techniques, like electron microscopy or AFM [31], are usefull tools to image loops. Single molecule techniques [24, 34, 35] have also been applied to this problem. The use of these techniques made the real time observation of the looping possible, allowing informations about the kinetics of the looping. Moreover, with the micromanipulation techniques, a new control parameter is introduced : the force (see Fig. 4). Applying a force will disfavor the formation of the loop. So the higher the force the smaller the time spent in the closed loop conformation and the longer the time spent in the open loop form. In fact the mechanical constrains allow the modification of the kinetics barriers of the looping reaction without modifying the physico-chemistry parameters. Simple thermodynamics allows then to have access to some kinetics, thermodynamics and structural parameters [35].

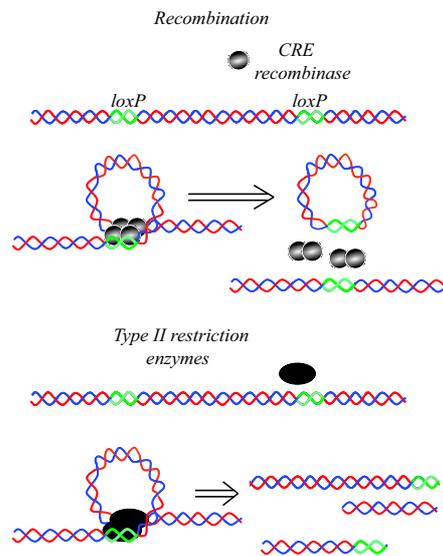


FIG. 2: Cre recombinase is one of the most well known of the site specific recombinases. This family of enzymes bind to two specific sites, the loxP sites (34 bp long) in the case of Cre, and excise the DNA in between after the formation of a transient loop, called synapse. Loop formation may enhance the activity of the well known type II restriction enzymes, the ones used in gene cloning to cut DNA at specific positions.

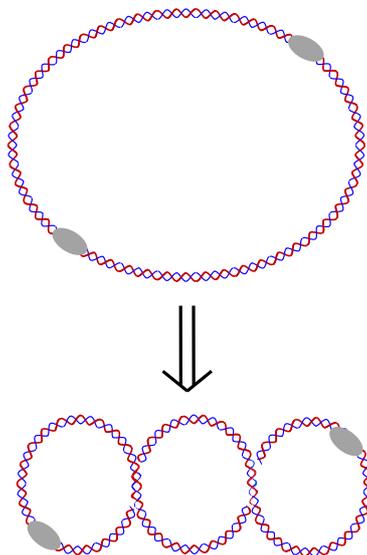


FIG. 3: DNA isolated from living organisms always present some torsional strain, *i.e.* torque. This is required for DNA processing and this strain is highly regulated by various enzymes in the cell. This torsional stress leads to the formation of structures called plectonemes. In the figure two turns have been added to the relaxed, *i.e.* without torsional stress, circular DNA drawn on top giving the lower form. These plectonemes will strongly affect the probability and kinetics of loop formation between two or more specific sites (bold ellipsoids). In this case superposition of the two sites, necessary for loop formation requires slithering.

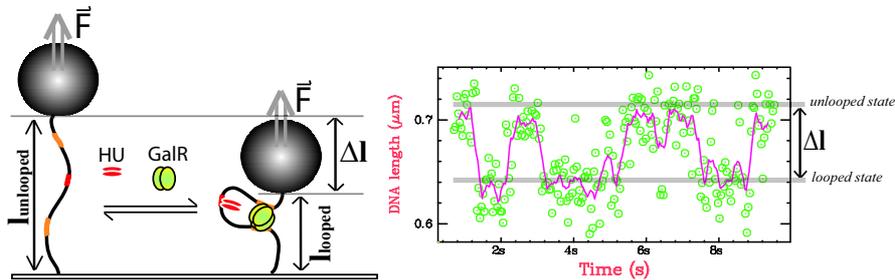


FIG. 4: Illustration of the experiment of Lia *et al.* A magnetic bead submitted to a magnetic field gradient is submitted to a force (left panel). The force will disfavor the closed loop form. If F is the force and δl the change in extension due to loop formation: when $F \times \delta l$ is equal to the energy of the loop formation then the proportion of the time spend in the closed or open loop are equal. This two level system then alternates between the two conformations as illustrated on the time trace on the right. Without the force the loop would remain in the closed state almost all the time. As described in [35] supercoiling (torsion) is important in this process. Experimental measured value gives $\delta l = 55$ nm. This distance is bigger than the distance between the sites incating an antiparallel conformation. The force where the loop is formed during 50% of the time is 0.9 pN thus leading to an energy of $-12 k_B T$ (where $k_B T$ is the thermal energy) for the loop formation in the experimental conditions. Measuring the average duration of each state leads to the kinetic constants k_{loop} and k_{unloop} . These depends on the force with an Arrhenius factor $\exp\left(-\frac{F \times \delta l}{k_B T}\right)$. Using this factor one may extrapolate the looped time to 21 s and the unlooped time 0.1 ms at zero force.

III. DNA LOOPING: THE PROBLEMATICS

The study of DNA loops has been an active field of research since the eighties. It was always characterized by a very strong interaction between experiments and theory.

The experimental and theoretical main goal is to calculate the probability that two sites of the same DNA molecule will meet by looping. For experimental convenience the looping probability is commonly expressed in terms of the local molar concentration of one site with respect to the other one (j , or j_M in $M = \text{mol} \cdot \text{L}^{-1}$). j_M is also related to the free energy of loop formation. The loop stability depends on the comparison of j_M with the dissociation constant of the protein-protein or protein-DNA interaction that stabilize the loop formation.

The difficulty for the DNA to make a loop are two-fold, and depend on the loop length. The formation of a small loop gives rise to an elastic problem because of the bending energy, while the formation of a large loop leads to an entropic problem because far appart sites have to come into collision.

A central point in the past studies of DNA looping has been to derive the probability distribution of loop formation as a function of the loop length. The most probable loop size for a given DNA molecule is encountered when the elastic energy starts to decrease (as the length increases) and the entropic loss is not too large yet; this corresponds to about 500 bp for a bare DNA. However, as we have underlined in the introduction, observed loop lengths *in vivo* spread over a large range of sizes. A central question is therefore to determine what kinds of protein action can modify the most probable loop length to make possible a given DNA processing.

IV. ON THE CYCLISATION FACTOR

A. Some Historical Background

Loop formation was first studied by Shore *et al.* through cyclisation or ring closure experiments [30]. These experiments are the simplest ones to analyse for two reasons. First, a linear DNA is entirely cyclized in the loop; secondly, the ligase proteins are not involved in the loop formation itself but only stabilize the loop once it is formed. In these experiments the loop length is varied from 130 to 4300 bp from the preparation of different linear DNA. These experiments have been first used to determined the elastic properties of DNA *e.g.* the persistence length, the twist modulus, and the pitch of the double helix.

From a theoretical point of view the theory of elasticity of a stiff chain was developed after the 1950s. The Wormlike Chain (WLC) of Kratky and Porod [41] turned out to be a very good model in this context, and was widely studied. The absence of exact analytical solution for the end-to-end distribution of the WLC model pushed workers in the field to develop various expansion approaches [42–50, 52, 53, 55].

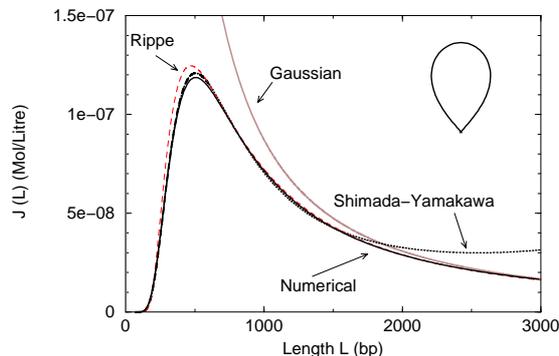


FIG. 5: Cyclization factor as a function of the length of DNA with the Gaussian model (gray line), WLC model and the Shimada-Yamakawa formula (dotted black line) [46], WLC model and a numerical calculation (full black line) [86], Rippe-Ringrose approximation (thin gray dashed) [28, 84]. The most probable length is 500 bp. Inset: the low bending energy configuration of a closed loop.

Shimada and Yamakawa applied in 1984 [46] the calculation of the end-to-end distribution of the WLC model to DNA cyclisation and compare with Shore *et al.* experiments. They derived a simple formula for j_M with numerical coefficients, which can be directly used by experimentalists. In the same period Shore *et al.*'s experimental results and Shimada and Yamakawa theoretical results have also been compared to Monte Carlo simulations by Hagermann *et al.* [62], Levene and Crothers *et al.* [27, 29, 63] and Frank-Kamenetskii *et al.* [64].

Very recently cyclisation experiments and theory have known a renewal of interest, stemming from Cloutier and Widom's results on an unexpectedly high cyclisation probability for small DNA molecules (94-116 bp) [77]. Though these experimental findings were merely an artifact due to an erroneous concentration of ligases, they have triggered a lively debate on the issue. Several theoretical works have been developed based on the idea that for short DNA the formation of a few base pairs of single-stranded DNA can facilitate loop formation; single-stranded (ss) DNA is indeed very flexible and therefore can provide a "flexible hinge". The WLC model has therefore been modified to include the possibility for formation of such thermally fluctuating bubbles [78-82]. In this context the sequence should play a crucial role [82]. The debate came to a conclusion with the work of the group of Vologodski [65, 66] who experimentally reinvestigated the cyclization of 100-130 bp DNA, and showed that the j_M factor is in full agreement with theoretical prediction and Monte-Carlo simulations [65, 67] based on the WLC model.

B. Theoretical Approaches to Cyclisation Probability

The Wormlike Chain model [41] has been successfully used to describe the elasticity of a single DNA molecule under a force. The correctness of the theoretical prediction has been verified by the experimental force extension curve obtained by stretching of single DNA molecules [56, 57].

Numerical calculations of the WLC partition function, combined with expansions at small and high forces have provided simple formulas, widely used by experimentalists [58]. The WLC model is therefore a good model to calculate the loop formation probability too. This calculation is very difficult to perform because it amounts to the evaluation of a path integral, the action of which is the bending energy of the polymer with closed end. The latter condition introduces a global constraint which can be expressed through the introduction in the action of an imaginary force field (Lagrange multiplier vector). The calculation of the path integral with this imaginary force field is in fact identical to the one of the WLC model with a real force, with the additional step that the path integral is then integrated over the Lagrange multiplier values. A second difficulty comes from the fact that loop lengths of the order of 100 bp are too short for the thermodynamic limit to hold in the calculation of the constrained partition function. As we have already discussed two effects contribute to the difficulty that DNA has to make a loop, and their relative importances depend on the length of the molecule.

For long molecules, of at least about 10 persistence lengths, the contact of the two ends of the molecules is difficult due to the entropy loss expressing the decrease in the number of configurations accessible when the loop is formed.

The looping probability can be simply calculated with the Gaussian polymer model and reads:

$$j_M(L) = 1.66 \left(\frac{3}{4\pi LA} \right)^{3/2} \quad (1)$$

where $A = 50$ nm is the DNA persistence length. Note that $1 \text{ M} = 1 \text{ mol} \cdot \text{L}^{-1} = 1.66 \text{ nm}^{-3}$. Short molecules (up to some persistence lengths), are much more rigid. The energetic cost to bend these molecules and put in contact the two ends is very large. Optimally closed configuration, with a ring shape, have a bending energy given by:

$$\Delta E_c(L) = AL/R^2 = 4\pi^2 \frac{A}{L} \quad (2)$$

where $R = L/(2\pi)$ is the ring radius. The real configuration is indeed not a circle if, as is generally the case, the tangent vector of the first and the last bases are to aligned along the same direction. The optimal configuration and its energy can be calculated in the continuous limit because they are related to the solution of the Mathieu equation for the motion of the pendulum [45]. The resulting bending energy is somewhat smaller than the ring one (2), and reads

$$\Delta E(L) = 14.055 \frac{A}{L}. \quad (3)$$

The optimal configuration for a 100 bp molecule is shown in Fig. 5 (top). The contribution of quadratic fluctuations around the optimal configuration to the cyclisation factor has been calculated by Shimada and Yamakawa with the outcome

$$C_{SY}(L) = \frac{1.66 \times 112.04}{A^3} \left(\frac{L}{A} \right)^{-5} e^{0.246 \frac{L}{A}} \quad (4)$$

leading to the final expression for the cyclisation factor:

$$j_M(L) = C_{SY}(L) e^{-\Delta E(L)}. \quad (5)$$

Expression (4) for $C_{SY}(L)$ has been obtained through a very involved calculation, in part analytical and in part numerical. While the saddle point configuration and its related elastic energy are easy to re-derive [72, 86], the calculation of the fluctuations is tricky, mainly because the fluctuations are calculated in the discrete space while the saddle point configuration is calculated in the continuous space limit. It would be very interesting to have an alternative derivation for this entropic cost factor. A similar calculation for the quadratic fluctuations was recently done by Kulic *et al.* around the optimal configuration of a looped DNA stretched at high force [87]. The looping free energy can be separated into a bending and an entropic contributions

$$\Delta G(L) = \Delta E(L) - \Delta S(L). \quad (6)$$

The entropic part $\Delta S(L)$ (in units of $k_B T$) in the looping formation is given by $\Delta S(L) = -\ln [C_{SY}(L) \times dv/1.66]$ with $dv = \frac{4}{3}\pi r^3$ (in nm^{-3}) the interaction volume in which the ring closes. As a numerical illustration, for a 100 bp loop, the elastic cost is of $\Delta E(100 \text{ bp}) = 21 k_B T$; the entropic loss is of $\Delta S(100 \text{ bp}) = -3 k_B T$ for an interaction radius of $r = 1$ nm (typical dimension of a binding site). Note that the entropic loss strongly depends on the interaction volume. For example if one considers the typical protein linear dimension of $r = 10$ nm a contribution $3 \times \ln 10 = 7 k_B T$ has to be added to the entropy $\Delta S(L)$. In Fig. 5 we show the Gaussian polymer formula (1) valid for long loops (full gray line) and the short-loop Shimada-Yamakawa formula (5) (full black line).

Fig. 5 also shows the curve obtained with two alternative approaches: the brownian dynamics simulations [37, 61] (dotted line) and the numerical calculation of the path integral under the global constraint of zero end-to-end distance [78, 82, 86]. The cyclisation probability goes to zero for small and long molecules and exhibits a maximum in between. This maximum is located at about 3.3 persistence lengths (500 bp), see Fig. 5.

As DNA is made of two strands it not only has a bending modulus but also a torsional modulus. This was included in the Shimada-Yamakawa framework [46]. In fact cyclisation also requires an angular phasing between the two extremities for the number of helical pitches to agree with the ~ 10.5 bp turn of circular DNA. The linking number dependent j_M factor increases and oscillates up to about 3 persistence lengths, then decreases monotonically.

As shown in Fig. 5 the Shimada-Yamakawa saddle point calculation works very well for lengths up to 1500 bp, while the Gaussian approximation is valid for DNA having more than 3000 bp. An approximate formula including both short and large length scale behavior has been obtained by combining the Gaussian behavior at large length scales

with an exponential decay at short length scales [28, 84, 85]. This formula is shown for comparison in Fig. 5 (red dotted line). The Shimada-Yamakawa approach is so far the best expansion in the short-stiff chain limit; Wilhelm and Frey have also developed an expansion in this limit [52], but the validity domain of their approach is limited to lengths of the molecule up to about 300 bp (two persistence lengths). Highest order corrections have been calculated by perturbation theory but do not substantially improve Wilhelm and Frey's result [53–55]. On the other hand, expansions around the long-flexible limit have also been developed by Daniels in 1952 [43]. The first-order correction to the Gaussian chain works well for molecule lengths larger than about 1000 bp (10 persistence lengths). Other approaches consist in relaxing the inextensibility constraint of the chain [48–51] but so far, did not succeed in catching the rigid rod-like regime in a systematic way (Ha *et al.* works give an accurate estimate for the most probable end-to-end, but not for the whole distribution for instance).

V. PROTEIN MEDIATED LOOPS

A. Some Historical Background

A step forward in complexity and biological interest was the study of the protein-protein interactions mediated by DNA loops [1]. The aforementioned *lac*, *gal* [34, 35, 38] and recombinase [28] *in vitro* experiments were carried out in the last decade, through the preparation of DNA fragments with a distance between the *lac* operators of about 300 and 100 bp in the first two cases, and a distance between the recombinase binding sites varying from 74 to 1500 bp in the third case.

Protein mediated loop experiments can be more complex than cyclisation experiments because of the presence of both protein-protein and DNA-proteins interactions. Each interaction has its own affinity constant, and competitions between these interactions with different possible populations of complexes are possible [1, 34, 38]. An example is the GalR-DNA complex shown in Fig. 1, and made of two dimers bound to each other to form a tetramer and to DNA. If, for example, the two DNA operators carry respectively a dimer and a tetramer the loop cannot be formed.

Proteins may in addition impose geometrical constraints to DNA (see for instance the parallel or antiparallel configurations in Fig. 1) [1]. These constraints have to be taken into account at different stages of simplification in theoretical approaches.

Theoretical investigations have mainly focused on how proteins can enhance the looping probability [37]. Three important factors have been studied with brownian dynamic and Monte Carlo simulations by the Langowski and Vologodskii groups since 1992 [22, 60, 61].

First a purely geometrical factor is relevant for small *i.e.* up to about 300 bp DNA loops: the protein has a linear size r , typically of the order of 10 nm. This is a non negligible distance with respect to the DNA persistence length ($A = 50$ nm). As firstly pointed out in the work of Brenowitz *et al.* the protein can act as a bridge in the loop formation. The loop gets clamped when the end-to-end distance of the molecule is about r , and not zero. The bending energy in presence of this protein bridge is therefore reduced with respect to a closed loop.

Secondly, the presence of additional proteins that sharply bend the DNA fragment can also facilitate the loop formation (see the HU protein in Fig. 1). Proteins can also produce a kink on the DNA in an indirect way by twisting it. This twist can give rise to a denaturation bubble and therefore a zone of high flexibility on the DNA segment. This action has been modeled in brownian dynamical simulations [37, 60, 61] by introducing a kink, of 120° or 80° , in the middle of the chain. The presence of a kink has been shown to greatly enhance the looping probability for molecular lengths of less than roughly 500 bp.

Thirdly when a torque is applied, DNA may also buckle in much the same way as an elastic tube, and create plectonemes. These structures, which also exist *in vivo* may strongly affect the end-to-end collision probability, as first pointed out by Vologodskii *et al.* [22]. The presence of superhelical conformation can enhance the loop probability for loop lengths up to about 1000 bp [37].

The geometrical and topological constraints due to the proteins, in general, make analytical treatments more difficult. A first, numerical method, has been proposed by Balaeff *et al.* to calculate the bending energy given precise boundary conditions obtained from the knowledge of the LacR protein structure [74, 75]. In these calculations the twist and electrostatic energies are included but the entropic contribution to the free energy is neglected. A second method recently developed by Yan *et al.* [78] and Douarche *et al.* [86], consist in the numerical calculation of the WLC path integral with the r end-to-end distance constraint. It allowed the authors to calculate the loop probability with a protein bridge, and to impose particular boundary conditions (for example, the parallel or anti-parallel configuration in Fig. 1) [78], or a kink in the middle of the sequence. A third approach that we detail in the next section is the extension of the Shimada-Yamakawa calculation to the case of an open and kinked loop.

B. Theoretical Approaches for the Effects of Protein Bridge and Kink at Half-Length

A protein bridge can easily be accounted for Shimada and Yamakawa saddle point calculation [86]. The optimal configuration for a 100 bp (34 nm) DNA with an end-to-end separation $r = 10$ nm is shown in Fig. 6. For the entropic contribution C due to quadratic fluctuations around the optimal configuration a simple and surprisingly accurate estimate is obtained by adapting Shimada-Yamakawa formula (4), and approximating

$$C(L, r) = C_{\text{SY}}(L + 2r). \quad (7)$$

This approximation can be understood if one compares (see Fig. 6) the open configuration with the closed configuration with a length increased by twice the end-to-end separation. The closing part has no elastic cost because it is essentially not curved, and contributes to fluctuations only. The choice of the factor 2 can be explained in geometrical terms [86]. As shown in Fig. 6 the cyclisation factor for a 100 bp DNA is enhanced by a factor 10^2 in presence of a protein bridge of 10 nm between the DNA extremities. For a DNA length more than about 10 times the end-to-end separation the curve goes to the Shimada-Yamakawa closed loop formula (5). Fig. 6 also shows the good agreement between the saddle point extension of Shimada and Yamakawa formula (top panel) numerical calculation (bottom panel), and the brownian dynamics (open squares) and Monte Carlo simulation (filled circles) points. Note that the brownian dynamics points are obtained from a model which includes twist rigidity and electrostatic effects. The free energy of loop formation is shown in Fig. 6 (left). A kink in the middle of the molecule can be treated in an analogous manner. The saddle point kinked configuration are shown in Fig. 7 (top). A kink of 90° increases the loop probability by a factor 10^5 . The most probable loop length decreases also from 500 bp to the minimal length of 35 bp for a kink of 90° . The free energy of loop formation is shown on the left of Fig. 7. A loop of 100 bp with a kink of 150° in the middle of the sequence and an end-to-end distance $r = 10$ nm has a free energy cost $\Delta G = 9 k_B T$ to be compared with the value of $\Delta G = 12 k_B T$ found from the loop formation mediated by the GalR and HU proteins [35] (see Fig. 4). An even simpler approach is to directly calculate the cyclization factor with the Shimada-Yamakawa formula on the equivalent loop of length $(L + 2r)$. This approach is similar to what was suggested in 1991 by Brenowitz *et al.* to interpret their experimental data *i.e.* to directly consider the protein as part of the length of the loop, directly in the Shimada-Yamakawa formula. The proposed formula for the cyclisation factor as a function of the protein size r , the length L of the DNA, and the kink angle κ reads

$$j_{\text{M}}(L, r, \kappa) = C_{\text{SY}}(L + 2r) e^{-\Delta E(L+2r, \kappa)} \quad (8)$$

A linear fit of the bending energy for the optimal configuration in the presence of a kink is

$$\Delta E(L, \kappa) = (-7.1 + 20.8 \frac{\kappa}{\pi}) \frac{A}{L} \quad (9)$$

Formula (8) allows us to obtain a simple prediction for the loop probability in presence of a kink in the middle of the sequence and a finite separation between the extremities.

As shown in Fig. 6 this formula (gray lines) is in good agreement with the loop probability obtained with the exact calculation of the saddle point energy of the open configuration (full line). Moreover the results are in good agreement with numerical calculations (graphics in the bottom). Rippe has also written down a simple formula that includes a kink angle κ and a finite end-to-end distance r in an effective way, but his formula contains a parameter that is fitted, for each values of r and κ , from the curve obtained by brownian dynamic simulations [84, 85]. The brownian dynamic data points, obtained from a model which includes twist rigidity and electrostatic effects, and the Rippe interpolation formula are shown in the inset of Fig. 6.

It is important to notice that protein-mediated site-to-site interaction differs from the end-to-end interactions because the DNA fragment is longer than the fragment involved in the loop. The extra DNA should reduce the contact probability between the two sites because it excludes certain conformations. The factor j_{M} calculated above through the end-to-end distribution is therefore an upper limit to the concentration of site-to-site interactions. These entropic effects have been included in theoretical approaches for long loops [59]. Moreover this extra DNA has been, for example, taken into account in the calculation of the formation of loop under an applied tension [71, 72, 87] described in the next section. Another effect that has been underlined in theoretical works is that the DNA stiffness depends on the sequence [83] and can be modified by proteins attached to DNA [76].

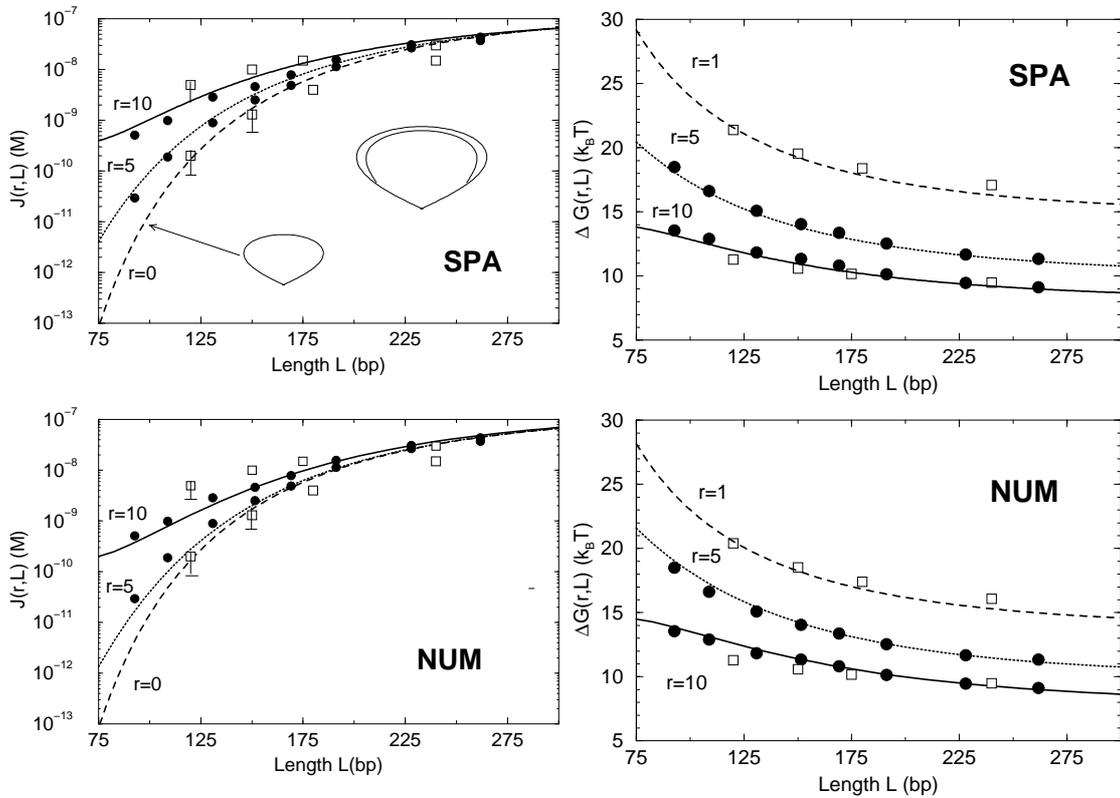


FIG. 6: Looping probability density (left panel) and free energy (right panel) within an interaction radius $r = 1$ nm (dashed line), 5 nm (dotted line), 10 nm (full line). Top: extension of the Shimada-Yamakawa calculation to finite r . Bottom: numerical calculation of the constrained path integral. The error bars are shown when they are larger than the symbol sizes. The $r = 1$ nm curve coincides with the $r = 0$ cyclization factor. Theoretical results are in good agreement with Monte Carlo simulations obtained by Vologodskii and Podtelezchnikov (full circle) [67] and brownian dynamics simulation obtained by Langowski and collaborators (open square) [61]. Inset of the top left panel: lowest bending energy configurations for 100 bp and the $r = 0$ nm (bottom), and $r = 10$ nm (top), the closure of the $r = 10$ nm configuration being shown by a thin line.

VI. CONCLUDING REMARKS: RECENT AND FURTHER WORKS

A. Loop Probability and Dynamics Under a Stretching Force

The single molecule experiments, started with Lia *et al.* work in 1995 [34, 35], of loop formation under a stretching force, have recently stimulated several theoretical works, in particular by the Marko group [70–72]. In their analysis the loops give rise to a length loss in the stretching work done by the force. The loop geometry has to be taken into account; for example in the antiparallel configuration of Fig. 1 the extra DNA cannot be fully stretched. The extension of Shimada-Yamakawa calculation of the quadratic fluctuations around the saddle point configuration has been performed in presence of a large external force by Kulić *et al.* [87]. The dynamics of loop formation, which can be directly studied by single strand experiments has been investigated by Monte Carlo simulations [61, 67]. From a theoretical point of view a Langevin dynamics on the end-to-end distance has been studied in [68], and calculation of the mean first passage of cyclisation has been performed in [69]. Clearly the application of polymer dynamics to the problem of loops formation deserves more investigations.

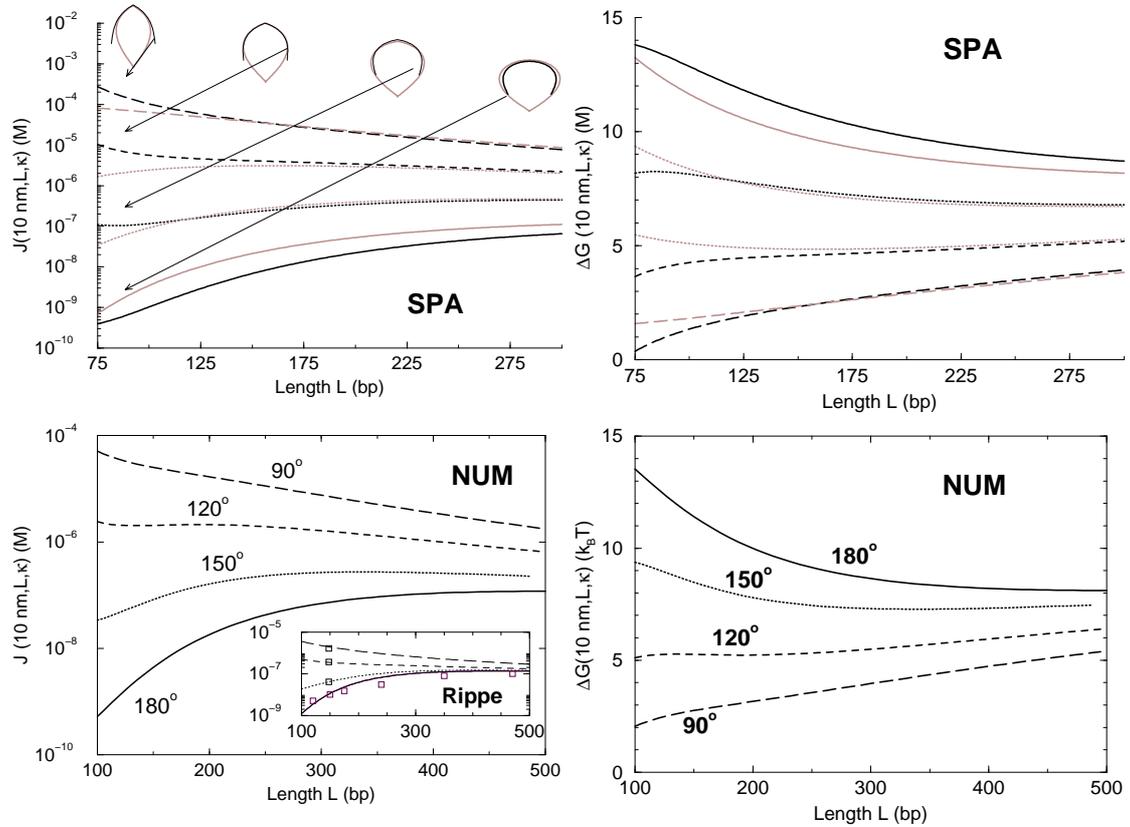


FIG. 7: Probability density (left) and free energy for a loop with $r = 10$ nm and a kink κ in the middle of the chain. $\kappa = 180^\circ$ (full lines), $\kappa = 150^\circ$ (dotted lines), $\kappa = 120^\circ$ (dashed line), $\kappa = 90^\circ$ (long dashed lines). Top: extension of Shimada-Yamakawa calculation (black lines), approximate formula (gray line) given in the text. Bottom: numerical calculation of the constrained path integral. Inset on the bottom left panel: brownian dynamics simulation points obtained by Langowski and collaborators (open square) [61], fitted by Rippe formula [84].

B. Towards *in vivo* Modeling

Of course the *in vivo* problem is more complicated because DNA is not free but attached to chromatin structural proteins and the uncontrolled presence of proteins may affect DNA elasticity. The experiments of Ringrose [28] show that the most probable loop length is smaller in chromatin than in bare DNA. Considering its importance this problem clearly deserves some work even at a simplified level. A very recent work of Sankararaman and Marko has studied the interaction between multiple loops on the same DNA. A simple effect for non site specific loops is that the most probable loop length can decrease due to the rearrangement entropy between multiple loops on the same DNA[73].

-
- [1] S. E. Halford, A. J. Welsh, M. D. Szczelkun, *Annu. Rev. Biophys. Biomol. Struct.* **33**, 1 (2004).
 - [2] R. Schleif, *Annu. Rev. Biochem.* **61**, 199 (1992).
 - [3] S. Adhya, *Annu. Rev. Genet.* **23**, 227 (1989).
 - [4] T. M. Dunn, S. Hahn, S. Ogden, and R. F. Schleif, *PNAS* **81**, 5017 (1984).
 - [5] K. S. Matthews, *Microbiol. Rev.* **56**, 123 (1992).
 - [6] M. Amouyal, L. Mortensen, H. Buc, and K. Hammer, *Cell* **58**, 545 (1989).
 - [7] M. Ptashne, Cold Spring Harbor Laboratory Press **edn.3** (2004).
 - [8] T. Tahirov, K. Sato, E. Ichikawa-Iwata, M. Sasaki, T. Inoue-Bungo, M. Shiina, K. Kimura, S. Takata, A. Fujikawa, H. Morii *et al.*, *Cell* **108**, 57 (2002).

- [9] P. Jackson, I. Mastrangelo, M. Reed, P. Tegtmeyer, G. Yardley, and J. Barrett, *Oncogene* **16**, 283 (1998).
- [10] W. Su, S. Porter, S. Kustu, and H. Echols, *PNAS* **87**, 5504 (1990).
- [11] C. Wyman, I. Rombel, A. K. North, C. Bustamante, and S. Kustu, *Science* **275**, 1658 (1997).
- [12] J. C. Becker, A. Nikroo, T. Brabletz, R. A. Reisfeld, *PNAS* **92**, 9727 (1995).
- [13] B. Tolhuis, R. J. Palstra, E. Splinter, F. Grosveld, and W. de Laat, *Mol. Cell* **10**, 1453 (2002).
- [14] J. Mahillon, and M. Chandler, *Microbiol. Mol. Biol. Rev.* **62**, 725 (1998).
- [15] Y. Voziyarov, S. Pathania, and M. Jayaram, *Nucleic Acids Res.* **27**, 930 (1999).
- [16] A. Nagy, *Genesis* **26**, 99 (2000).
- [17] M. L. Embleton, V. Siksnys, and S. E. Halford, *J. Mol. Biol.* **311**, 503 (2001).
- [18] M. L. Embleton, S. A. Williams, M. A. Watson, S. E. Halford, *J. Mol. Biol.* **289**, 785 (1999).
- [19] N. E. Murray, *Microbiol. Mol. Biol. Rev.* **64**, 412 (2000).
- [20] D. Dryden, N. Murray, and D. Rao, *Nucleic Acids Res.* **29**, 3728 (2000).
- [21] L. Postow, C. D. Hardy, J. Arsuaga, and N. R. Cozzarelli, *Genes Dev.* **18**, 1766 (2004).
- [22] A. Vologodskii, N. R. Cozzarelli, *Biophys. J.* **70**, 2548 (1996).
- [23] Y. L. Lyubchenko, L. S. Shlyakhtenko, B. Chernov, and R. E. Harrington, *PNAS* **88**, 5331 (1991).
- [24] Z. Katiliene, E. Katilius, and N. W. Woodbury, *Biophys. J.* **84**, 4053 (2003).
- [25] L. J. Parkhurst, K. M. Parkhurst, R. Powell, J. Wu, and S. Williams, *Biopolymers* **61**, 180 (2003).
- [26] S. Chatterjee, Y. N. Zhou, S. Roy, and S. Adhya, *PNAS* **94**, 2957 (1997).
- [27] J. D. Kahn, and D. M. Crothers, *PNAS* **89**, 6343 (1992).
- [28] L. Ringrose, S. Chabanis, P. Angrand, C. Woodroffe, and A. F. Stewart, *EMBO J.* **18**, 6630 (1999).
- [29] D. M. Crothers, J. Drak, J. D. Kahn, and S. D. Levene, *Methods Enzymol.* **212**, 3 (1992).
- [30] D. Shore, J. Langowski, and R. L. Baldwin, *PNAS* **78**, 4833 (1981).
- [31] K. Rippe, M. Guthold, P. H. von Hippel, and C. Bustamante, *J. Mol. Biol.* **270**, 125 (1997).
- [32] Q. Du, C. Smith, N. Shiffeldrin, M. Vologodskia, and A. Vologodskii, *PNAS* **102**, 5397 (2005).
- [33] D. L. Hamilton, K. Abremski, *J. Mol. Biol.* **178**, 481 (1984).
- [34] L. Finzi, *J. Gelles*, *Science* **267**, 378 (1995).
- [35] G. Lia, D. Bensimon, V. Croquette, J.-F. Allemand, D. Dunlap, D. E. Lewis, S. Adhya, and L. Finzi, *PNAS* **100**, 11373 (2003).
- [36] J. R. Marsden, and U.K. Laemmli, *Cell* **17**, 849 (1979).
- [37] K. Rippe, P. H. von Hippel, and J. Langowski, *Trends Biochem. Sci.* **20**, 500 (1995).
- [38] M. Brenowitz, A. Pickar, and E. Jamison, *Biochemistry* **30**, 5986 (1991).
- [68] S. Jun and J. Bechhoefer, *Phys. in Canada* **59**, 85 (2003).
- [40] M. Lewis, G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan, and P. Lu, *Science* **271**, 1247 (1996). The Protein Data Bank, *Molecules of the Month: Lac Repressor*, http://www.rcsb.org/pdb/molecules/pdb39_1.html (2003).
- [41] O. Kratky, and G. Porod, *Rec. Trav. Chim.* **68**, 1106 (1949).
- [42] J. J. Hermans and R. Ullman, *Physica* **18**, 951 (1952).
- [43] H. E. Daniels, *Proc. Roy. Soc. Edinburgh A* **63**, 290 (1952).
- [44] W. Gobush, H. Yamakawa, W. H. Stockmayer, and W. S. Magee, *J. Chem. Phys.* **57**, 2839 (1972).
- [45] H. Yamakawa, and W. H. Stockmayer, *J. Chem. Phys.* **57**, 2843 (1972).
- [46] J. Shimada, and H. Yamakawa, *Macromolecules* **17**, 689 (1984).
- [47] H. Yamakawa, *Helical Wormlike Chains in Polymer Solutions*, Springer Verlag (Berlin, 1997).
- [48] R. A. Harris, and J. E. Hearst, *J. Chem. Phys.* **44**, 2595 (1966). J. E. Hearst, R. A. Harris, and E. Beals, *J. Chem. Phys.* **45**, 3106 (1966).
- [49] J. B. Lagowski, J. Noolandi and B. Nickel, *J. Chem. Phys.* **95**, 1266 (1991).
- [50] B.-Y. Ha, and D. Thirumalai, *J. Chem. Phys.* **103**, 9408 (1995). J. K. Battacharjee, D. Thirumalai, and J. D. Bryngelson, *cond-mat/9709345* (1997).
- [51] R. G. Winkler, P.Reineker, and L. Harnau, *J. Chem. Phys.* **101**, 8119 (1994). R. G. Winkler, *J. Chem. Phys.* **118**, 2919 (2003).
- [52] J. Wilhelm and E. Frey, *Phys. Rev. Lett.* **77**, 2581 (1996).
- [53] S. Stepanow, and G. M. Schütz, *Europhys. Lett.* **60**, 546 (2002). S. Stepanow, *J. Phys.: Condens. Matter* **17**, S1799 (2005).
- [54] A. J. Spakowitz, and Z.-G. Wang, *J. Chem. Phys.* **119**, 13113 (2003), *Macromolecules* **37**, 5814 (2004).
- [55] H. Kleinert, and A. Chervayakov, *cond-mat/0503199* (2005).
- [56] C. Bustamante, J. F. Marko, E. D. Siggia and S. Smith, *Science* **265**, 1599 (1994).
- [57] J. F. Marko and E. D. Siggia, *Macromolecules* **28**, 8759 (1995).
- [58] M. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block, *Biophys. J.* **72**, 1335 (1997).
- [59] A. Hanke, and R. Metzler, *Biophys. J.* **85**, 167 (2003).
- [60] K. Klenin, H. Merlitz, and J. Langowski, *Biophys. J.* **74**, 780 (1998).
- [61] H. Merlitz, K. Rippe, K. Klenin, and J. Langowski, *Biophys. J.* **74**, 773 (1998).
- [62] P. J. Hagerman, and V. A. Ramadevi, *J. Mol. Biol.* **212**,351 (1990).
- [63] S. D. Levene, and D. M. Crothers *J. Mol. Biol.***189**,61 (1986), *J. Mol. Biol.***189**,73 (1986).
- [64] M. D. Frank-Kamenetskii, A.V. Lukashin, V.V. Anshelevich, and A.V. Vologodskii, *J. Bimol. Struct. Dyn.* **2**, 1005 (1985).
- [65] Q. Du, N. Shiffeldrin, M. Vologodskia, and A. Vologodskii, *PNAS* **102**, 5397 (2005).
- [66] Q. Du, M. Vologodskia, H. Kuhn, M. Frank-Kamenetskii, A. Vologodskii, *PNAS* **102**, 5397 (2005).

- [67] A. A. Podtelezchnikov, and A.V. Vologodskii, *Macromolecules* **33**, 2767 (2000).
- [68] S. Jun, J. Bechhoefer, and B.-Y. Ha, *Europhys. Lett.* **64**, 420 (2003).
- [69] A. Dua, and B. Cherayl, *J. Chem. Phys.* **117**,7765,(2002), *J.Chem. Phys.* **116**, 399 (2002).
- [70] J. F. Marko, and E. Siggia, *Biophys. J.* **73**, 2173 (1997).
- [71] S. Blumberg, A.V. Tkachenko, and J.C. Meiners, *Biophys. J.* **88**, 1692 (2005).
- [72] S. Sankararaman, and J. Marko, *Phys. Rev E* **71**, 021911 (2005)
- [73] S. Sankararaman, and J. F. Marko, *Phys. Rev. Lett.* **95**, 078104 (2005).
- [74] A. Balaeff, L. Mahadevan, and K. Schulten, *Phys. Rev. Lett.* **83**, 4900 (1999), *Structure* **12**, 123 (2004). A. Balaeff, C. R. Koudella, L. Mahadevan, and K. Schulten, *Phil. Trans. R. Soc. Lond. A* **362**, 1355 (2004).
- [75] E. Villa, A. Balaeff, and K. Schulten, *PNAS* **102**, 6783 (2005).
- [76] R. Bruinsma, and J. Rudnick, *Biophys. J.* **76**, 1725 (1999).
- [77] T. E. Cloutier, and J. Widom, *Mol. Cell* **14**, 355 (2004).
- [78] J. Yan, and J. F. Marko, *Phys. Rev. E* **68**, 011905 (2003), *Phys. Rev. Lett.* **93**, 108108 (2004). J. Yan, R. Kawamura, and J. F. Marko, *Phys. Rev. E* **71**, 061905 (2005).
- [79] P. A. Wiggins, R. Phillips, and P. C. Nelson, *Phys. Rev E* **71**, 021909 (2005).
- [80] B. Chakrabarti, and A. J. Levine, *Phys. Rev E* **71**, 031905 (2005).
- [81] Y. O. Popov, and A. V. Tkachenko, *Phys. Rev E* **71**, 051905 (2005).
- [82] P. Ranjith, P. B. Sunil Kumar, and G. I. Menon, *Phys. Rev. Lett.* **94**, 138102 (2005).
- [83] D. Garrivier, and B. Fourcade, *Europhys. Lett.***49**, 390 (2000).
- [84] K. Rippe, *Trends Biochem. Sci.* **26**, 733 (2001).
- [85] V. A. Bloomfield, D. M. Crothers, and J. I. Tinoco, *Physical Chemistry of Nucleic Acids*, Harper and Row (1974).
- [86] N. Douarache, and S. Cocco, cond-mat/0508288 (2005), submitted to *Phys. Rev. E*.
- [87] J. M. Kulić, H. Mohrbach, V. Lobaskin, R. Thakkar, and H. Schiessel, cond-mat/0501379 (2005).

Deuxième partie

Régulation à l'étape de réplication : bases d'une étude dynamique

Chapitre 8

Expression d'un gène

8.1 Retour sur le dogme central

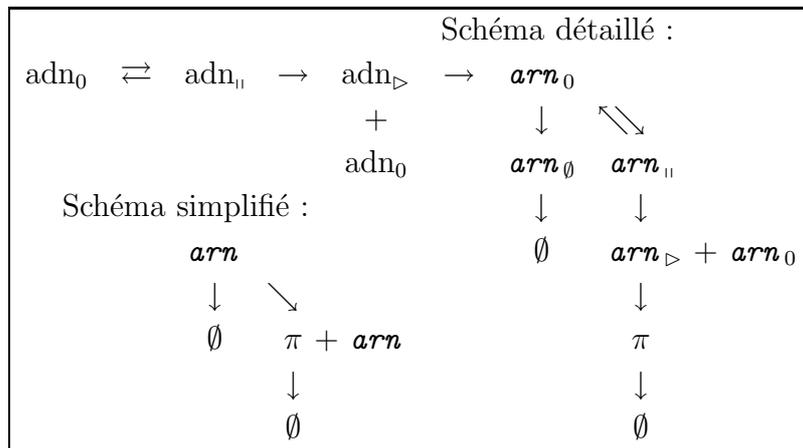
Dans l'avant propos, le dogme central a été introduit comme un schéma *simple* nous permettant de relier un *gène* à son expression, c'est-à-dire aux *protéines* dont il gouverne la synthèse. La GFP – pour « Green Fluorescent Protein » – est une protéine capable d'absorber puis de réémettre dans une longueur spécifique un rayonnement par *fluorescence* [Tsi98]. Elle permet donc une observation directe du niveau d'expression de son gène. En moins de dix ans la GFP a ainsi inspiré quantité d'expériences visant à caractériser l'expression de gènes particuliers.

Il n'est pas raisonnable au premier abord que le dogme central permette de décrire convenablement la *diversité* des mécanismes dont dispose *E Coli* pour réguler son expression génétique. Nous aimerions donc le « compléter », par exemple en décrivant les complexes que forment les ARN polymérases avec l'ADN ou le ribosome avec l'ARN, en incluant les gènes dont dépendent la transcription ou la traduction, etc. En fait nous voudrions tendre vers le réseau génétique le plus *exhaustif* possible. Comme il faut également pouvoir comprendre, résoudre ce réseau nous devons nous limiter au sous-réseau le plus *pertinent*. Selon le gène que nous observons, ou bien encore la propriété qui nous intéresse dans son expression, cette notion de pertinence n'a toutefois *rien de triviale*.

Dans la littérature nous trouvons en effet beaucoup de schémas [Pau05, SES02, Swa04, ARM98, BKR02, PK04] dont les avantages respectifs ne sont d'ailleurs pas toujours très clairs. Nous adoptons pour ce travail le schéma employé en 2002 par Swain, Elowitz et Siggia [SES02] : c'est une espèce de « dogme central détaillé » (table 8.1)

- il comprend peu de types de molécules : ADN, ARN polymérase, ARN (messager), ribosome, dégradosome et protéine. Il tient compte de la formation de complexes correspondant aux étapes de transcription et de traduction entre l'ARN polymérase, le ribosome, le dégradosome, l'ADN et l'ARN. Il

- n'inclut explicitement que le gène que l'on désire étudier ;
- il est donc constitué d'un nombre raisonnable de réactions entre ces différentes molécules. L'article fournit une estimation de toutes les constantes de réactions, basée sur des données biologiques. Cela diminue le nombre de paramètres libres et permet de se concentrer sur d'autres aspects du modèle (réplication de l'ADN, etc.).



TAB. 8.1 – Schémas de l'expression d'un gène selon [SES02]. Le schéma simplifié substitue au schéma détaillé le sous-schéma de l' arn_0 par celui en arn .

En général, les données biologiques sont insuffisantes et/ou largement variables selon les conditions expérimentales. La quasi-majorité des simulations utilisent donc des constantes de réaction *ad hoc* [BKR02, PE01]. Elles doivent alors être accompagnées d'une étude de l'espace des constantes de réactions, ce qui constitue un lourd travail étant donnée la dimension de cet espace. Nous réécrivons individuellement les réactions du schéma dans la table 8.2. Les symboles des types de molécules étudiés sont

- adn_0 : promoteur du gène dont nous observons l'expression ;
- $adn_{||}$: complexe formé par l'accrochage de l'ARN polymérase au promoteur. La transcription n'est pas initiée (pas de lecture) ;
- adn_{\triangleright} : transcription par l'ARN polymérase de la séquence d'ADN correspondant au promoteur, qui est libéré ;
- arn_0 : séquence d'ARN initiant l'expression du gène ;
- arn_{\emptyset} : complexe formé par l'accrochage du dégradosome à la séquence initiatrice ;
- $arn_{||}$: complexe formé par l'accrochage du ribosome à la séquence initiatrice. La traduction n'est pas initiée (pas de lecture) ;
- arn_{\triangleright} : traduction par le ribosome de la séquence d'ARN codant la protéine. La séquence initiatrice est libérée ;
- π : protéine correspondant au gène étudié.

Le produit d'une dégradation, peu importe de quoi, est désigné par \emptyset . Le schéma peut en fait se réduire au dogme central (schéma simplifié), auquel cas *arn* signifie simplement la séquence d'ARN d'intérêt qui peut soit se dégrader, soit synthétiser une protéine. Précisons nos conventions d'écriture. Nous distinguons les molécules ou les complexes relatifs aux molécules d'ADN et d'ARN au moyen de la police de caractère : adn_α et arn_α , où l'indice α est choisit par analogie avec les sigles des magnétophones.¹

Schéma détaillé		Schéma simplifié	
Réactions			
(1)	$\text{adn}_0 \rightarrow \text{adn}_{\text{II}}$	idem	(1)
(2)	$\text{adn}_{\text{II}} \rightarrow \text{adn}_0$	idem	(2)
(3)	$\text{adn}_{\text{II}} \rightarrow \text{adn}_{\triangleright} + \text{adn}_0$	idem	(3)
(4)	$\text{adn}_{\triangleright} \rightarrow \text{arn}_0$	$\text{adn}_{\triangleright} \rightarrow \text{arn}$	(4)
(5)	$\text{arn}_0 \rightarrow \text{arn}_{\emptyset}$	$\text{arn} \rightarrow \emptyset$	(5)
(6)	$\text{arn}_{\emptyset} \rightarrow \emptyset$		
(7)	$\text{arn}_0 \rightarrow \text{arn}_{\text{II}}$	$\text{arn} \rightarrow \pi + \text{arn}$	(6)
(8)	$\text{arn}_{\text{II}} \rightarrow \text{arn}_0$		
(9)	$\text{arn}_{\text{II}} \rightarrow \text{arn}_{\triangleright} + \text{arn}_0$		
(10)	$\text{arn}_{\triangleright} \rightarrow \pi$		
(11)	$\pi \rightarrow \emptyset$	idem	(7)
Diagrammes de corrélation			

TAB. 8.2 – Réactions des schémas proposés en table 8.1. Diagrammes de corrélation de ces réactions (§ 8.4, [GB00, Gib00]).

8.2 Sensibilité au bruit

En section 8.3 nous explicitons les valeurs des constantes de réaction et nous montrons la réduction du schéma détaillé à celui simplifié. Avant cela faisons une rapide mise au point. La présentation du travail adoptée occulte volontaire-

¹Cette convention peut se révéler ambiguë, mais nous la préférons au P de : Promoteur ? Polymérase ? Protéine ?

ment la véritable problématique, à savoir caractériser la *robustesse* d'un réseau génétique au(x) *bruit(s)*.

Si le travail de Swain et coll. est connu ce n'est pas pour avoir proposé un schéma d'expression génétique, plus détaillé que le dogme central et d'avoir montré sous quelles conditions il s'y réduit. En biologie les systèmes à petit nombre de molécules sont monnaie courante. Intuitivement nous attendons une forte sensibilité de leur fonctionnement au bruit, inévitable, dans le nombre de ces molécules. Une modélisation convenable du bruit semble être un ingrédient essentiel à une bonne compréhension des mécanismes de régulation de l'expression génétique. L'article de Swain et coll. contient une caractérisation expérimentale, numérique et théorique du bruit associé à l'expression d'un gène. Ils arrivent ainsi à séparer les composantes *intrinsèque* (propre au gène observé) et *extrinsèque* (influence du milieu²) du bruit : cette distinction est très utile expérimentalement. Notons que dans une telle étude, le réseau le plus pertinent n'est pas nécessairement le plus complet. Le dogme central peut parfaitement convenir, et la régulation du gène peut être grossièrement reproduite par un interrupteur « activé/réprimé » aléatoire [Pau05].

Dans le reste de ce travail, la question du bruit ne sera pas adressée directement. Le but est dans un premier temps plus modeste : fournir un outil numérique qui complète les expériences en cours dans l'équipe de Didier Chatenay.³ Nous choisissons des méthodes de simulations (§ 8.4) adaptées à la description de ce bruit, bien que nous ne l'exploitions pas ici.

8.3 Constantes de réaction. Cycle cellulaire

La constante de la i ème réaction est notée : k_i pour le schéma détaillé, k'_i pour le simplifié. Remarquons que ce sont toutes des « vitesses » de réactions, de dimension l'inverse d'un temps. En effet, tel que nous l'avons formulé (table 8.2) le schéma ne contient que des réactions unimoléculaires, ou du premier ordre [Ans]. Nous reviendrons sur ce point plus bas, lors de l'évaluation des k_i . Montrons tout d'abord comment se ramener au schéma simplifié, c'est-à-dire comment exprimer les k'_j en fonction des k_i . Nous pouvons reconduire immédiatement $k'_{1-4} = k_{1-4}$ et $k'_7 = k_{11}$, mais pour k'_5 et k'_6 ? Considérons les équations d'évolution des concentrations $[arn_0]$ et $[arn_{ii}]$.⁴ Nous utilisons pour cela le sous-schéma formé des

²Le bruit extrinsèque peut donc être vu comme une conséquence des bruits intrinsèques associés aux autres éléments du milieu.

³Qui utilisent aussi des protéines fluorescentes.

⁴Ce sont les classiques équations bilan. Nous verrons en section 8.4 que cette description requiert des hypothèses fortes sur les concentrations.

réactions 5, 7, 8 et 9 :

$$\frac{d}{dt} \begin{Bmatrix} [\mathit{arn}_0] \\ [\mathit{arn}_{II}] \end{Bmatrix} = \begin{Bmatrix} -(k_5 + k_7) & k_8 + k_9 \\ k_7 & -(k_8 + k_9) \end{Bmatrix} \begin{Bmatrix} [\mathit{arn}_0] \\ [\mathit{arn}_{II}] \end{Bmatrix}$$

dont les solutions s'obtiennent par exemple par diagonalisation

$$[\mathit{arn}_0] = A_0 \exp(\lambda_+ t) + B_0 \exp(\lambda_- t) \quad [\mathit{arn}_{II}] = A_{II} \exp(\lambda_+ t) + B_{II} \exp(\lambda_- t)$$

avec les valeurs propres

$$\lambda_{\pm} = \frac{-s \pm \delta}{2} \quad \text{où} \quad \begin{cases} s = k_5 + k_7 + k_8 + k_9 \\ \delta^2 = s^2 - 4k_5(k_8 + k_9). \end{cases}$$

Nous évaluons k'_5 à partir de la demi-durée de vie de arn_0 .⁵ Nous considérons pour cela le mode propre qui fournit la borne supérieure de $t_{1/2} = -\ln 2 / \lambda_{\pm}$. comme $\lambda_- < \lambda_+ < 0$, nous choisissons $k'_5 \simeq -\lambda_+$ que nous minorons encore par $k'_5 \simeq -\ln 2 \times \lambda_+ = \frac{\ln 2}{2}(s - \delta)$.

Enfin, $k'_5 \times \Delta t$ représente la probabilité d'observer la dégradation d'une molécule d' arn au bout d'un temps $\Delta t \ll 1$. C'est-à-dire qu'il ne se produit pas de dégradation avant. Quel est le nombre moyen de π produites pendant ce temps par la molécule d' arn ? Soit Π le nombre de π produites par une molécule d' arn . Cela signifie Π accrochages réussis (attention, ils sont réversibles) d'un ribosome (réactions 7 puis 9) suivis d'un accrochage (irréversible cette fois) de dégradosome (réaction 5). Considérons que la loi de probabilité de Π est géométrique : si r et $d = 1 - r$ sont les probabilités d'un accrochage réussi à la molécule d' arn , d'un ribosome ou d'un dégradosome respectivement, alors la loi de Π est $r^{\Pi} \times d$ et sa valeur moyenne est $\langle \Pi \rangle = r/d$. Considérant des processus de Markov (l'état suivant ne dépends que de l'actuel) les chemins conduisant à l'accrochage d'un ribosome ou d'un dégradosome sont :

- partant de arn_0 on peut réaliser l'accrochage réussit d'un ribosome avec le poids $k_7 \times k_9$;
- partant de arn_{II} un dégradosome peut s'accrocher, après qu'un ribosome est réussi ou non son accrochage, avec des poids respectifs $k_9 \times k_5$ et $k_8 \times k_5$.

Mais alors

$$r = \frac{k_7 k_9}{k_7 k_9 + k_5(k_8 + k_9)} \quad \text{d'où} \quad \langle \Pi \rangle = \frac{k_7 k_9}{k_5(k_8 + k_9)},$$

et en première approximation k'_6 est donnée par $k'_6 \times \Delta t = \langle \Pi \rangle k'_5 \times \Delta t$.

Ces évaluations sont reproduites de l'article de Swain et coll. [SES02], de même que les valeurs des constantes de réactions qui suivent. Avant cela nous avons besoin des réactions bimoléculaires, ou du deuxième ordre [Ans], dont sont issues les réactions 1 et 7 qui décrivent la formation des complexes d'initiation de la transcription et de la traduction respectivement :

⁵Temps au bout duquel la concentration est moitié moindre que celle de l'instant initial.

(1bi) $\text{adn}_0 + \text{poly} \xrightarrow{K_1} \text{adn}_n$ avec « *poly* » l'ARN polymérase ;

(7bi) $\text{arn}_0 + \text{ribo} \xrightarrow{K_7} \text{arn}_n$ avec « *ribo* » le ribosome.

Les approximations unimoléculaires supposent un excès significatif et quasi-constant en ARN polymérase ainsi qu'en ribosomes dans le milieu, ainsi $k_1 = K_1[\text{poly}]$ et $k_7 = K_7[\text{ribo}]$. Notons que les constantes de réaction K_1 et K_7 ont la dimension du produit d'un volume par l'inverse d'un temps, c'est-à-dire l'inverse du produit d'une concentration par un temps (exprimées en $\text{M}^{-1} \cdot \text{s}^{-1}$). En reprenant alors [SES02] et ses références :

1. le taux d'accrochage d'une polymérase au promoteur est de $1.4 \times 10^7 \text{ M}^{-1} \cdot \text{s}^{-1} = K_1$. La concentration en polymérase libres dans la cellule est d'environ 30 nM d'où la constante effective $k_1 = K_1 \times 30 \times 10^{-9} = 0.42 \text{ s}^{-1}$;
2. la constante d'équilibre d'accrochage d'une polymérase $K_1/k_2 = 1.4 \times 10^8 \text{ M}^{-1}$ implique $k_2 = 0.1 \text{ s}^{-1}$;
3. varie largement en fonction du promoteur. Prendre $k_3 \in [10^{-3}; 0.1] \text{ s}^{-1}$;
4. le ribosome (ou le dégradosome) ne pouvant s'accrocher qu'à un ARN d'au moins 130 nt (nucléotides), transcrit par une polymérase dont la vitesse est de $40 \text{ nt} \cdot \text{s}^{-1}$ d'où la constante $k_4 = 40/130 \simeq 0.3 \text{ s}^{-1}$;
5. cette constante est déduite de $k_5 = k_7 k_9 / [\langle \Pi \rangle (k_8 + k_9)]$ (voir réaction 9) ;
6. choix empirique $k_6 = 0.1 \text{ s}^{-1}$;
7. le taux d'accrochage d'un ribosome à la séquence d'ARN initiant l'expression du gène est de $10^7 \text{ M}^{-1} \cdot \text{s}^{-1} = K_7$. La concentration en ribosomes libres dans la cellule est de 400 nM, d'où la constante effective $k_7 = 4 \text{ s}^{-1}$;
8. la constante d'équilibre de l'accrochage d'un ribosome $K_7/k_8 = 2.5 \times 10^7 \text{ M}^{-1}$ implique $k_8 = 0.4 \text{ s}^{-1}$;
9. estimée à $k_9 = 0.3 \text{ s}^{-1}$. Pour un nombre moyen de protéines $\langle \Pi \rangle = 15$ alors $k_5 = 0.114 \text{ s}^{-1}$;
10. une protéine est codée par environ 1000 nt, qui sont lus par un ribosome à la vitesse de $48 \text{ nt} \cdot \text{s}^{-1}$, soit $k_{10} = 48/1000 = 0.048 \text{ s}^{-1}$;
11. la demi-durée de vie de la protéine est d'environ $\ln 2/k_{11} = 3 \text{ h}$, d'où $k_{11} = 6.42 \times 10^{-5} \text{ s}^{-1}$.

Pour le schéma simplifié :

5'. $k'_5 = 11.53 \times 10^{-3} \text{ s}^{-1}$ (dégradation ARN) ;

6'. et $k'_6 = 0.173 \text{ s}^{-1}$ (traduction).

Pour simuler le cycle cellulaire d'*E Coli*, nous devons en spécifier la durée $\tau = 1 \text{ h}$, ainsi que :

- sa croissance volumique, choisie linéaire $V(t_{\text{rel}}) = V_0(1 + t_{\text{rel}}/\tau)$ avec t_{rel} le temps relatif au début du cycle et $V_0 = 2.5 \times 10^{-15} \text{ L}$ soit $1505.6 \mu\text{M}^{-1}$;

- le mécanisme de réplication de l'ADN (le chromosome). Nous doublons le matériel génétique, ce qui revient à ajouter une molécule d' adn_0 (promoteur libre), à $t_{\text{rel}} = 0.4 \times \tau = 24$ min fixé arbitrairement.

Nous devons également nous donner une recette de répartition des molécules entre les deux bactéries filles à la division (fin du cycle cellulaire). La conservation de l'information génétique entre générations signifie l'héritage d'exactly un chromosome, soit une molécule d' adn_0 , par bactérie fille. Les autres espèces sont réparties binômialement avec un paramètre p : sur les N molécules d'un type à la fin du cycle, chacune peut être héritée soit par l'une des filles avec la probabilité p , soit par l'autre avec la probabilité $q = 1 - p$. La probabilité pour qu'elles récupèrent respectivement N_0 et $M_0 = N - N_0$ molécules au total est donc $B(N_0; p, N) = C_{N_0}^N p^{N_0} q^{M_0}$ soit $B(N_0; 1/2, N) = \frac{N!}{N_0! M_0!} \frac{1}{2^N}$ si $p = q = 1/2$, ce qui constitue *a priori* notre choix. En pratique, à chaque division nous ne suivons qu'une des deux bactéries filles. Nous espérons alors que la lignée engendrée au cours des générations permet une description statistique fidèle d'une population cultivée expérimentalement.

8.4 Dynamique stochastique : algorithmes de Gillespie

Pour décrire le schéma nous pouvons penser écrire puis résoudre, numériquement si besoin est, le système d'équations différentielles gouvernant l'évolution des diverses concentrations. Dans cette optique l'évolution des concentrations est alors déterministe et continue, ce qui constitue deux hypothèses fortes. La continuité des concentrations peut se justifier à grand nombre de molécules, ce qui n'est pas notre cas : au cours du cycle cellulaire le nombre de promoteur libre adn_0 ne prend par exemple que trois valeurs $N = 0, 1$ ou 2 . Dans ces conditions nous avons également souligné en section 8.2 la potentielle sensibilité du schéma aux fluctuations, mêmes faibles, dans son nombre de molécules. Une description déterministe n'est donc pas appropriée. C'est pourquoi nous optons pour une description stochastique de la dynamique du schéma.

Plutôt que d'introduire un terme de bruit dans le système d'équations différentielles, nous considérons l'équation maîtresse du système. Par définition, celle-ci gouverne l'évolution temporelle de la probabilité d'avoir à un instant donné, un nombre spécifié de molécules de chacun des huit types considérés : adn_0 , adn_{I} , etc.⁶ Nous attirons l'attention du lecteur sur l'importance de discerner explicitement le nombre de molécules (d'exemplaires, de copies) du nombre de types (d'espèces) de molécules. L'équation maîtresse comprend donc les termes de peuplement et de dépeuplement correspondants aux 11 réactions du schéma étudié. Son écriture est

⁶Nous n'incluons pas les dégradations.

lourde, et une résolution analytique n'est pas possible. Une alternative consiste à construire une chaîne de Markov dont chaque évènement est la réalisation d'une des 11 réactions du schéma, en un temps donné. Avant de préciser cette chaîne, notons une caractéristique commune à toutes ces descriptions : l'absence d'information spatiale sur les molécules. Elle se justifie dans un premier temps en supposant que le milieu cellulaire est bien homogénéisé.

Le caractère markovien de la chaîne signifie que l'état à venir du système ne dépend que de l'actuel. À un instant de la dynamique nous devons alors être capable de répondre aux deux questions :

- quelle est la prochaine réaction ?
- Quand se produit-elle ?

Nous retenons pour cela l'algorithme proposé en 1977 par Gillespie [Gil97], qui reproduit en fait la distribution exacte du système.⁷ Précisons notre description du système : plutôt que de considérer les 11 réactions pour chacune des molécules disponibles de chaque type, nous envisageons l'ensemble des réalisations disponibles de chacune des 11 réactions. Par exemple, si nous disposons de N *arn*₀ la réaction 7 peut se produire avec la constante k_7 pour chacune de ces molécules. Alternativement, cette réaction est susceptible de se produire pour les molécules du type *arn*₀ avec la constante $a_7 = N \times k_7$, appelée propensité. Cela est vrai pour l'approximation unimoléculaire de la réaction 7. Considérée comme bimoléculaire (7bi), cette dernière se produira avec la propensité $A_7 = N \times M \times K_7$ pour les molécules de types *arn*₀ et *ribo* (en nombre M). Notons qu'il est parfaitement possible que k_7 et donc a_7 dépendent explicitement du temps, par exemple au travers du volume de la bactérie [LVTH04]. Nous aurons l'occasion de revenir sur ce point plus tard (§ 9.3.3) : pour le moment nous supposons que la dépendance en temps se trouve uniquement dans les nombres de molécules. Nous pouvons alors montrer que le label i de la prochaine réaction et le temps θ au bout duquel elle se produit forment un couple de variables aléatoires, discrète et continue respectivement, de distribution

$$f_i(\theta) = a_i \exp(-s\theta) \quad \text{avec} \quad s = \sum_{j=1}^{11} a_j.$$

La loi et la distribution marginales sont alors

$$R_i = \frac{a_i}{s} \quad \text{et} \quad \varepsilon(\theta; s) = s \exp(-s\theta),$$

soit pour θ une distribution exponentielle de paramètre la somme s des propensités. C'est la *méthode directe* – en anglais « Direct Method », abrégée en DM – de Gillespie [Gil97, GB00, Gib00], sur laquelle est basée l'algorithme suivant :

⁷Une démonstration est disponible dans l'article original [Gil97] ou la thèse de Gibson [Gib00].

- $t = t_{\text{rel}} = 0$: initialisation des constantes de réaction k_{1-11} , du nombre de chaque type de molécules et du volume de la bactérie. Calcul des propensités a_{1-11} et de leur somme;
- choix de la réaction à exécuter j selon la loi R_i . Temps $\hat{\theta}$ au bout duquel elle se réalise, distribué selon $\varepsilon(\theta; s)$;
- sauvegarde $t_{\text{old}} = t_{\text{rel}}$ et mise à jour $t_{\text{rel}} = t_{\text{rel}} + \hat{\theta}$. Modification du nombre de chaque type de molécules affecté par l'exécution de la réaction j . Calcul des propensités ayant changées, données par le diagramme de corrélation [GB00, Gib00] de la table 8.2;
- réplication si $t_{\text{old}} < 0.4\tau \leq t_{\text{rel}}$: désassemblage des complexes adn_{ll} et adn_{D} . Le nombre d' adn_0 passe de 1 à 2. Calcul des propensités a_{1-4} ;
- division si $t_{\text{old}} < \tau \leq t_{\text{rel}}$: désassemblage de tout les complexes, transmission d'un chromosome à chaque bactérie fille et répartition binomiale du reste des molécules. Calcul de toutes les propensités. Mises à jour $t = t + t_{\text{rel}}$ et $t_{\text{rel}} = t_{\text{rel}} - \tau$;
- mesure à une fréquence assurant une faible corrélation entre mesures successives : calcul du volume, sauvegarde permanente⁸ du nombre et de la concentration des types de molécules que nous voulons étudier à la fin de l'expérience;
- calcul de la somme des propensités puis retour en • jusqu'au t voulu, c'est-à-dire suffisant pour accumuler une bonne statistique de mesures.

Nous pouvons montrer qu'une manière équivalente de choisir (i, θ) consiste à générer autant de temps que de réactions θ_{1-11} selon des distributions exponentielles de paramètre leur propensité $\varepsilon(\theta_i; a_i)$, puis d'exécuter celle de temps minimum. C'est la *méthode de la première réaction* – en anglais « First Reaction Method », abrégée en FRM – également proposée par Gillespie. Par itération (•–○), elle nécessite la génération d'autant de nombres aléatoires que de réactions, soient 11 contre 2 pour la DM. En revanche elle n'utilise pas la somme des propensités. Cette somme constitue également un processus de durée proportionnelle au nombre de réactions. La question du générateur aléatoire est également cruciale. Les diverses distributions peuvent être générées à l'aide d'un générateur uniforme sur l'intervalle $[0; 1[$. Connaissant l'expression analytique exacte de la distribution nous utilisons la méthode dite « de transformation » [PTVF02]. Nous ne discutons pas ici le choix du générateur : nous avons utilisé un programme nous ayant précédemment donné entière satisfaction [PTVF02, DCJP03, Dou02].

⁸La sauvegarde n'est *a priori* pas judicieuse, comparée à un calcul dynamique des moments de la grandeur observée. Dans notre cas, elle ne pose de problème ni pour le stockage (mémoire), ni pour le temps de calcul.

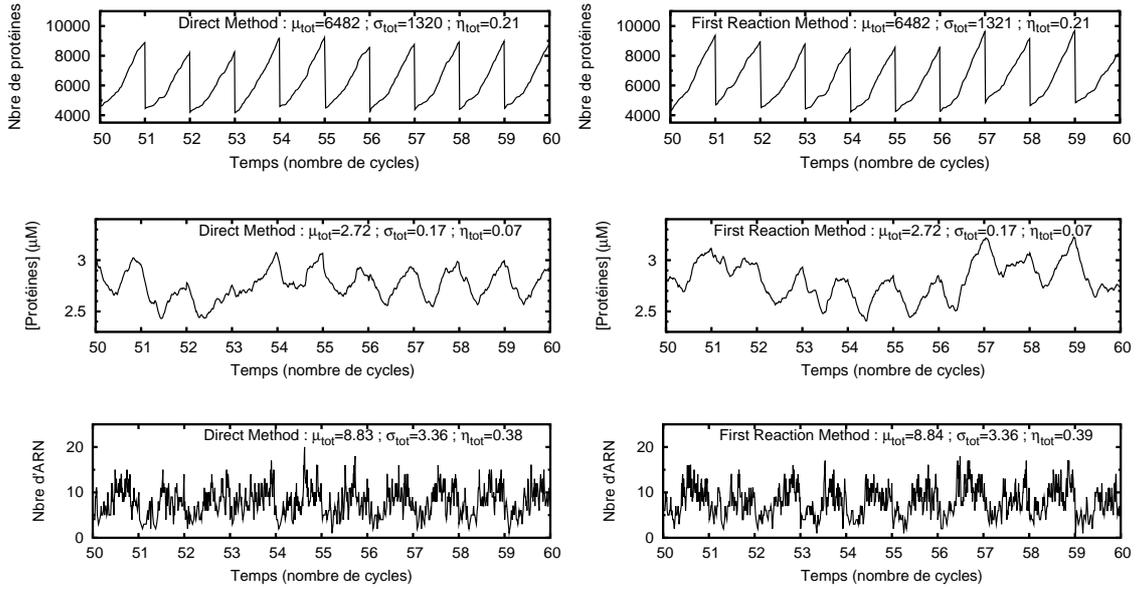


FIG. 8.1 – Extrait de la dynamique en nombre et concentration de protéines ainsi qu'en nombre d'ARNs, en employant le schéma détaillé (table 8.2). Les deux algorithmes utilisant la DM (gauche) et la FRM (droite) sont considérés.

8.5 Résultat : l'« étalon » chromosome

Nous réalisons cent simulations, en jouant pour chacune sur le générateur aléatoire (condition initiale), de 10000 cycles cellulaires chacune. Nous utilisons les algorithmes inspirés de la DM et de la FRM, et ce pour les schémas détaillé (figure 8.1) et simplifié (figure 8.2). En moyenne nous réalisons 100 mesures par cycle, au cours desquelles nous enregistrons le temps, les nombres d'ADNs, d'ARNs et de protéines ainsi que leur concentration. La réplication, la division et la croissance volumique font que ces quantités atteignent un cycle limite aux grands temps, plutôt qu'un état stationnaire (figure 8.3). Nous estimons (visuellement) que ce régime est atteint à partir d'une dizaine de cycles. Disposant de N mesures de la quantité X , on calcule ensuite moyenne et écart-type totaux au moyen des estimateurs empiriques :

$$\mu_{\text{tot}} = \frac{1}{N} \sum_{i=1}^N X_i \quad \text{et} \quad \sigma_{\text{tot}}^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \mu_{\text{tot}})^2.$$

Remarquons que σ_{tot}^2 utilise μ_{tot} et non une valeur moyenne exacte connue *a priori*. C'est pourquoi il emploie le facteur $(N-1)$ au lieu du N attendu. Numériquement l'évaluation de σ_{tot}^2 est sensible aux erreurs d'arrondis, particulièrement lorsque X est de grande amplitude. Comme nous sommes à termes intéressés par une évaluation fine des bruits totaux $\eta_{\text{tot}} = \sigma_{\text{tot}}/\mu_{\text{tot}}$, nous utilisons

plutôt l'estimateur [PTVF02]

$$\sigma_{\text{tot}}^2 = \frac{1}{N-1} \left\{ \sum_{i=1}^N (X_i - \mu_{\text{tot}})^2 - \frac{1}{N} \left[\sum_{i=1}^N (X_i - \mu_{\text{tot}}) \right]^2 \right\}.$$

Les résultats sont en accord avec [SES02]. Nous constatons pour le schéma détaillé que la FRM est sensiblement plus rapide que la DM : environ 30 contre 40 minutes de temps de calcul. Le schéma simplifié permet un gain de temps significatif : la durée des simulations tourne autour de 5–10 minutes. L'expression du gène, c'est-à-dire le nombre ou la concentration en protéines, obtenue par ce dernier schéma est fidèle à celle réalisée par le schéma détaillé : mêmes moyenne et écart-type, et surtout bruit identique. Seul le nombre d'ARNs est affecté : il est en moyenne plus élevé (de 2 molécules) et varie un peu plus largement. En revanche il fluctue de manière équivalente (bruit identique). Cela reflète le principe du schéma simplifié qui condense l'étape de traduction (faisant intervenir les molécules d'ARNs) du schéma détaillé, tout en gardant les ingrédients essentiels du point de vue stochastique. Sauf mention du contraire nous utilisons donc ce dernier dans la suite de notre travail.

Expérimentalement une visualisation intuitive, permettant une étude quantitative de l'expression du gène, est donnée par les histogrammes en intensité de fluorescence des protéines. Les simulations réalisées permettent de construire des histogrammes en nombre de protéines à n'importe quel instant précis du cycle,

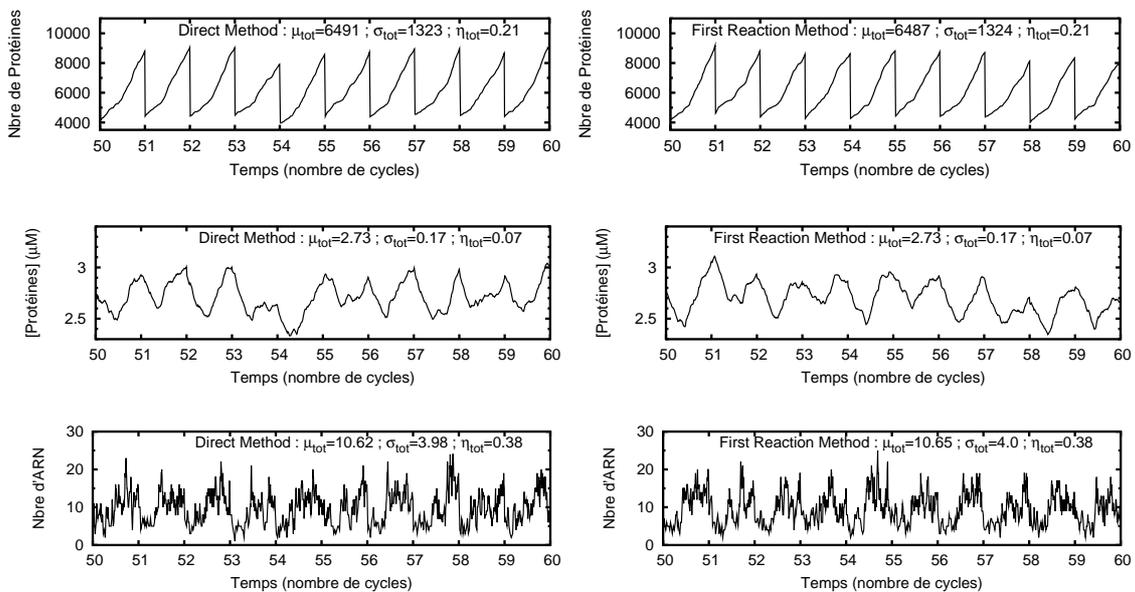


FIG. 8.2 – Extrait de la dynamique en nombre et concentration de protéines ainsi qu'en nombre d'ARNs, en employant le schéma simplifié (table 8.2). Les deux algorithmes utilisant la DM (gauche) et la FRM (droite) sont considérés.

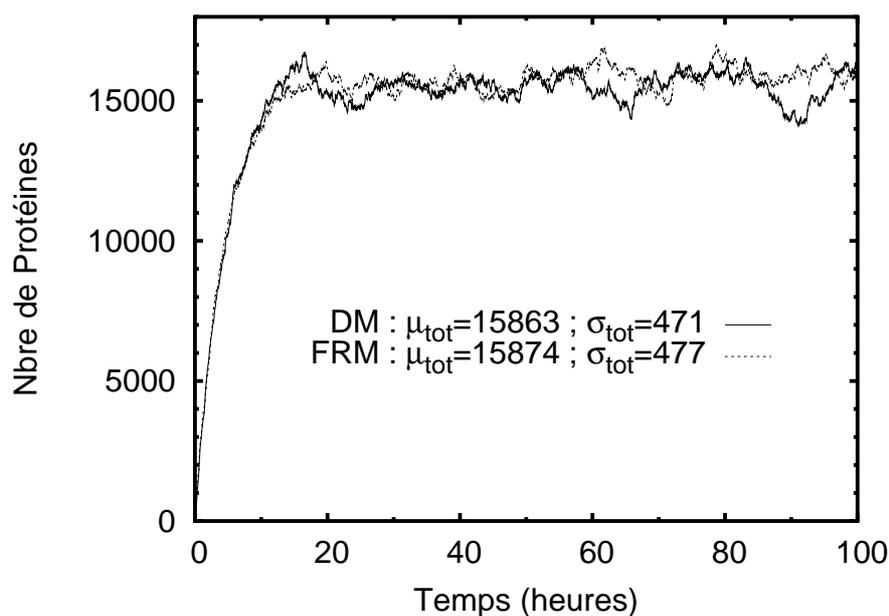


FIG. 8.3 – Le nombre de protéines atteint, au bout d’une cinquantaine d’heures, un état stationnaire pour un cycle de durée infinie. Le nombre d’ARNs (non montré) est quasi-stationnaire dès le début de l’expérience. Il est caractérisé par $\mu_{\text{tot}} = 6.27$, $\sigma_{\text{tot}} = 2.33$ et $\eta_{\text{tot}} = 0.38$ (DM comme FRM). Nous simulons le schéma simplifié (table 8.2.), avec les deux algorithmes de DM et FRM, sur 10^8 itérations soit une dynamique s’étalant sur environ 10000 h. Nous réalisons une mesure toutes les 100 itérations. Les bruits sont identiques $\eta_{\text{tot}} = 0.03$.

quitte à interpoler. Ils correspondent *grosso modo* aux courbes d’intensité⁹ à condition que les bactéries de la population soient synchronisées : si nous figeons pour réaliser une mesure d’intensité elles sont toutes à la même étape du cycle cellulaire. Si elles ne le sont pas, on peut espérer estimer l’intensité en moyennant le nombre de protéines sur chaque cycle de la simulation (figure 8.4).

⁹Il peut être intéressant de rajouter une réaction pour l’expression en fluorescence de la protéine, ou encore supprimer celle de dégradation pour avoir de l’accumulation, comme c’est le cas expérimentalement.

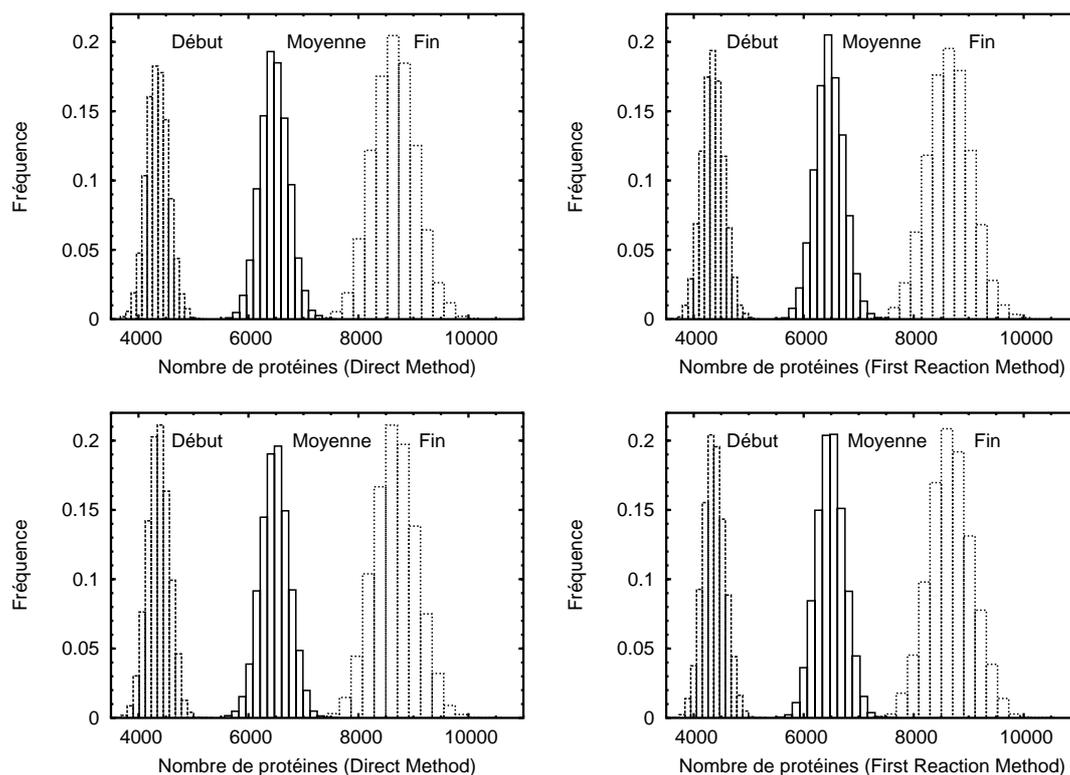


FIG. 8.4 – Histogrammes (15 canaux, 100 simulations de 10000 cycles.) en nombre de protéines en début et fin de cycle. La moyenne sur chaque cycle est également envisagée. Les schémas détaillé et simplifié (courbes du haut et du bas respectivement) donnent : $\mu \simeq 4300$ et $\sigma \simeq 200$ en début de cycle, $\mu \simeq 6500$ et $\sigma \simeq 250$ pour la moyenne sur le cycle, $\mu \simeq 8700$ et $\sigma \simeq 400$ en fin de cycle. Les bruits sont identiques $\eta \simeq 0.04$.

Chapitre 9

Contrôle du nombre de copies d'un plasmide

9.1 Généralités [SC02, SR01, Weg05]

Un *plasmide* est une séquence d'ADN extra-chromosomale qui « parasite » la bactérie.¹ Comme le chromosome, les plasmides sont en général circulaires. Ils sont de taille diverse mais sont globalement plus courts que le chromosome : de 1 à 100 kb contre environ 1 Mb. Si le chromosome contient l'information génétique essentielle au bon fonctionnement de la bactérie, les plasmides ne portent comparativement que peu de gènes, offrant « des plus » à leur hôte comme la résistance à certains antibiotiques. Le plasmide parasite la bactérie dans la mesure où l'expression de ses gènes se fait en utilisant les ressources de la bactérie. Ainsi, les étapes de transcription et de traduction, mais aussi la réplication des plasmides d'une bactérie constituent pour elle des poids supplémentaires à celui déjà exercé par le chromosome.

Une observation fondamentale est alors que d'une génération à l'autre, le nombre de plasmides – en début de cycle pour fixer les idées – reste en moyenne constant. Ce nombre est une caractéristique du plasmide considéré. On parle de son *nombre de copies*², que l'on note c . Remarque : on considère des situations où la cohabitation bactérie/plasmide est bien acceptée. Cela peut se produire si la bactérie ne contient qu'un type de plasmides, ou plus généralement des types de plasmides de *groupes d'incompatibilité* distincts. Par souci de simplicité nous ne considérerons que la première éventualité. La conservation de c implique l'existence de mécanismes de contrôle de la réplication au cours du cycle ainsi que de la répartition des exemplaires à la division. Ils permettent au plasmide de s'adapter aux fluctuations inévitables – par exemple sur la durée du cycle – et significatives d'un tel système constitué d'un petit nombre de molécules.

¹Attention : ce n'est pas un parasite (organisme) au sens usuellement entendu en biologie.

²Ne pas confondre avec le nombre d'exemplaires du plasmide à un instant donné.

Sans contrôle de sa réplication le plasmide risque soit d'écraser son hôte sous un poids trop élevé, soit d'être perdu suite à une division cellulaire³. On parle de mécanismes de *Contrôle du Nombre de Copies*, ou CNC. Au niveau microscopique, le mécanisme contrôlant la réplication est l'expression de séquences d'ADN qui avec l'*origine de réplication* (ori) constituent le *réplicon* du plasmide. Un CNC donné fixe c , mais la réciproque n'est *a priori* pas vraie. Selon que l'initiation de la réplication est activée par un ARN ou une protéine on qualifie le plasmide de *relâché* ou de *strict*.

Avant de décrire les principes de ces deux catégories de CNC, autorisons nous la digression suivante : comment caractériser un plasmide ? Historiquement la classification s'est faite au moyen du phénotype initialement observé. Ainsi le plasmide ColE1 exprime la protéine colicine E1, toxique pour les bactéries n'ayant pas ce plasmide. De même les plasmides apportant une résistance à leur hôte furent désignés avec des noms commençant par « R ». Comme un plasmide exprime en fait plusieurs gènes, ce système s'est vite révélé inadapté. Leur utilisation et modification massives ont posé la question légèrement différente : quelle(s) fonctionnalité(s) d'un plasmide nous intéresse(nt) ? Ils sont utilisés comme *vecteurs de clonage* pour étudier l'expression de gènes *a priori* non originels, dont les séquences sont insérées artificiellement : l'accent est alors mis sur leur nombre de copies, ou plus précisément sur leur CNC. Nous caractérisons donc avant tout un plasmide par son réplicon. Il est salutaire de mentionner l'abus de langage suivant : on confond parfois « plasmide » et « réplicon » de même que « réplicon » et « ori ». Offrons nous alors le luxe d'expliquer la convention actuelle de classification : les plasmides sont distingués par un « p » suivi de majuscules reprenant le plus souvent les initiales des découvreurs du plasmide dont le réplicon a été extrait ou des premiers utilisateurs du réplicon comme vecteur de clonage. Un nombre permet ensuite de différencier selon les gènes insérés. Par exemple Bolivar et Rodriguez utilisèrent le réplicon (de) ColE1 dans diverses constructions dont le très employé pBR322.

9.2 ColE1 est un plasmide relâché

La régulation de la réplication du plasmide ColE1 est du type relâché. L'amorce permettant l'accrochage de l'ADN polymérase dans le voisinage de l'ori est le résultat de l'hybridation entre le brin d'ADN et une partie de l'ARN II, dont la transcription commence 550 nt en amont de l'ori, se poursuit à travers pour se terminer environ 150 nt en aval (figure 9.1). Sans donner de détails, mentionnons que cette hybridation est rendue possible par le repliement de l'extrémité 5'

³On parle parfois de « guérison » d'une des bactéries filles. Afin d'éviter toute ambiguïté avec le sens classiquement admis en biologie on utilisera le mot *perte*.

de l'ARN en une structure secondaire spécifique. L'amorce est alors formée des 550 premiers nt de l'ARN, coupés du reste par une RNase. Les ARN II ne peuvent servir qu'à l'exemplaire de ColE1 les ayant transcrits : on parle de *cis*-action.

L'inhibition de l'initiation de la réplication est assurée par l'ARN I complé-

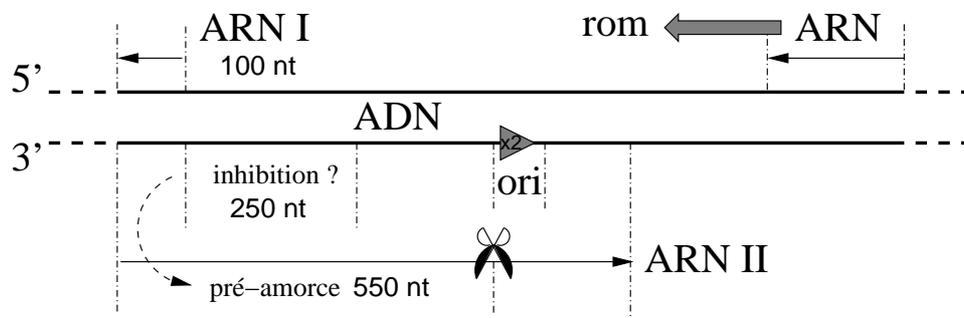


FIG. 9.1 – Réplicon du plasmide ColE1. L'ARN II participe à l'amorce potentielle de la réplication à l'ori (*cis*-action). La formation de l'amorce est inhibée efficacement lorsque l'ARN polymérase traverse une zone d'environ 250 nt, par complexation de l'ARN II transcrit avec l'ARN I complémentaire (*trans*-action). Le complexe est stabilisé par un dimère de protéines rom (*trans*-action).

mentaire des 100 premiers nt de l'ARN II car transcrit du brin d'ADN opposé (figure 9.1), ce qui leur permet de complexer. Là encore, sans en donner les raisons notons que cette complexation n'empêche le repliement convenable de l'ARN II de manière efficace que lorsqu'il est transcrit à hauteur de 100/350 nt, soit une fenêtre de transcription par l'ARN polymérase de 250 nt. Ce complexe d'ARNs peut être stabilisé⁴ par un dimère de protéines rom⁵, également exprimées par ColE1. Cette protéine ainsi que l'ARN I ne sont spécifiques que du type de plasmide (ColE1), et non de l'exemplaire : on parle de *trans*-action.

Ainsi, chaque exemplaire de ColE1 peut initier sa réplication en transcrivant de l'ARN II (activateur/*cis*-action). Parallèlement ColE1 transcrit de l'ARN I pouvant inhiber l'effet de l'ARN II (répresseur/*trans*-action). Sa demi-durée de vie de l'ordre de la minute rend sa concentration directement proportionnelle à celle en ColE1 (§§ 9.3.1 et 9.3.2) : l'ARN I quantifie le nombre d'exemplaires de ColE1⁶. Ce dernier exprime également des protéines rom facilitant cette inhibition entre ARNs (*trans*-action). Elles ne paraissent pas indispensables à la régulation de la réplication mais elles influent sur le nombre de copies, ce qui semble les confiner dans un rôle de catalyseur. Néanmoins des travaux théoriques [PE01, PNE98, GP98] indiquent que sous certaines conditions physiologiques ces protéines permettraient au CNC de répondre différemment selon que

⁴En anglais on parle du passage du « kiss » au « hug complex ».

⁵Certains auteurs préfèrent le nom de rop.

⁶En fait sa concentration dans la bactérie.

le nombre d'exemplaires du plasmide en début de cycle est faible ou non (§ 9.3.2). Les CNCs des plasmides relâchés sont très semblables à celui de ColE1 : activation par un ARN, réprimée par un ARN complémentaire avec l'aide d'une protéine. De tels CNCs conviennent aux réplicons de grands nombres de copies, qui correspondent à des plasmides originels courts car exerçant un faible poids sur la bactérie. Il suffit « au plasmide » d'inhiber toute réplication dès que son nombre d'exemplaires est en moyenne suffisant pour assurer la conservation de c à la division⁷. Pour finir signalons que le terme « relâché » s'entend par opposition au chromosome, qui nécessite des protéines (en plus des enzymes comme les polymérases, les RNases, etc.) pour initier sa réplication : ces plasmides n'ont fondamentalement besoin d'exprimer aucune protéine pour se répliquer.

9.3 Modélisation du CNC

9.3.1 Schémas hyperbolique/exponentiel

Nous commençons la modélisation de l'initiation de la réplication d'un plasmide relâché par le schéma naïf présenté en table 9.1. Les 100 premiers nt des

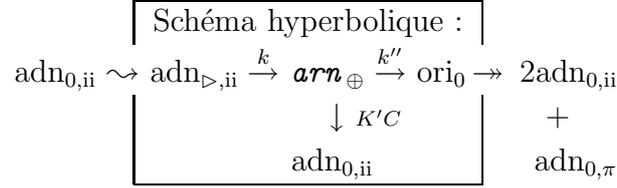
Transcriptions : $\text{adn}_{0,\text{ii}} \rightsquigarrow \text{adn}_{\triangleright,\text{ii}} \rightarrow \text{arn}_{\text{ii}} \rightarrow \text{arn}_{\oplus}$ $\quad \quad \quad \quad \quad \quad \quad \downarrow \quad \quad \quad \downarrow$ $\quad \quad \quad \quad \quad \quad \quad \emptyset \quad \quad \quad \emptyset$ $\text{adn}_{0,\text{i}} \rightsquigarrow \text{adn}_{\triangleright,\text{i}} \rightarrow \text{arn}_{\text{i}}$ $\quad \quad \quad \quad \quad \quad \quad \downarrow$ $\quad \quad \quad \quad \quad \quad \quad \emptyset$	Initiation réplication : <i>cis</i> -action $\text{arn}_{\oplus} + \text{ori} \mapsto \text{ori}_0 \rightarrow \text{adn}_{0,\dots}$ "...": ii, i, ϱ (ou π)
Dimérisation : $2\varrho \rightleftharpoons \varrho^2 \rightarrow \emptyset$	Inhibition : $\text{arn}_{\oplus} + \text{arn}_{\text{i}} \rightleftharpoons \text{arn}_{\heartsuit} \rightarrow \emptyset$ $\text{arn}_{\heartsuit} + \varrho^2 \rightarrow \text{arn}_{\text{?}} \rightarrow \emptyset$

TAB. 9.1 – Un premier schéma pour l'initiation de la réplication d'un plasmide relâché. L'initiation de la transcription « \rightsquigarrow » et l'expression de ϱ se modélisent selon la table 8.1. Le mécanisme de réplication « \rightarrow » n'est pour le moment pas précisé.

ARN I et II sont symbolisés par $\text{arn}_{\text{i,ii}}$ et sont transcrits de manière analogue à l' arn du schéma simplifié (table 8.1). L'état arn_{\oplus} désigne l'ARN II sensible à l'accrochage d'ARN I (polymérase transcrivant la fenêtre d'inhibition potentielle de 250 nt) : leur complexation réversible en arn_{\heartsuit} peut être stabilisée en $\text{arn}_{\text{?}}$ par un dimère ϱ^2 de protéines rom exprimées du gène $\text{adn}_{0,\varrho}$ suivant le schéma simplifié (non repris, voir table 8.1). Autrement l'ARN II participe à l'amorce ori_0

⁷Plus précisément d'assurer dynamiquement, c'est-à-dire à chaque instant du cycle un nombre d'exemplaires en rapport avec le volume de la bactérie.

permettant l'initiation de la réplication en l'ori de l'exemplaire l'ayant transcrit (schéma non précisé). En ne gardant que les étapes limitantes, ce schéma peut se réduire à celui qualifié d'« hyperbolique » de la table 9.2. Plus précisément,

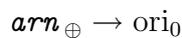


TAB. 9.2 – Schéma hyperbolique obtenu de la table 9.1, proposé pour l'initiation de la réplication d'un plasmide relâché. Si k'' ou $K'C \gg k$ il reste $\text{adn}_{\triangleright,\text{ii}} \rightarrow \text{ori}_0$ de constante kH dépendant hyperboliquement de la concentration en plasmides C .

nous négligeons les dégradations des \mathbf{arn}_{ii} et \mathbf{arn}_{\oplus} . Le premier donne quasi-instantanément le second, l'étape limitante étant l'initiation de la transcription. Quand au second, s'il ne participe pas à l'amorce, il est essentiellement dégradé par complexation avec \mathbf{arn}_{i} . La courte demi-durée de vie de l'ARN I rend à chaque instant sa concentration proportionnelle à celle en plasmide

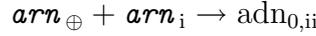
$$[\mathbf{arn}_{\text{i}}] \propto C = n/V,$$

que nous supposons observable. Par exemple en insérant le gène de la GFP (protéine notée π) dans le plasmide, à partir de la table 8.1 nous avons $C = [\text{adn}_{0,\pi}] + [\text{adn}_{\text{i},\pi}]$. Nous pouvons donc nous passer d'une modélisation explicite de la transcription de l'ARN I. Nous réexaminons plus tard (§ 9.3.2) la relation de proportionnalité en détail. Si \mathbf{arn}_{\oplus} n'est pas inhibé, la *cis*-action de l'ARN II implique qu'il ne peut participer qu'à l'amorce de la réplication de l'exemplaire l'ayant transcrit. La notation $\mathbf{arn}_{\oplus} + \text{ori} \mapsto \text{ori}_0$ n'est pas à prendre au sens strict. Dire que l'ARN II agit en *cis* signifie entre autre qu'il n'a pas le temps de diffuser « loin » de l'exemplaire l'ayant transcrit : il est quasi-instantanément transformé en amorce ou dégradé par complexation avec l'ARN I, et les \mathbf{arn}_{\oplus} ne sont donc pas distribués de manière homogène dans la bactérie. Nous pouvons décider qu' \mathbf{arn}_{\oplus} représente en fait un exemplaire tentant de se répliquer. Par conséquent, l'état $\text{adn}_{0,\text{ii}}$ ne doit être de nouveau créé qu'à la fin de cette tentative (qu'elle réussisse ou non). Nous pouvons alors noter sans ambiguïté



dont la constante traduit basiquement la transcription des 200 nt suivant la fenêtre d'inhibition potentielle. Pour finir nous considérons un large excès, quasi-constant en protéines rom sous forme dimérisée. Dans ces conditions elles assurent simplement la dégradation systématique des complexes d'ARNs formés. Ce rôle

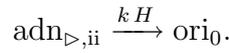
est également réinspecté plus tard (§ 9.3.2). Avec la nouvelle interprétation de l'état arn_{\oplus} cette dégradation se réécrit



qui termine de justifier le schéma présenté en table 9.1.⁸ Ce dernier est qualifié d'hyperbolique car la probabilité, par molécule d'ARN II transcrite, d'initier la réplication est hyperbolique en C la concentration en plasmide

$$H = \frac{k k''}{k k'' + k K' C} = \frac{1}{1 + \kappa C} \quad \text{avec} \quad \kappa = \frac{K'}{k''}.$$

Si l'étape limitante est l'initiation de la transcription de l'ARN II, c'est-à-dire que $k \ll k''$ ou $K' C$, nous pouvons alors complètement omettre l'état arn_{\oplus} et écrire



Ce modèle d'initiation de la réplication est-il finalement pertinent pour un plasmide relâché? En particulier nous avons dès le début condensé la transcription des 250 nt de la fenêtre d'inhibition potentielle, en l'unique état arn_{\oplus} . Ceci implique une durée de traversée de cette fenêtre exponentiellement distribuée avec le paramètre k'' (§ 8.4), c'est-à-dire susceptible de fluctuer largement. Remarquons qu'une distribution piquée autour de la même moyenne $1/k''$ correspond à un comportement exponentiel et non hyperbolique en C , de la probabilité d'initier la réplication par molécule d'ARN II transcrite. En effet, si la durée typique au bout de laquelle on observe l'inhibition est t_{\oplus} il lui correspond alors une énergie libre $(K' C \times t_{\oplus}) k_B T$, soit une probabilité $\exp(-K' C t_{\oplus})$ de ne pas inhiber avant t_{\oplus} . Si nous connaissons la distribution f de durée de la traversée, alors la probabilité d'initiation est :

$$\int_0^{\infty} \exp(-K' C t_{\oplus}) f(t_{\oplus}) dt_{\oplus}.$$

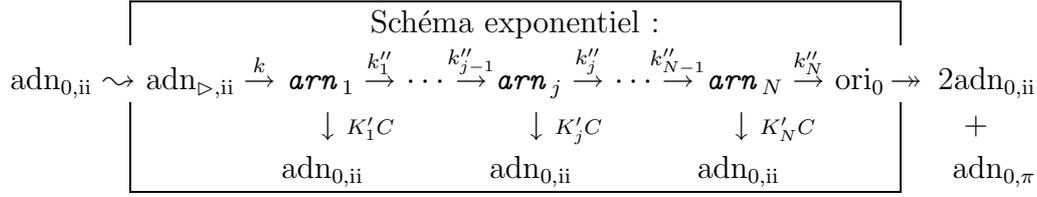
Une distribution exponentielle implique en effet un comportement hyperbolique :

$$\int_0^{\infty} \exp(-K' C t_{\oplus}) \varepsilon(t_{\oplus}; k'') dt_{\oplus} = H,$$

tandis que la suppression des fluctuations de t_{\oplus} autour de la moyenne $1/k''$ donne un comportement limite de type exponentiel :

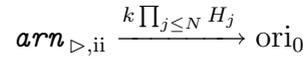
$$\int_0^{\infty} \exp(-K' C t_{\oplus}) \delta\left(t_{\oplus} - \frac{1}{k''}\right) dt_{\oplus} = \exp(-\kappa C) = E.$$

⁸La plupart des articles le prennent comme point de départ. Il se justifie en effet intuitivement, et une « dérivation » comme celle que nous venons d'exposer n'est ni plus rigoureuse ni nécessairement plus claire : le lecteur jugera.



TAB. 9.3 – Le comportement exponentiel de l’initiation de la réplication d’un plasmide relâché, est reproduit par un grand nombre d’étapes hyperboliques (table 9.2). Ce schéma se condense en $\text{adn}_{\triangleright,\text{ii}} \rightarrow \text{ori}_0$ de constante $k \exp(-\kappa C)$ si tous les rapports K'_j/k''_j sont pris égaux à κ/N et que $k \ll k''_j$ ou $K'_1 C$.

Un moyen d’obtenir ce comportement consiste à multiplier les états arn_{\oplus} d’inhibition potentielle en utilisant le schéma hyperbolique. Pour N d’entre elles le schéma exponentiel (table 9.3) se réduit là encore à l’unique réaction



à condition que l’initiation de la transcription de l’ARN II soit l’étape limitante, c’est-à-dire $k \ll k''_j$ ou $K'_j C$. Si nous choisissons les constantes de ces réactions telles que leurs rapports soient tous égaux à $\kappa_j = \kappa/N$ alors un grand nombre de ces étapes fournit le comportement exponentiel désiré, car

$$\prod_{j=1}^N H_j = \left(\frac{1}{1 + \kappa C/N} \right)^N \stackrel{(N \gg 1)}{\sim} \exp(-\kappa C) = E.$$

Nous pouvons penser à une étape pour chacun des 250 nt lus, mais aucune donnée biologique ne permet de justifier ce choix. Plus généralement nous pouvons voir ces étapes comme des pauses plus ou moins longues⁹ lors de la transcription de l’ARN II. Une valeur précise pour N a toutefois peu d’importance étant donné que dès $N = 5$ étapes, le comportement obtenu est déjà quasi-exponentiel (figure 9.2).

Pour finir, soulignons que c’est Ehrenberg qui en 1996 a fait le lien formel entre les deux comportements limites : hyperbolique et exponentiel. Son article [Ehr96] a permis de concilier deux modèles, jusque là opposés entre eux : celui de Brenner et Tomizawa [BT91] et celui de Brendel et Perelson [BP93].

9.3.2 Réponse à la concentration en plasmide

Avant de passer à la simulation de ce CNC, revenons sur la proportionnalité de la concentration en ARN I avec celle en plasmide. Nous considérons l’équation bilan

$$\frac{d[\text{arn}_i]}{dt} = k_i C(t) - d_i [\text{arn}_i] - k_V [\text{arn}_i]$$

⁹Nous les prendrons, pour simplifier, toutes égales en § 9.3.3.

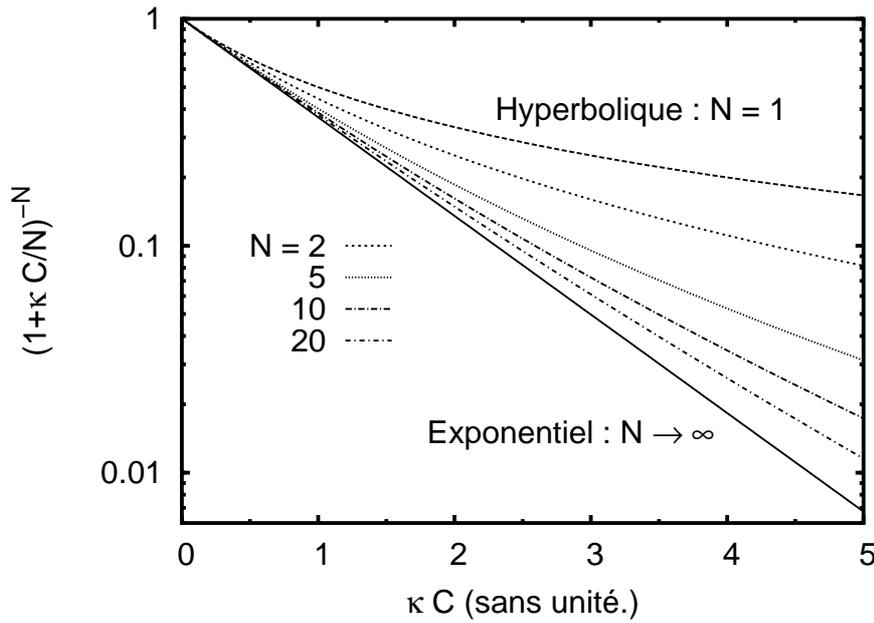


FIG. 9.2 – Dès $N = 5$ étapes hyperboliques ($N = 1$), on obtient un comportement quasi-exponentiel pour l'initiation de la réplication d'un plasmide relâché.

avec les deux premiers termes, de peuplement et de dépeuplement qui correspondent respectivement à la transcription d'ARN I et à sa dégradation. Le troisième et dernier terme diminue les concentrations par effet de dilution, dû à la croissance cellulaire $V(t)$. Cette dilution concerne toutes les espèces présentes

$$\frac{d}{dt} \left[\frac{N(t)}{V(t)} \right] = \frac{dN(t)/dt}{V(t)} - \frac{dV(t)/dt}{V(t)} \frac{N(t)}{V(t)}.$$

Le premier terme prend en compte la variation dans la concentration due à l'exécution de réactions seulement (variation en N), tandis que le second contient celle due à la croissance de la cellule seule (variation de V). Nous considérerons dans la suite de ce travail une croissance volumique exponentielle, bien que nous l'ayons précédemment considérée comme linéaire, en accord avec [SES02]. Ce choix se justifie par une approche différente du problème : nous idéalisons la population de bactérie comme une seule macro-bactérie. Nous voyons qu'avec ce choix

$$V(t) = V_0 \exp(\ln 2 t / \tau) \quad \text{avec} \quad k_V = \ln 2 / \tau.$$

Ceci étant dit, pour que le CNC soit efficace, il faut que la concentration en ARN I soit représentative de celle en plasmide. La diminution de la concentration doit donc être dominée par la dégradation et non la dilution : $d_i \gg k_V$. L'équation bilan devient l'équation différentielle du premier ordre à coefficient constant, avec second membre

$$\frac{d}{dt} [arn_i] + d_i [arn_i] = k_i C(t)$$

dont la solution générale est¹⁰

$$[\mathbf{arn}_i] = \alpha \exp(-d_i t) + k_i \int_0^t \exp[-d_i(t-x)] C(x) dx.$$

Supposons que sur l'intervalle de temps $[0; t]$, la concentration en plasmide soit quasi-constante. Alors

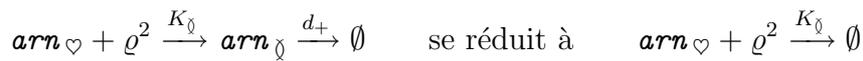
$$k_i \int_0^t \exp[-d_i(t-x)] C(x) dx \simeq \frac{k_i}{d_i} C(t) (1 - e^{-d_i t}).$$

Le terme exponentiel est négligeable aux grands $d_i \gg 1$. Pour que le rapport k_i/d_i reste borné nous nous plaçons également aux grands $k_i \gg 1$, et nous obtenons la proportionnalité

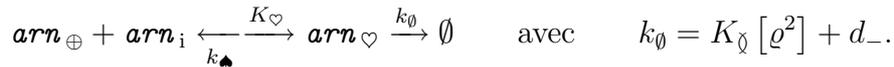
$$[\mathbf{arn}_i] \simeq \frac{k_i}{d_i} C(t).$$

Pour cela nous avons supposé une dégradation de l'ARN I, en rapport étroit avec sa transcription, et la plus rapide possible. Cela minimise les effets de mémoires (ARN I non dégradé et transcrit longtemps avant) et assure en effet une concentration la plus à jour possible avec celle en plasmide. Le cas idéal $k_i = d_i \rightarrow \infty$ n'est pas physique : en pratique la concentration en ARN I oscille autour de celle, optimale, suivant parfaitement la concentration en plasmide. Quiconque a déjà essayé de régler la température de sa douche, avec pour seule information la température de l'eau qu'il reçoit sur le moment peut assez bien comprendre ce genre d'oscillation ainsi que la problématique globale d'adapter une réponse à un signal.

Revenons pour finir sur le rôle de la protéine rom. La chaîne de réactions (nous reprenons le schéma de la table 9.1)



si l'étape limitante est la formation du complexe ($d_+ \gg K_\emptyset$). En supposant connue l'évolution de la concentration en dimères ϱ^2 , la dégradation de l'ARN II se réécrit



Là encore, si c'est la formation du complexe qui est l'étape limitante ($K_\heartsuit \gg k_\spadesuit$ ou k_\emptyset) nous pouvons condenser



¹⁰Somme de la solution générale de l'équation sans second membre, avec une solution particulière de celle avec second membre.

avec la constante K' qui dépend maintenant du temps à travers la concentration en dimères de protéines rom

$$K' = \alpha K_{\heartsuit} \times \underbrace{\frac{C_1 + [\varrho^2]}{C_2 + [\varrho^2]}}_{=P} \quad \text{où} \quad \begin{cases} \alpha = [\mathbf{arn}_i]/C & (\text{proportionnalité}) \\ C_1 = d_-/K_{\emptyset} \\ C_2 = (d_- + k_{\spadesuit})/K_{\emptyset}. \end{cases}$$

La définition des concentrations C_1 et C_2 montre que P est la probabilité, par ARN II transcrit, qu'une complexation entre ARNs se concrétise en inhibition de l'ARN II. Précédemment (§ 9.3.1), nous avons considéré un large excès en dimères de protéines rom, soit $[\varrho^2] \gg C_2$. Comme $C_2 > C_1$, nous vérifions que le rôle des dimères est alors bien d'assurer une dégradation quasi-systématique des complexes d'ARNs formés :

$$P = \frac{1 + C_1/[\varrho^2]}{1 + C_2/[\varrho^2]} \simeq 1$$

Qu'implique la réciproque ? Supposons que $C_2 \gg [\varrho^2]$, mais également que $[\varrho^2] \gg C_1$. Alors $k_{\spadesuit} \gg 1$, ce qui correspond aux situations de grande instabilité du complexe d'ARNs, et la probabilité de sa dégradation devient

$$P = \frac{[\varrho^2]}{C_2} \frac{1 + C_1/[\varrho^2]}{1 + [\varrho^2]/C_2} \simeq \frac{[\varrho^2]}{C_2} \simeq \frac{K_{\emptyset}}{k_{\spadesuit}} [\varrho^2].$$

Si les dimères de protéines rom répondent comme l'ARN I à la concentration en plasmides $[\varrho^2] \propto C$, la dégradation de l' \mathbf{arn}_{\oplus} se fait avec la constante

$$K'C \equiv \bar{K}'C^2,$$

répondant quadratiquement et non plus proportionnellement à C . Par conséquent la probabilité par molécule d'ARN II d'initier la réplication n'est plus hyperbolique en C :

$$\frac{k k''}{k k'' + k K'C} = \frac{1}{1 + \bar{\kappa}^2 C^2} \quad \text{avec} \quad \bar{\kappa}^2 = \frac{\bar{K}'}{k''}.$$

Il est donc possible d'imaginer des situations pour lesquelles la réplication est plus souvent initiée aux faibles concentrations en ColE1, ce qui réduirait fortement le risque de perte du plasmide.

En effet, si les n_{τ} exemplaires en fin de cycle sont répartis binomialement, de paramètre $p = 1/2$ entre les deux bactéries filles¹¹, l'une des deux peut perdre le plasmide avec une probabilité $1/2^{n_{\tau}}$ soit pour les deux $1/2^{n_{\tau}-1}$. En l'absence de mécanisme de répartition¹² de ses exemplaires à la division, le plasmide a donc tout « intérêt » à s'être répliqué le plus possible.

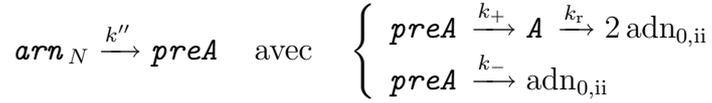
Réciproquement l'initiation est plus efficacement inhibée pour des nombres d'exemplaires suffisant à conserver c . Notons toutefois que de telles situations où la complexation des ARNs est très instable et la demi-durée de vie des dimères de protéines rom est courte, ne bénéficient pour l'instant d'aucun appui expérimental.

¹¹À la manière donc des molécules, en section 8.3.

¹²Nous précisons ces mécanismes en section (§ 9.4).

9.3.3 Constantes de réaction

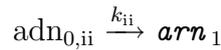
Nous reprenons le schéma exponentiel de la table 9.3. Nous précisons le processus de réplication : arrivé en arn_N , l'ARN II forme alors une pré-amorce $preA$ qui peut se concrétiser ou non, en l'amorce A qui initie vraiment la réplication. C'est-à-dire



Dans la mesure où c'est la formation de la pré-amorce qui est limitante ($k'' \ll k_{\pm}$) nous pouvons réécrire plus simplement



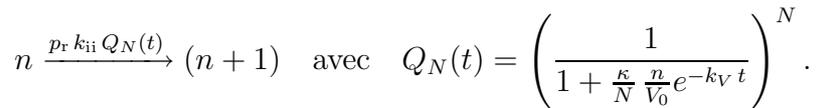
la probabilité d'initier la réplication. Nous pouvons également omettre la réplication à proprement parler, car $k_r \gg k'' p_r$. D'autre part, nous supposons que l'initiation de la transcription de l'ARN II est en fait directe



Enfin la forme contractée du schéma hyperbolique permet d'écrire



Nous avons donc supprimé tout les états intermédiaires, et $\text{ adn}_{0,ii}$ représente alors le plasmide lui-même. Nous sommes donc en train de décrire le processus de peuplement



Son équation bilan

$$\frac{dC(t)}{dt} = p_r k_{ii} Q_N(t) C(t) - k_V C(t)$$

admet l'état stationnaire

$$p_r k_{ii} \left(\frac{1}{1 + \kappa C_*/N} \right)^N = k_V \quad \text{soit} \quad C_* = \frac{N}{\kappa} \left[(p_r k_{ii}/k_V)^{1/N} - 1 \right].$$

Le paramètre déterminant la dynamique de réponse du CNC est en fait $\lambda = p_r k_{ii}/k_V$.¹³ Exiger de l'état stationnaire que $C_* = c/V_0$, où c est le nombre de copie du plasmide¹⁴, permet d'obtenir pour chaque valeur de λ une estimation de κ .

¹³Il réapparaît dans la solution formelle de l'équation maîtresse du système.

¹⁴En début de cycle par convention.

L'expression ci-dessus est valable quelque soit N , mais elle n'est pas très parlante. En guise d'illustration les cas hyperbolique ($N = 1$) et exponentiel ($N \rightarrow \infty$) correspondent respectivement à

$$\kappa = \frac{\lambda - 1}{c/V_0} \quad \text{et} \quad \kappa = \frac{\ln \lambda}{c/V_0}.$$

Nous pouvons *a posteriori* attribuer une valeur à chacune des constantes de réactions. Dans leur article [BP93], Brendel et Perelson considèrent un cycle cellulaire de $\tau = 48000$ s, soit 1 h et 20 min contre 1 h utilisée en section 8.3 et prise de [SES02]. La constante de dilution associée $1.4 \cdot 10^{-5} \text{ s}^{-1}$ est très faible. Toujours dans [BP93], la probabilité d'initier la réplication est prise $p_r = 1/2$ avec $k_+ = k_- = 4.3 \text{ min}^{-1}$, tandis qu'un plasmide¹⁵ d'environ 6000 bp lues par l'ADN polymérase à la vitesse¹⁶ de $500 \text{ nt} \cdot \text{s}^{-1}$ fournit $k_r = 8.333 \cdot 10^{-2} \text{ s}^{-1}$. Les 250 nt de la fenêtre d'inhibition sont lus par l'ARN polymérase à la vitesse de $50 \text{ nt} \cdot \text{s}^{-1}$, contre les $40 \text{ nt} \cdot \text{s}^{-1}$ de la section 8.3. Cela signifie un temps de transcription d'environ 5 s pour la fenêtre entière, valeur confirmée dans [Ehr96]. Dans [PE01], il est proposé l'encadrement $5 \leq \lambda \leq 100$. Une fois ce paramètre fixé nous pouvons en déduire k_{ii} , puis en précisant N (déjà nécessaire pour k'') et c nous obtenons K' .

L'équation maîtresse de ce schéma n'est pas plus résoluble analytiquement.¹⁷ Nous avons donc construit un algorithme de dynamique stochastique, semblable dans son principe à celui de la section 8.4. Nous avons choisit la FRM, et nous avons individuellement considéré les cas $N < \infty$ et $N \rightarrow \infty$. Dans le premier cas nous simulons l'ensemble des N étapes du schéma, tandis que le deuxième cas n'est réellement qu'un processus de peuplement.

Décrivons d'abord le cas $N < \infty$. À chaque instant de la dynamique nous générons autant de temps putatifs que de réactions, puis nous exécutons la « plus rapide ». Ces temps sont distribués exponentiellement comme en section 8.4 pour : l'initiation de la transcription de l'ARN II, le passage entre chacune des N étapes ainsi que la formation de l'amorce et l'initiation de la réplication. En revanche l'inhibition potentielle en chacune des N étapes se fait avec une constante de réaction dépendant du temps. À l'étape « j » l'inhibition a une propensité¹⁸ $a(N_j, n; t) = N_j K' C(t)$ et le temps putatif associé est alors distribué selon [GB00, GB00, LVTH04]

$$f(\theta | N_j, n; t) = a(N_j, n; t) \exp \left[- \int_t^{t+\theta} a(N_j, n; x) dx \right]$$

L'intégrale se calcule exactement, et la génération de la variable aléatoire θ selon cette distribution ne pose donc pas de problème majeur (méthode dite de

¹⁵C'est une estimation convenable au regard des 4361 bp du plasmide pBR322.

¹⁶Dans [ES95], Ehrenberg et Sverredal propose $1 \text{ kb} \cdot \text{s}^{-1}$.

¹⁷Sauf dans le cas hyperbolique où moyennant certaine condition nous retrouvons un processus de Poisson.

¹⁸Nous exigeons en fait plus que $\kappa_j = \kappa/N$ (§ 9.3.1) : nous prenons $K'_j = K'$ et $k''_j = k''/N$.

transformation [PTVF02]). Elle s'exprime

$$f(\theta|N_j, n; t) = \alpha(N_j, n) e^{-k_V(t+\theta)} \exp\left[\frac{\alpha(N_j, n)}{k_V} e^{-k_V(t+\theta)}\right] \exp\left[-\frac{\alpha(N_j, n)}{k_V} e^{-k_V t}\right]$$

en posant $\alpha(N_j, n) = K' N_j n / V_0$. Avant de passer au cas exponentiel, notons que l'approximation adiabatique [LVTH04] consistant à générer θ comme si la propensité ne dépendait pas du temps est, pour les paramètres considérés ici, totalement erronée.

Pour $N \rightarrow \infty$, l'unique réaction est celle de réplication du plasmide. La FRM est donc identique à la DM. À chaque instant de la dynamique, le temps de réaction doit être généré en prenant en compte la dépendance temporelle de la propensité

$$a(n; t) = n p_r k_{ii} \exp[-\kappa C(t)] = k_V n \lambda \exp\left[-\ln \lambda \frac{n}{c} e^{-k_V t}\right],$$

où nous avons introduit λ , en utilisant notamment l'estimation de κ pour le cas exponentiel. Explicitement, la distribution est alors

$$f(\theta|n; t) = k_V n \lambda \exp\left[-\ln \lambda \frac{n}{c} e^{-k_V(t+\theta)}\right] \times \\ \times \exp\left\{-k_V n \lambda \int_t^{t+\theta} \exp\left[-\ln \lambda \frac{n}{c} e^{-k_V x}\right] dx\right\}.$$

L'intégrale s'exprime comme la différence de deux fonctions spéciales (des intégrales exponentielles), mais n'est pas calculable analytiquement. Pour générer θ selon cette distribution nous utilisons la méthode dite de rejet (« rejection method » en anglais) [PTVF02]. Nous devons pour cela proposer une fonction majorante f . Cette fonction doit à la fois pouvoir s'exprimer simplement tout en approchant le plus possible la distribution f , et ce afin de conserver un temps de calcul raisonnable. Nous avons utilisé une lorentzienne

$$\frac{F}{1 + [(\theta - M)/S]^2}$$

dont les paramètres F , M et S dépendent *a priori* du temps t , du nombre d'exemplaires n ainsi que des constantes c , λ et k_V . L'optimisation de ces paramètres fait intervenir la fonction W de Lambert [Hay05]. Le générateur ainsi obtenu a été testé individuellement avec succès. Nous préférons revenir au problème physique, remettant à plus tard la description de la procédure. Signalons tout de même au lecteur le peu de succès là encore de l'approximation adiabatique.

9.4 Distribution conditionnelle

Plutôt que de chercher à calculer directement l'histogramme en nombre d'exemplaires du système à un instant donné, nous divisons le problème en deux parties.

Si $P_n(\alpha)$ est la probabilité d'observer α exemplaires (noté $n(0)$ sur les figures) en début du cycle n , elle s'exprime alors comme la composition

$$P_{n+1}(\gamma) = \sum_{\beta=0}^{\infty} D(\gamma|\beta) \sum_{\alpha=0}^{\infty} R(\beta|\alpha) P_n(\alpha)$$

avec

- la probabilité conditionnelle d'obtenir β exemplaires en fin de cycle, sachant qu'il y en avait α en début : $R(\beta|\alpha)$. Elle correspond au phénomène de répllication ;
- la probabilité conditionnelle d'hériter pour le début du cycle de γ exemplaires sur les β disponibles à la fin du cycle précédent : $D(\gamma|\beta)$. Elle correspond au mécanisme de répartition lors de la division cellulaire.

Notre algorithme nous permet de calculer R , en fixant une fois pour toute le nombre d'exemplaires α en début de cycle, puis en emmagasinant l'histogramme du nombre d'exemplaires β à la fin des cycles (figure 9.3). Nous avons fixé arbitrairement $\lambda = 50$ et considéré un nombre de copies $c = 10$ pouvant faire penser à celui du plasmide ColE1. Que ce soit pour le cas hyperbolique ($N = 10$) ou exponentiel, les histogrammes correspondant à chaque α considéré¹⁹, sont apparus convergés dès 10000 cycles (figure 9.4).

Nous pouvons alors préciser à notre guise le mécanisme de répartition D . La dynamique de cycle à cycle s'obtient alors comme

$$P_{n+1}(\gamma) = \sum_{\alpha=0}^{\infty} P_n(\alpha) \sum_{\beta=0}^{\infty} D(\gamma|\beta) R(\beta|\alpha) = \sum_{\alpha=0}^{\infty} T(\gamma|\alpha) P_n(\alpha).$$

En pratique D peut être une répartition 50/50 idéale, à la manière du chromosome

$$D(\gamma|\beta) = \delta_{\beta, 2\gamma},$$

situation toutefois peu réaliste. Nous pouvons considérer une loi binomiale, comme pour la répartition des molécules

$$D(\gamma|\beta) = B(\gamma; 1/2, \beta) = \frac{\beta!}{\gamma!(\beta - \gamma)!} \frac{1}{2^\beta}.$$

Nous avons déjà mentionné le risque de perte du plasmide (§ 9.3.2). Les plasmides du type relâché, dont fait partie ColE1 n'ont en général pas de système de partition (par^-). Leur tactique pour se maintenir repose sur leur relativement faible taille et leur nombre de copie élevé (§ 9.2). Réciproquement, les plasmides possédant un système de partition (par^+) sont plus longs tandis que leur nombre de copies est moindre. Un bon modèle de système par^+ est d'assurer l'héritage d'au moins un exemplaire à chaque cellule fille. Dans ce cas

¹⁹Nous avons couvert $1 \leq \alpha \leq 35$.

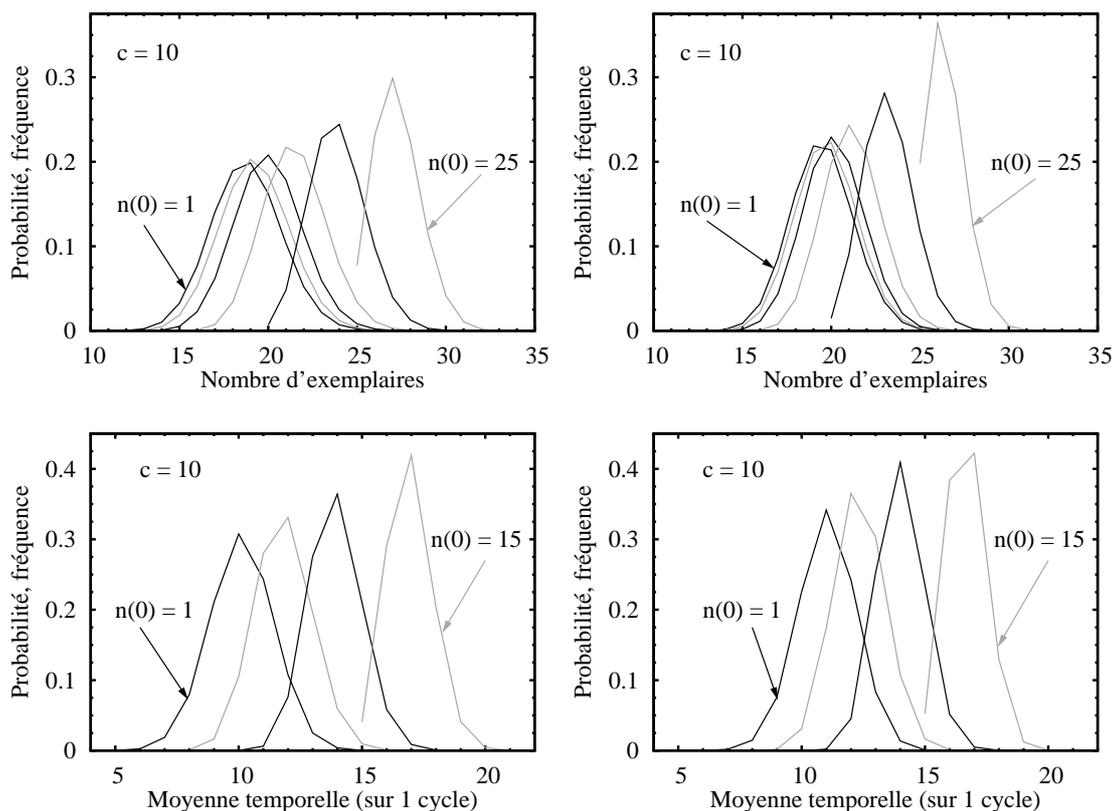


FIG. 9.3 – Haut : Distribution conditionnelle $R(\beta|\alpha)$ pour divers $\alpha = n(0)$. Elle correspond au phénomène de répliation. Selon que le nombre d'exemplaires en début de cycle est faible ou important, le CNC du plasmide active plus ou moins la répliation. Le schéma exponentiel (droite) assure un meilleur contrôle que l'hyperbolique ($N = 10$, gauche). Bas : nous pouvons aussi accumuler les histogrammes, de la moyenne temporelle du nombre d'exemplaires au cours du cycle (hyperbolique : gauche. Exponentiel : droite).

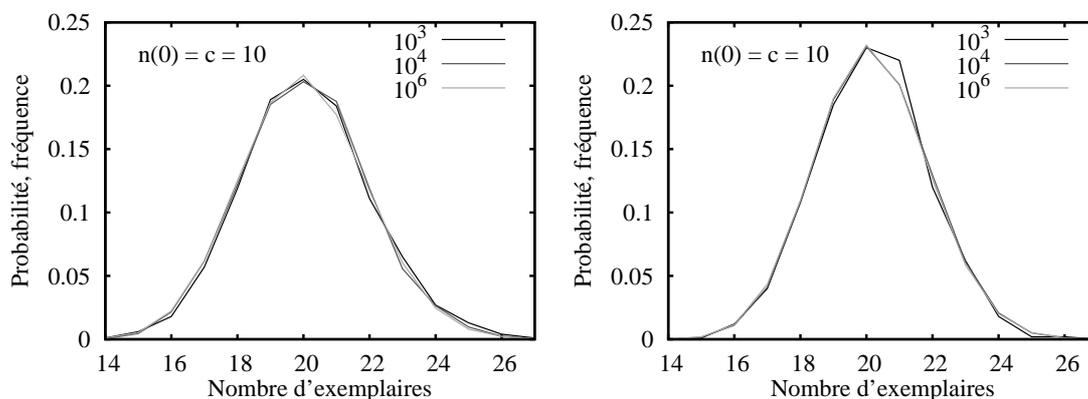


FIG. 9.4 – Convergence de la distribution conditionnelle R avec le nombre de cycles accumulés, illustrée pour $n(0) = c = 10$.

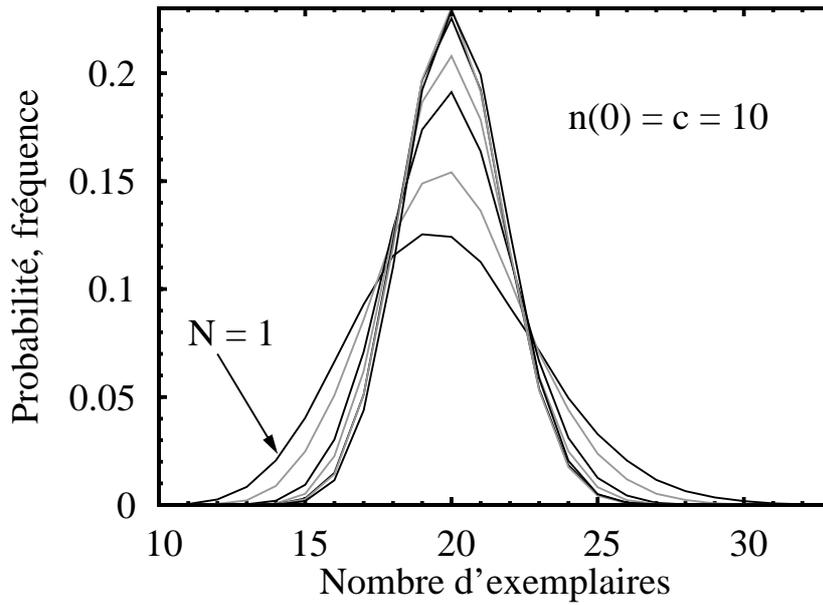


FIG. 9.5 – Un grand nombre d'étapes hyperboliques redonne le schéma exponentiel. Illustration sur la distribution conditionnelle R de $n(0) = c = 10$. Sont montrés, en partant du bas : $N = 1, 2, 5, 10, 100, 250$ et le résultat du schéma exponentiel.

$$D(\gamma|\beta) = B(\gamma - 1; 1/2, \beta - 2).$$

Fixant l'une de ces trois règles, pour les cas hyperbolique et exponentiel, nous disposons encore de deux méthodes pour évaluer P . Soit nous construisons un histogramme initial arbitraire, que nous composons successivement avec T , jusqu'à convergence de l'histogramme (état stationnaire) [Ehr96]. Ce dernier doit alors être renormalisé à chaque itération. Soit, partant d'un nombre d'exemplaires arbitraire, nous réalisons une dynamique stochastique avec la loi de probabilité T , tout en emmagasinant l'histogramme du nombre d'exemplaires à chaque « réaction » (changement de cycle en fait).

Ces deux méthodes produisent les mêmes résultats (figure 9.6), la première convergeant néanmoins beaucoup plus rapidement. Terminons sur une question : dans le cas du chromosome, ou du plasmide de répartition 50/50, chaque cellule fille hérite du même nombre d'exemplaires d'ADN. Nous pouvons donc suivre l'une des deux au hasard, sans risque de perdre de l'information. Dans le cas d'une répartition plus réaliste, comme la binomiale, il n'est pas *a priori* évident de pouvoir décrire les propriétés statistiques d'une population de bactérie en observant au cours du temps une seule lignée, descendance. La réponse se trouve dans un article de Seneta et Tavaré [ST83], publié en 1982. Il y est démontré que les cellules de lignés différentes ne sont pas corrélées.

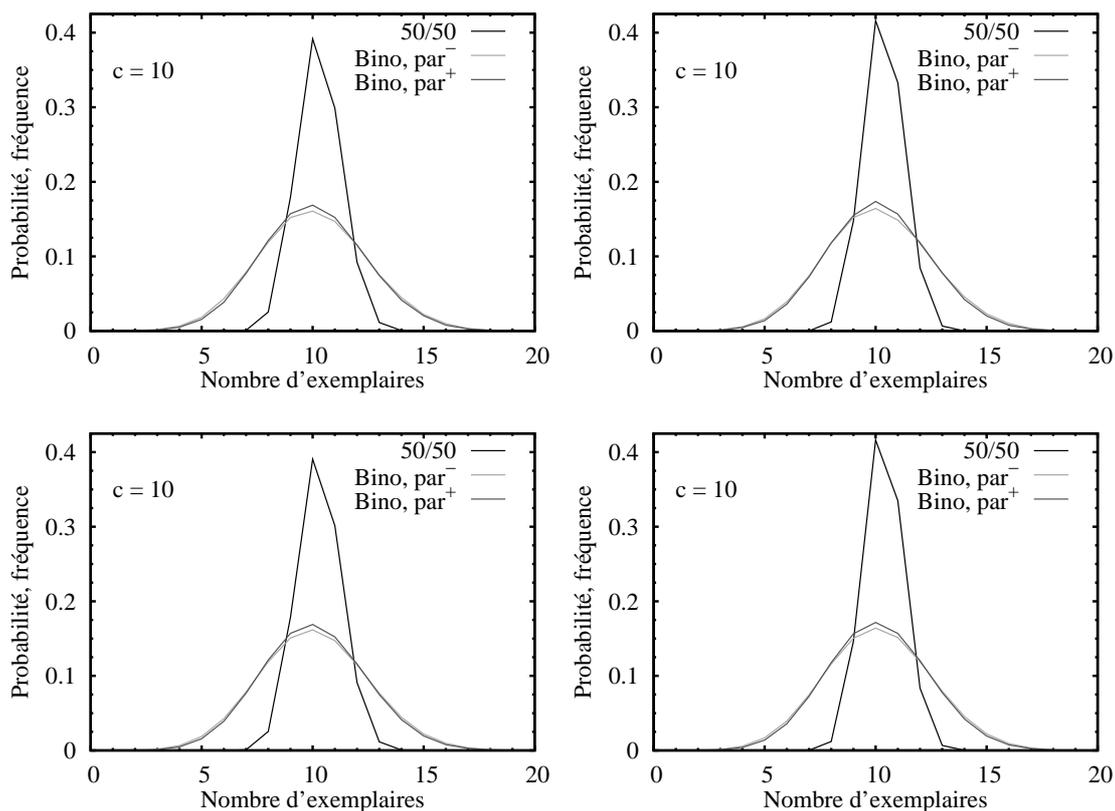


FIG. 9.6 – Obtention de la distribution du nombre d'exemplaires du plasmide en début de cycle, par composition de R et D . Pour D , sont considérés une répartition 50/50, binomiale (par^{\pm}). Les schémas hyperbolique (gauche) et exponentiel (droite) sont considérés. Bas : distribution stationnaire obtenue par itération de la composition jusqu'à convergence. Haut : accumulation de l'histogramme du nombre d'exemplaires au cours de simulations de dynamique stochastique.

Chapitre 10

Perspectives : signal plasmide/chromosome

Une perspective théorique déjà mentionnée est la caractérisation de la robustesse au bruit de ce réseau génétique : comment le bruit s'y propage t'il ? Pour revenir aux expériences réalisées dans le groupe de Didier Chatenay, il serait intéressant d'inclure dans un même code de dynamique stochastique les plasmides et le chromosome, exprimant chacun une protéine différente (de couleur de fluorescence différente typiquement). Le chromosome sert alors en quelque sorte d'étalonnage au bruit, et permet de déduire le nombre d'exemplaires du plasmide par simple rapport des deux signaux de fluorescence. Enfin, mentionnons l'existence d'un modèle pour le CNC du plasmide R1 [ES95]. Bien qu'expérimentalement les dérivés du ColE1 (comme le pBR322) soient les plus fréquents, l'étude d'un autre CNC peut se révéler instructive.

Bibliographie

- [AA97] T. Aki and S. Adhya, *Repressor Induced Site-Specific Binding of HU for Transcriptional Regulation*, The EMBO Journal **16** (1997), 3666.
- [ACA96] T. Aki, H. E. Choy, and S. Adhya, *Histone-Like Protein HU as a Specific Transcriptional Regulator : Co-Factor Role in Repression of Gal Transcription by Gal Repressor*, Genes to Cells **1** (1996), 179.
- [Ans] A. Ansari, *Physics 450 : Molecule and Cell Biophysics*, Consultable sur <http://tigger.uic.edu/classes/phys/phys461/phys450/>.
- [ARM98] A. Arkin, J. Ross, and H. H. McAdams, *Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage λ -Infected Escherichia Coli Cells*, Genetics **149** (1998), 1633.
- [AS70] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, Dover, 1970, Consultable sur <http://www.convertit.com>.
- [BBS⁺00] C. G. Baumann, V. A. Bloomfield, S. B. Smith, C. Bustamante, M. D. Wang, and S. M. Block, *Stretching of Single Collapsed DNA Molecules*, Biophysical Journal **78** (2000), 1965.
- [BF85] M. G. Bawendi and K. F. Freed, *A Wiener Integral Model for Stiff Polymer Chains*, Journal of Chemical Physics **83** (1985), 2491.
- [BKMS04] A. Balaeff, C. R. Koudella, L. Mahadevan, and K. Schulten, *Modeling DNA Loops Using Continuum and Statistical Mechanics*, Philosophical Transactions of the Royal Society of London A **362** (2004), 1355.
- [BKR02] I. Bose, R. Karmakar, and S. Roy, *Stochastic Simulation of Gene Expression in a Single Cell*, cond-mat/0210183 (2002).
- [BM98] C. Bouchiat and M. Mézard, *Elasticity Model of a Supercoiled DNA Molecule*, Physical Review Letters **80** (1998), 1556.

- [BM00] ———, *Elastic Rod Model of a Supercoiled DNA Molecule*, The European Physical Journal E **2** (2000), 377.
- [BM02] ———, *Reply to the Comment on : “elasticity Model of a Supercoiled DNA Molecule”*, Physical Review Letters **88** (2002), 089802–1.
- [BMRJ88] G. R. Bellomy, M. C. Mossing, and M. T. Record Jr, *Physical Properties of DNA in Vivo as Probed by the Length Dependence of the Lac Operator Looping Process*, Biochemistry **27** (1988), 3900.
- [BMS94] C. Bustamante, J. F. Marko, and E. D. Siggia, *Entropic Elasticity of λ -Phage DNA*, Science **265** (1994), 1599.
- [BMS99] A. Balaeff, L. Mahadevan, and K. Schulten, *Elastic Rod Model of a DNA Loop in the Lac Operon*, Physical Review Letters **83** (1999), 4900.
- [BMS03] ———, *Modeling DNA Loops Using the Theory of Elasticity*, physics/0301006 (2003).
- [BMS04] ———, *Structural Basis for Cooperative DNA Binding by CAP and Lac Repressor*, Structure **12** (2004), 123.
- [BMW00] C. Bustamante, J. C. Macosko, and G. J. L. Wuite, *Grabbing the Cat by the Tail : Manipulating Molecules One by One*, Nature Reviews, Molecular Cell Biology **1** (2000), 130.
- [BP93] V. Brendel and A. S. Perelson, *Quantitative Model of ColE1 Plasmid Copy Number Control*, Journal of Molecular Biology **229** (1993), 860.
- [BPJ91] M. Brenowitz, A. Pickar, and E. Jamison, *Stability of a Lac Repressor Mediated “Looped Complex”*, Biochemistry **30** (1991), 5986.
- [BS00] D. A. Beard and T. Schlick, *Modeling Salt-Mediated Electrostatics of Macromolecules : The Discrete Surface Charge Optimization Algorithm and Its Application to the Nucleosome*, Biopolymers **58** (2000), 106.
- [BSBB97] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, *Ionic Effects on the Elasticity of Single DNA Molecules*, Proceedings of the National Academy of Sciences **94** (1997), 6185.
- [BT91] M. Brenner and J.-I. Tomizawa, *Quantization of ColE1-Encoded Replication Elements*, Proceedings of the National Academy of Sciences **88** (1991), 405.
- [BTB97] J. K. Bhattacharjee, D. Thirumalai, and J. D. Bryngelson, *Distribution Function of the End-to-End Distance of Semiflexible Polymers*, cond-mat/9709345 (1997).

- [BTM05] S. Blumberg, A. V. Tkachenko, and J.-C. Meiners, *Disruption of Protein-Mediated DNA Looping by Tension in the Substrate DNA*, *Biophysical Journal* **88** (2005), 1692.
- [BWA⁺99] C. Bouchiat, M. D. Wang, J.-F. Allemand, T. Strick, S. M. Block, and V. Croquette, *Estimating the Persistence Length of a Worm-Like Chain Molecule From Force-Extension Measurements*, *Biophysical Journal* **76** (1999), 409.
- [Car04] G. A. Carri, *Statistical Mechanics of Worm-Like Polymers From a New Generating Function. (II) The Force-Elongation Relationship*, *Journal of Chemical Physics* **121** (2004), 12112.
- [CHN90] D. M. Crothers, T. E. Haran, and J. G. Nadeau, *Intrinsically Bent DNA*, *Journal of Biological Chemistry* **265** (1990), 7093.
- [CL05] B. Chakrabarti and A. J. Levine, *Nonlinear Elasticity of an α -Helical Polypeptide*, *Physical Review E* **71** (2005), 031905–1.
- [CM04] G. A. Carri and M. Marucho, *Statistical Mechanics of Worm-Like Polymers From a New Generating Function*, *Journal of Chemical Physics* **121** (2004), 6064.
- [CP02] L. Cannavecciuolo and J. S. Pedersen, *Properties of Polyelectrolyte Chains From Analysis of Angular Correlation Functions*, *Journal of Chemical Physics* **117** (2002), 8973.
- [CP04] ———, *Moments and Distribution Function of Polyelectrolyte Chains*, *Journal of Chemical Physics* **120** (2004), 8862.
- [CW04] T. E. Cloutier and J. Widom, *Spontaneous Sharp Bending of Double-Stranded DNA*, *Molecular Cell* **14** (2004), 355.
- [CW05] ———, *DNA Twisting Flexibility and the Formation of Sharply Looped Protein-DNA Complexes*, *Proceedings of the National Academy of Sciences* **102** (2005), 3645.
- [CYL⁺04] S. Cocco, J. Yan, J.-F. Léger, D. Chatenay, and J. F. Marko, *Overstretching and Force-Driven Strand Separation of Double-Helix DNA*, *Physical Review E* **70** (2004), 011910–1.
- [DA00] H. Diamant and D. Andelman, *Binding of Molecules to DNA and Other Semiflexible Polymers*, *Physical Review E* **61** (2000), 6740.
- [Dan52] H. E. Daniels, *The Statistical Theory of Stiff Chains*, *Proceedings of the Royal Society of Edinburgh A* **63** (1952), 290.
- [DC02a] A. Dua and B. J. Cherayil, *The Dynamics of Chain Closure in Semiflexible Polymers*, *Journal of Chemical Physics* **116** (2002), 399.
- [DC02b] ———, *The Thermodynamics of Reversible Cyclization in Semiflexible Polymers*, *Journal of Chemical Physics* **117** (2002), 7765.

- [DCJP03] N. Douarche, F. Calvo, P. J. Jensen, and G. M. Pastor, *Model Simulations of Ground-State and Finite-Temperature Properties of Disordered Magnetic Nanostructure*, The European Physical Journal D **24** (2003), 77.
- [DG02] R. T. Dame and N. Goosen, *HU : Promoting Or Counteracting DNA Compaction ?*, Federation of European Biochemical Societies Letters **529** (2002), 151.
- [DMR01] D. T. F. Dryden, N. E. Murray, and D. N. Rao, *Nucleoside Triphosphate-Dependent Restriction Enzymes*, Nucleic Acids Research **29** (2001), 3728.
- [Dou02] N. Douarche, *Effets d'un champ magnétique intense sur l'auto-organisation de nanoparticules ferromagnétiques*, Master's thesis, Université Paul Sabatier (Toulouse III), 2002.
- [DRDK02] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner, *Capturing Chromosome Conformation*, Science **295** (2002), 1306.
- [DSS⁺05] Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskiaia, and A. Vologodskii, *Cyclization of Short DNA Fragments and Bending Fluctuations of the Double Helix*, Proceedings of the National Academy of Sciences **102** (2005), 5397.
- [DVK⁺05] Q. Du, M. Vologodskiaia, H. Kuhn, M. Frank-Kamenetskii, and A. Vologodskii, *Gapped DNA and Cyclization of Short DNA Fragments*, Biophysical Journal **88** (2005), 4137.
- [DvML⁺05] R. T. Dame, J. van Mameren, M. S. Luijsterburg, M. E. Mysiak, A. Janićijević, G. Pazdzior, P. C. van der Vliet, C. Wyman, and G. J. L. Wuite, *Analysis of Scanning Force Microscopy Images of Protein-Induced DNA Bending Using Simulations*, Nucleic Acids Research **33** (2005), e68.
- [Ehr96] M. Ehrenberg, *Hypothesis : Hypersensitive Plasmid Copy Number Control of ColE1*, Biophysical Journal **70** (1996), 135.
- [ES95] M. Ehrenberg and A. Sverredal, *A Model for Copy Number Control of the Plasmid R1*, Journal of Molecular Biology **246** (1995), 472.
- [FG95] L. Finzi and J. Gelles, *Measurement of Lactose Repressor-Mediated Loop Formation and Breakdown in Single DNA Molecules*, Science **267** (1995), 378.
- [FH65] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals*, McGraw-Hill, 1965.
- [FKW98] E. Frey, K. Kroy, and J. Wilhelm, *Physics of Solutions and Networks of Semiflexible Macromolecules and the Control of Cell Function*, cond-mat/9808022 (1998).

- [FKWS98] E. Frey, K. Kroy, J. Wilhelm, and E. Sackmann, *Lecture 9. Statistical Mechanics of Semiflexible Polymers : Theory and Experiment*, cond-mat/9707021 (1998).
- [Fre71] K. F. Freed, *Wiener Integrals and Models of Stiff Polymer Chains*, *Journal of Chemical Physics* **54** (1971), 1453.
- [GA02] M. Geanacopoulos and S. Adhya, *Genetic Analysis of GalR Tetramerization in DNA Looping During Repressosome Assembly*, *Journal of Biological Chemistry* **277** (2002), 33148.
- [GB00] M. A. Gibson and J. Bruck, *Efficient Exact Stochastic Simulation of Chemical Systems With Many Species and Many Channels*, *Journal of Physical Chemistry A* **104** (2000), 1876.
- [GCM01] K. Ghosh, G. A. Carri, and M. Muthukumar, *Configurational Properties of a Single Semiflexible Polyelectrolyte*, *Journal of Chemical Physics* **115** (2001), 4367.
- [Gib00] M. A. Gibson, *Computational Methods for Stochastic Biological Systems*, Ph.D. thesis, California Institute of Technology, 2000.
- [Gil97] D. T. Gillespie, *Exact Stochastic Simulation of Coupled Chemical Reactions*, *Journal of Physical Chemistry* **81** (1997), 2340.
- [Gol80] H. Goldstein, *Classical Mechanics*, 2ème ed., Addison Wesley, 1980.
- [GP98] P. J. E. Goss and J. Peccoud, *Quantitative Modeling of Stochastic Systems in Molecular Biology by Using Stochastic Petri Nets*, *Proceedings of the National Academy of Sciences* **95** (1998), 6750.
- [GVZA01] M. Geanacopoulos, G. Vasmatzis, V. B. Zhurkin, and S. Adhya, *Gal Repressosome Contains an Antiparallel DNA Loop*, *Nature Structural Biology* **8** (2001), 432.
- [GYSM72] W. Gobush, H. Yamakawa, W. H. Stockmayer, and W. S. Magee, *Statistical Mechanics of Wormlike Chains. (I) Asymptotic Behavior*, *Journal of Chemical Physics* **57** (1972), 2839.
- [Har66] S. Haroche, *Forme lagrangienne de la mécanique quantique*, Consultable sur <http://www.phys.ens.fr/cours/notes-de-cours/cct-dea/>, 1966.
- [Hay05] B. Hayes, *Why W?*, *American Scientist* **93** (2005), 104.
- [HH66] R. A. Harris and J. E. Hearst, *On Polymer Dynamics*, *Journal of Chemical Physics* **44** (1966), 2595.
- [HH67] J. E. Hearst and R. A. Harris, *On Polymer Dynamics. (III) Elastic Light Scattering*, *Journal of Chemical Physics* **46** (1967), 398.

- [HH05] M. Hochbruck and M. E. Hochstenbach, *Subspace Extraction for Matrix Functions*, Consultable sur la page personnelle <http://www.case.edu/artsci/math/hochstenbach/>, 2005.
- [HHB66] R. A. Harris, J. E. Hearst, and E. Beals, *On Polymer Dynamics (II)*, Journal of Chemical Physics **45** (1966), 3106.
- [HJK04] B. Hamprecht, W. Janke, and H. Kleinert, *End-to-End Distribution Function of Two-Dimensional Stiff Polymers for All Persistence Lengths*, Physics Letters A **330** (2004), 254.
- [HK05] B. Hamprecht and H. Kleinert, *End-to-End Distribution Function of Stiff Polymers for All Persistence Lengths*, Physical Review E **71** (2005), 031803.
- [HL97] M. Hochbruck and C. Lubich, *On Krylov Subspace Approximations to the Matrix Exponential Operators*, Society for Industrial and Applied Mathematics Journal on Numerical Analysis **34** (1997), 1911.
- [HM03] A. Hanke and R. Metzler, *Entropic Loss in Long-Distance DNA Looping*, Biophysical Journal **85** (2003), 167.
- [HMS88] L. Huo, K. J. Martin, and R. Schleif, *Alternative DNA Loops Regulate the Arabinose Operon in Escherichia Coli*, Proceedings of the National Academy of Sciences **85** (1988), 5444.
- [HP01] P. L. Hansen and R. Podgornik, *Wormlike Chains in the Large- d Limit*, Journal of Chemical Physics **114** (2001), 8637.
- [HR90] P. J. Hagerman and V. A. Ramadevi, *Application of the Method of Phage T4 DNA Ligase-Catalited Ring-Closure to the Study of DNA Structure*, Journal of Molecular Biology **212** (1990), 351.
- [HS62] J. E. Hearst and W. H. Stockmayer, *Sedimentation Constants of Broken Chains and Wormlike Coils*, Journal of Chemical Physics **37** (1962), 1425.
- [HSB99] M. Hegner, S. B. Smith, and C. Bustamante, *Polymerization and Mechanical Properties of Single RecA-DNA Filaments*, Proceedings of the National Academy of Sciences **96** (1999), 10109.
- [HSPP99] P. L. Hansen, D. Sventek, V. A. Parsegian, and R. Podgornik, *Buckling, Fluctuations, and Collapse in Semiflexible Polyelectrolytes*, Physical Review E **60** (1999), 1956.
- [HT95] B.-Y. Ha and D. Thirumalai, *A Mean-Field Model for Semiflexible Chains*, Journal of Chemical Physics **103** (1995), 9408.
- [HT99] ———, *Persistence Length of Flexible Polyelectrolyte Chains*, Journal of Chemical Physics **110** (1999), 7533.

- [HT03] ———, *Bending Rigidity of Stiff Polyelectrolyte Chains : A Single Chain and a Bundle of Multichains*, *Macromolecules* **36** (2003), 9658.
- [HU52] J. J. Hermans and R. Ullman, *The Statistics of Stiff Chains, With Applications to Light Scattering*, *Physica* **18** (1952), 951.
- [HWMW87] W.-T. Hsieh, P. A. Whitson, K. S. Matthews, and R. D. Wells, *Influence of Sequences and Distance Between Two Operators on Interaction With the Lac Repressor*, *Journal of Biological Chemistry* **262** (1987), 14583.
- [JAAB⁺01] B. M. Jaffar Ali, R. Amit, I. Braslavsky, A. B. Oppenheim, O. Gileadi, and J. Stavans, *Compaction of Single DNA Molecules Induced by Binding of Integration Host Factor (IHF)*, *Proceedings of the National Academy of Sciences* **98** (2001), 10658.
- [Jac98] J. D. Jackson, *Classical Electrodynamics*, 3rd ed., Wiley, 1998.
- [JBH03] S. Jun, J. Bechhoefer, and B.-Y. Ha, *Diffusion-Limited Loop Formation of Semiflexible Polymers : Kramers Theory and the Interwined Time Scales of Chain Relaxation and Closing*, *Europhysics Letters* **64** (2003), 420.
- [JHBB04] S. Jun, J. Herrick, A. Bensimon, and J. Bechhoefer, *Persitence Length of Chromatin Determines Origin Spacing in Xenopus Early-Embryo DNA Replication : Quantitative Comparisons Between Theory and Experiment*, *Cell Cycle* **3** (2004), 223.
- [Jun04] S. Jun, *Kinetic Model of DNA Replication and the Looping of Semiflexible Polymers*, Ph.D. thesis, Simon Fraser University, 2004.
- [KA01] S. Kar and S. Adhya, *Recruitment of Hu by Piggyback : A Special Role of GalR in Repressosome Assembly*, *Genes and Development* **15** (2001), 2273.
- [KBAG98] A. L. Kholodenko, M. Ballauff, and M. Agüero Granados, *Conformational Statistics of Semiflexible Polymers : Comparison Between Different Models*, *Physica A* **260** (1998), 267.
- [KC98] J. D. Kahn and D. M. Crothers, *Measurement of the DNA Bend Angle Induced by the Catabolite Activator Protein Using Monte Carlo Simulation of Cyclization Kinetics*, *JMB* **276** (1998), 287.
- [KC05] H. Kleinert and A. Chervyakov, *Perturbation Theory for Path Integrals of Stiff Polymers*, cond-mat/0503199 (2005).
- [KEL⁺01] A. N. Kapanidis, Y. W. Ebricht, R. D. Ludescher, S. Chan, and R. H. Ebricht, *Mean DNA Bend Angle and Distribution of DNA Bend Angles in the CAP-DNA Complex in Solution*, *Journal of Molecular Biology* **312** (2001), 453.

- [Kho93] A. L. Kholodenko, *Analytical Calculation of the Scattering Function for Polymers of Arbitrary Flexibility Using the Dirac Propagator*, *Macromolecules* **26** (1993), 4179.
- [KL04] K. V. Klenin and J. Langowski, *Modeling of Intramolecular Reactions of Polymers : An Efficient Method Based on Brownian Dynamics Simulations*, *Journal of Chemical Physics* **121** (2004), 4951.
- [Kle04] H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics, and Financial Markets*, 4th ed., World Scientific, 2004, Consultable sur <http://www.physik.fu-berlin.de/kleinert/>.
- [KML98] K. V. Klenin, H. Merlitz, and J. Langowski, *A Brownian Dynamics Program for the Simulation of Linear and Circular DNA and Other Wormlike Chain Polyelectrolytes*, *Biophysical Journal* **74** (1998), 780.
- [KML⁺05] I. M. Kulić, H. Mohrbach, V. Lobaskin, R. Thakkar, and H. Schiessel, *Apparent Persistence Length Renormalization of Bent DNA*, *Physical Review E* **72** (2005), 041905–1.
- [KNA⁺87] H. Krämer, M. Niemöller, L. Amouyal, B. Revet, B. von Wilcken-Bergmann, and B. Müller-Hill, *Lac Repressor Forms Loops With Linear DNA Carrying Two Suitably Spaced Lac Operators*, *The EMBO Journal* **6** (1987), 1481.
- [KP49] O. Kratky and G. Porod, *Röntgenuntersuchung gelöster Fadenmoleküle*, *Recueil des Travaux Chimiques des Pays-Bas* **68** (1949), 1106.
- [KS03] I. M. Kulić and H. Schiessel, *Nucleosome Repositioning Via Loop Formation*, *Biophysical Journal* **84** (2003), 3197.
- [Kul04] I. Kulić, *Statistical Mechanics of Protein Complexed and Condensed DNA*, Ph.D. thesis, Johannes Gutenberg-Universität, 2004.
- [KV94] A. L. Kholodenko and T. Vilgis, *Elastic Response of the Dirac Chain*, *Physical Review E* **50** (1994), 1257.
- [Lad71] F. Lado, *Numerical Fourier Transforms in One, Two, and Three Dimensions for Liquid State Calculations*, *Journal of Computational Physics* **8** (1971), 417.
- [Lan86] C. Lanczos, *The Variational Principle of Mechanics*, 4th ed., Dover, 1986.
- [lB79] M. le Bret, *Catastrophic Variation of Twist and Writhing of Circular DNAs With Constraint ?*, *Biopolymers* **18** (1979), 1709.

- [LBC⁺03] G. Lia, D. Bensimon, V. Croquette, J.-F. Allemand, D. Dunlap, D. E. A. Lewis, S. Adhya, and L. Finzi, *Supercoiling and Denaturation in Gal Repressor/Heat Unstable Nucleoid Protein (HU)-Mediated DNA Looping*, Proceedings of the National Academy of Sciences **10** (2003), 11373.
- [LC86] S. D. Levene and D. M. Crothers, *Ring Closure Probabilities for DNA Fragments by Monte Carlo Simulation*, Journal of Molecular Biology **189** (1986), 61.
- [LCH⁺96] M. Lewis, G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan, and P. Lu, *Crystal Structure of the Lactose Operon Repressor and Its Complexes With DNA and Inducer*, Science **271** (1996), 1247.
- [LE95] T. B. Liverpool and S. F. Edwards, *Probability Distribution of Wormlike Polymer Loops*, Journal of Chemical Physics **103** (1995), 6716.
- [LGA99] D. E. A. Lewis, M. Geanacopoulos, and S. Adhya, *Role of HU and DNA Supercoiling in Transcription Repression : Specialized Nucleoprotein Repression Complex at Gal Promoter in Escherichia Coli*, Molecular Microbiology **31** (1999), 451.
- [LL] L. Landau and E. Lifchitz, *Physique théorique tome 1 : mécanique*, Librairie du globe, MIR.
- [LL90] L Landau and E. Lifchitz, *Physique théorique tome 7 : théorie de l'élasticité*, 2ème ed., Librairie du globe, MIR, 1990.
- [LNN91] J. B. Lagowski, J. Noolandi, and B. Nickel, *Stiff Chain Model – Functional Integral Approach*, Journal of Chemical Physics **95** (1991), 1266.
- [LRB⁺98] J.-F. Léger, J. Robert, L. Bourdieu, D. Chatenay, and J. F. Marko, *RecA Binding to a Single Double-Stranded DNA Molecules*, Proceedings of the National Academy of Sciences **95** (1998), 12295.
- [LRS⁺99] J.-F. Léger, G. Romano, A. Sarkar, J. Robert, L. Bourdieu, D. Chatenay, and J. F. Marko, *Structural Transitions of a Twisted and Stretched DNA Molecule*, Physical Review Letters **83** (1999), 1066.
- [LSAA97] Y. L. Lyubchenko, L. S. Shlyakhtenko, T. Aki, and S. Adhya, *Atomic Force Microscopic Demonstration of DNA Looping by GalR and HU*, Nucleic Acids Research **25** (1997), 873.
- [LVTH04] T. Lu, D. Volfson, L. Tsimring, and J. Hasty, *Cellular Growth and Division in the Gillespie Algorithm*, Systems Biology **1** (2004), 121.

- [LW95] H. Li and T. A. Witten, *Fluctuations and Persistence Length of Charged Flexible Polymers*, *Macromolecules* **28** (1995), 5921.
- [LWS02] Y. Lu, B. Weers, and N. C. Stellwagen, *DNA persistence length revisited*, *Biopolymers* **61** (2002), 261.
- [Man69] G. S. Manning, *Limiting Laws and Counterion Condensation in Polyelectrolyte Solutions. (i) Colligative Properties*, *Journal of Chemical Physics* **51** (1969), 924.
- [Mar98] J. F. Marko, *DNA Under High Tension : Overtwisting, Undertwisting, and Relaxation Dynamics*, *Physical Review E* **57** (1998), 2134.
- [Mat92] K. S. Matthews, *DNA Looping*, *Microbiological Reviews* **56** (1992), 123.
- [MC03a] M. Marucho and G. A. Carri, *On the Generalized Borel Transform and Its Application to the Statistical Mechanics of Macromolecules*, *Journal of Mathematical Physics* **44** (2003), 6020.
- [MC03b] ———, *Statistics of Semiflexible Polymer Chains and the Generalized Borel Transform*, cond-mat/0306748 (2003).
- [Mer97] E. Merzbacher, *Quantum Mechanics*, 3rd ed., Wiley, 1997.
- [Mes95] A. Messiah, *Mécanique quantique tomes 1 et 2*, 2ème ed., Dunod, 1995.
- [MH98] B. Müller-Hill, *Some Repressors of Bacterial Transcription*, *Current Opinion in Microbiology* **1** (1998), 145.
- [MK96] U. Micka and K. Kremer, *The Persistence Length of Polyelectrolyte Chains*, *Journal of Physics : Condensed Matter* **8** (1996), 9463.
- [MN97] J. D. Moroz and P. Nelson, *Torsional Directed Walks, Entropic Elasticity, and DNA Twist Stiffness*, *Proceedings of the National Academy of Sciences* **94** (1997), 14418.
- [MN98] ———, *Entropic Elasticity of Twist-Storing polymers*, *Macromolecules* **31** (1998), 6333.
- [MN04] M. Manghi and R. R. Netz, *Variational Theory For a Single Polyelectrolyte Chain Revisited*, *The European Physical Journal E* **14** (2004), 67.
- [MRKL98] H. Merlitz, K. Rippe, K. V. Klenin, and J. Langowski, *Looping Dynamics of Linear DNA Molecules and the Effect of DNA Curvature : A Study by Brownian Dynamics Simulation*, *Biophysical Journal* **74** (1998), 773.
- [MS95] J. F. Marko and E. D. Siggia, *Stretching DNA*, *Macromolecules* **28** (1995), 8759.

- [Mur00] N. E. Murray, *Type I Restriction Systems : Sophisticated Molecular Machines (A Legacy of Bertani and Weigle)*, *Microbiology and Molecular Biology Reviews* **64** (2000), 412.
- [MvL03] C. Moler and C. van Loan, *Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later*, *Society for Industrial and Applied Mathematics Review* **45** (2003), 3.
- [Nag00] A. Nagy, *Cre Recombinase : The Universal Reagent for Genome Tailoring*, *Genesis* **26** (2000), 99.
- [Nel98] P. Nelson, *Sequence-Disorder Effects on DNA Entropic Elasticity*, *Physical Review Letters* **80** (1998), 5810.
- [NMF78] N. Norisuye, H. Murakama, and H. Fujita, *Wormlike Chains Near the Rod Limit : Moments and Particle Scattering Function*, *Macromolecules* **11** (1978), 966.
- [Odi95] T. Odijk, *Stiff Chains and Filaments Under Tension*, *Macromolecules* **28** (1995), 7016.
- [Pau05] J. Paulsson, *Models of Stochastic Gene Expression*, *Physics of Life Reviews* **2** (2005), 157.
- [PE01] J. Paulsson and M. Ehrenberg, *Noise in a Minimal Regulatory Network : Plasmid Copy Number Control*, *Quarterly Reviews of Biophysics* **34** (2001), 1.
- [PK04] J. Puchalka and A. M. Kierzek, *Bridging the Gap Between Stochastic and Deterministic Regimes in the Kinetic of the Biochemical Reaction Networks*, *Biophysical Journal* **86** (2004), 1357.
- [PNE98] J. Paulsson, K. Nordström, and M. Ehrenberg, *Requirements for Rapid Plasmid ColE1 Copy Number Adjustments : A Mathematical Model of Inhibition Modes and RNA Turnover Rates*, *Plasmid* **39** (1998), 215.
- [PT05a] Y. O. Popov and A. V. Tkachenko, *Effects of Kinks on DNA Elasticity*, *Physical Review E* **71** (2005), 051905–1.
- [PT05b] ———, *Effects of Sequence Disorder on DNA Looping and Cyclization*, *cond-mat/0510302* (2005).
- [PTVF02] W. H. Press, S. A. Teulkolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C. The Art of Scientific Computing*, 2nd ed., Cambridge University Press, 2002, Consultable sur <http://www.library.cornell.edu/nr/>.
- [PV00] A. A. Podtelezhnikov and A. V. Vologodskii, *Dynamics of Small Loops in DNA Molecules*, *Macromolecules* **33** (2000), 2767.

- [RB99] J. Rudnick and R. Bruinsma, *DNA-Protein Cooperative Binding Through Variable-Range Elastic Coupling*, *Biophysical Journal* **76** (1999), 1725.
- [RCA⁺99] L. Ringrose, S. Chabanis, P.-O. Angrand, C. Woodroffe, and A. F. Stewart, *Quantitative Comparison of DNA Looping in Vitro and in Vivo : Chromatin Increases Effective DNA Flexibility at Short Distances*, *The EMBO Journal* **18** (1999), 6630.
- [RGvHB97] K. Rippe, M. Guthold, P. H. von Hippel, and C. Bustamante, *Transcriptional Activation Via DNA-Looping : Visualization of Intermediates in the Activation Pathway of E Coli*, *Journal of Molecular Biology* **270** (1997), 125.
- [RHMM03] A. Rosa, T. X. Hoang, D. Marenduzzo, and A. Maritan, *Elasticity of Semiflexible Polymers With and Without Self-Interactions*, *Macromolecules* **36** (2003), 10095.
- [Rip01] K. Rippe, *Making Contacts on a Nucleic Acid Polymer*, *Trends in Biochemical Sciences* **26** (2001), 733.
- [RM02] V. Rossetto and A. C. Maggs, *Comment on : "elasticity Model of a Supercoiled DNA Molecule"*, *Physical Review Letters* **88** (2002), 089801–1.
- [RSKM05] P. Ranjith, P. B. Sunil Kumar, and G. I. Menon, *Distribution Functions, Loop Formation Probabilities, and Force-Extension Relations in a Model for Short Double-Stranded DNA Molecules*, *Physical Review Letters* **94** (2005), 138102–1.
- [RvHL95] K. Rippe, P. H. von Hippel, and J. Langowski, *Action at Distance : DNA-Looping and Initiation of Transcription*, *Trends in Biochemical Sciences* **20** (1995), 500.
- [RWB98] C. Rivetti, C. Walker, and C. Bustamante, *Polymer Chain Statistics and Conformational Analysis of DNA Molecules With Bends Or Sections of Different Flexibility*, *Journal of Molecular Biology* **280** (1998), 41.
- [RYMN96] P. A. Rice, S.-W. Yang, K. Mizuuchi, and H. A. Nash, *Crystal Structure of an IHF-DNA Complex : A Protein-Induced DNA U-Turn*, *Cell* **87** (1996), 1295.
- [Sak94] J. J. Sakurai, *Modern Quantum Mechanics*, 2nd ed., Addison Wesley, 1994.
- [SB83] D. Shore and R. L. Baldwin, *Energetics of DNA Twisting. (i) Relation Between Twist and Cyclization Probability*, *Journal of Molecular Biology* **170** (1983), 957.

- [SC02] L. Snyder and W. Champness, *Molecular genetics of bacteria*, 2nd ed., American Society for Microbiology, 2002.
- [Sch68] L. I. Schiff, *Quantum Mechanics*, 3rd ed., McGraw-Hill, 1968.
- [SDC⁺03] T. 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 **66** (2003), 1.
- [SES02] P. S. Swain, M. B. Elowitz, and E. D. Siggia, *Intrinsic and Extrinsic Contributions to Stochasticity in Gene Expression*, Proceedings of the National Academy of Sciences **99** (2002), 12795.
- [SFB92] S. B. Smith, L. Finzi, and C. Bustamante, *Direct Mechanical Measurements of the Elasticity of Single DNA Molecules by Using Magnetic Beads*, Science **258** (1992), 1122.
- [SGLA02] S. Semsey, M. Geanacopoulos, D. E. A. Lewis, and S. Adhya, *Operator-Bound GalR Dimers Close DNA Loops by Direct Interaction : Tetramerization and Inducer Binding*, The EMBO Journal **21** (2002), 4349.
- [Sid98] R. B. Sidje, *Expokit : Software Package for Computing Matrix Exponentials*, Association for Computing Machinery : Transaction on Mathematical Software **24** (1998), 130, Sources disponibles gratuitement sur <http://www.maths.uq.edu.au/expokit/>.
- [Sir02] C. Sire, *Physique hors d'équilibre*, Cours du DEA « physique de la matière », université Paul Sabatier (Toulouse III), 2002.
- [SKOS01] M. Sano, A. Kamino, J. Okamura, and S. Shinkai, *Ring Closure of Carbon Nanotubes*, Science **293** (2001), 1299.
- [SLB81] D. Shore, J. Langowski, and R. L. Baldwin, *DNA Flexibility Studied by Covalent Closure of Short Fragments Into Circles*, Proceedings of the National Academy of Sciences **78** (1981), 4833.
- [SLR00] A. Schulz, J. Langowski, and K. Rippe, *The Effect of the DNA Conformation on the Rate of NtrC Activated Transcription of Escherichia Coli RNA Polymerase- σ^{54} Holoenzyme*, Journal of Molecular Biology **300** (2000), 709.
- [SLZR03] K. K. Swinger, K. M. Lemberg, Y. Zhang, and P. A. Rice, *Flexible DNA Bending in HU-DNA Crocrystal Structures*, The EMBO Journal **22** (2003), 3749.
- [SM05a] S. Sankararaman and J. F. Marko, *Entropic Compression of Interacting DNA Loops*, Physical Review Letters **95** (2005), 078104–1.

- [SM05b] ———, *Formation of Loops in DNA Under Tension*, Physical Review E (2005), 021911–1.
- [SN03] C. Storm and P. C. Nelson, *Theory of High-Force DNA Stretching and Overstretching*, Physical Review E **67** (2003), 051906–1.
- [Spa04] A. J. Spakowitz, *Semiflexible Polymers : Fundamental Theory and Applications in DNA Packaging*, Ph.D. thesis, California Institute of Technology, 2004.
- [SPKE90] W. Su, S. Porter, S. Kustu, and H. Echols, *DNA-Looping and Enhancer Activity : Association Between DNA-Bound NtrC Activator and RNA Polymerase at the Bacterial GlnA Promoter*, Proceedings of the National Academy of Sciences **87** (1990), 5504.
- [SR01] J. Sambrook and D. W. Russel, *Molecular Cloning, a Laboratory Manual Volume 1*, 3rd ed., Cold Spring Harbor, 2001.
- [SS02a] J. Samuel and Sinha S., *Elasticity of Semiflexible Polymers*, Physical Review E **66** (2002), 050801–1.
- [SS02b] S. Stepanow and G. M. Schütz, *The Distribution Function of a Semiflexible Polymer and Random Walks With Constraints*, Europhysics Letters **60** (2002), 546.
- [SSG05] J. Samuel, Sinha S., and A. Ghosh, *DNA Elasticity : Topology of Self-Avoidance*, cond-mat/0504189 (2005).
- [ST83] E. Seneta and S. Tavaré, *Some Stochastic Models for Plasmid Copy Number*, Theoretical Population Biology **23** (1983), 241.
- [Ste03] S. Stepanow, *Statistical Mechanics of Semiflexible Polymers*, cond-mat/0306149 (2003).
- [Ste05] ———, *On the Behaviour of the Short Kratky-Porod Chain*, Journal of Physics : Condensed Matter **17** (2005), S1799.
- [STRJ02] I. A. Shkel, O. V. Tsodikov, and T. M. Record Jr, *Asymptotic Solution of the Cylindrical Nonlinear Poisson-Boltzmann Equation at Low Salt Concentration : Analytical Expressions for Surface Potential and Preferential Interaction Coefficient*, Proceedings of the National Academy of Sciences **99** (2002), 2597.
- [STY67] N. Saitô, K. Takahashi, and Y. Yunoki, *The Statistical Mechanical Theory of Stiff Chains*, Journal of the Physical Society of Japan **22** (1967), 219.
- [SVA05] S. Semsey, K. Virnik, and S. Adhya, *A Gamut of Loops : Meandering DNA*, Trends in Biochemical Sciences **30** (2005), 334.
- [SW04] A. J. Spakowitz and Z.-G. Wang, *Exact Results for a Semiflexible Polymer Chain in an Aligning Field*, Macromolecules **37** (2004), 5814.

- [SW05] ———, *End-to-End Distance Vector Distribution With Fixed End Orientations for the Wormlike Chain Model*, Physical Review E **72** (2005), 041802–1.
- [Swa04] P. S. Swain, *Efficient Attenuation of Stochasticity in Gene Expression Through Post-Transcriptional Control*, Journal of Molecular Biology **344** (2004), 965.
- [SY84] J. Shimada and H. Yamakawa, *Ring-Closure Probabilities for Twisted Wormlike Chains. Application to DNA*, Macromolecules **17** (1984), 689.
- [TC02] M. Trabi and D. J. Craik, *Circular Proteins – No End in Sight*, Trends in Biochemical Sciences **27** (2002), 132.
- [TH97] D. Thirumalai and B.-Y. Ha, *Statistical Mechanics of Semiflexible Chains : A Meanfield Variational Approach*, cond-mat/9705200 (1997).
- [Tsi98] R. Y. Tsien, *The Green Fluorescent Protein*, Annual Reviews of Biochemistry **67** (1998), 509.
- [TW86] H. Tsuru and M. Wadati, *Elastic Model of Highly Supercoiled DNA*, Biopolymers **25** (1986), 2083.
- [UJPS97] M. Ullner, B. Jönsson, C. Peterson, and O. Sommelius, *The Electrostatic Persistence Length Calculated From Monte Carlo, Variational and Perturbation Methods*, Journal of Chemical Physics **107** (1997), 1279.
- [VBMS04] E. Villa, A. Balaeff, L. Mahadevan, and K. Schulten, *Multiscale Method for Simulating Protein-DNA Complexes*, Society for Industrial and Applied Mathematics : Multiscale Modeling and Simulation **2** (2004), 527.
- [VBS05] E. Villa, A. Balaeff, and K. Schulten, *Structural Dynamics of the Lac Repressor-DNA Complex Revealed by a Multiscale Simulation*, Proceedings of the National Academy of Sciences **102** (2005), 6783.
- [vdZ03] G. van der Zwan, *Kramer’s Theory*, Consultable sur la page personnelle <http://www.chem.vu.nl/acas/Homepages/zwan/>, 2003.
- [VLK⁺03] K. Virnick, Y. L. Lyubchenko, M. A. Karymov, P. Dahlgren, M. Y. Tolstorukov, S. Semsey, V. B. Zhurkin, and S. Adhya, “*Antiparallel*” *DNA Loop in Gal Repressosome Visualized by Atomic Force Microscopy*, Journal of Molecular Biology **334** (2003), 53.
- [vNvdHdJ⁺03] J. van Noort, T. van der Heijden, M. de Jager, C. Wyman, R. Kanaar, and C. Dekker, *The Coiled-Coil of the Human Rad50*

- DNA Repair Protein Contains Specific Segments of Increased Flexibility*, Proceedings of the National Academy of Sciences **100** (2003), 7581.
- [vNVG⁺04] J. van Noort, S. Verbrugge, N. Goosen, C. Dekker, and R. T. Dame, *Dual Architectural Roles of HU : Formation of Flexible Hinges and Rigid Filaments*, Proceedings of the National Academy of Sciences **101** (2004), 6969.
- [WBB⁺04] J. Watson, T. Baker, S. Bell, A. Gann, M. Levine, and Losick R., *Molecular Biology of the Gene*, 5th ed., Benjamin Cummings, 2004, Chapitres 6 et 16 consultables sur <http://www.pearson.ch/HigherEducation/BenjaminCummings/>.
- [Weg05] G. Wegrzin, *What Does "Plasmid Biology" Currently Mean? Summary of the Plasmid Biology 2004 Meeting*, Plasmid **53** (2005), 14.
- [WF96] J. Wilhelm and E. Frey, *Radial Distribution Function of Semiflexible Polymers*, Physical Review Letters **77** (1996), 2581.
- [Win03] R. G. Winkler, *Deformation of Semiflexible Chains*, Journal of Chemical Physics **118** (2003), 2919.
- [WN05] P. A. Wiggins and P. C. Nelson, *A Generalized Theory of Semiflexible Polymers*, cond-mat/0508155 (2005).
- [WPN05] P. A. Wiggins, R. Philips, and P. C. Nelson, *Exact Theory of Kinkable Elastic Polymers*, Physical Review E **71** (2005), 021909–1.
- [WR92] R. G. Winkler and P. Reineker, *Finite Size Distribution and Partition Functions of Gaussian Chains : Maximum Entropy Approach*, Macromolecules **25** (1992), 6891.
- [WRH94] R. G. Winkler, P. Reineker, and L. Harnau, *Models and Equilibrium Properties of Stiff Molecular Chains*, Journal of Chemical Physics **101** (1994), 8119.
- [WRN⁺97] C. Wyman, I. Rombel, A. K. North, C. Bustamante, and S. Kustu, *Unusual Oligomerization Required for Activity of NtrC, a Bacterial Enhancer-Binding Protein*, Science **275** (1997), 1658.
- [Yam73] H. Yamakawa, *Statistical Mechanics of Wormlike Chains : Path Integral and Diagram Methods*, Journal of Chemical Physics **59** (1973), 3811.
- [Yam97] ———, *Helical Wormlike Chains in Polymer Solutions*, 1st ed., Springer-Verlag, 1997.
- [YF74] H. Yamakawa and M. Fujii, *Light Scattering From Wormlike Chains. Determination of the Shift Factor*, Macromolecules **7** (1974), 649.

- [YKM05] J. Yan, R. Kawamura, and J. F. Marko, *Statistics of Loop Formation Along Double Helix DNAs*, Physical Review E **71** (2005), 061905–1.
- [YM03] J. Yan and J. F. Marko, *Effects of DNA-Distorting Proteins on DNA Elastic Response*, Physical Review E **68** (2003), 011905–1.
- [YM04] ———, *Localized Single-Stranded Bubble Mechanism for Cyclization of Short Double Helix DNA*, Physical Review Letters **93** (2004), 108108–1.
- [YS72] H. Yamakawa and W. S. Stockmayer, *Statistical Mechanics of Wormlike Chains. (ii) Excluded Volume Effects*, Journal of Chemical Physics **57** (1972), 2843.
- [ZC03] Y. Zhang and D. M. Crothers, *Statistical Mechanics of Sequence-Dependent Circular DNA and Its Application for DNA Cyclization*, Biophysical Journal **84** (2003), 136.
- [Zho01] H.-X. Zhou, *Loops in Proteins Can Be Modeled as Worm-Like Chains*, Journal of Physical Chemistry B **105** (2001), 6763.
- [ZRG02] R. Zandi, J. Rudnick, and R. Golestanian, *Radial Distribution Function of Rod-Like Polyelectrolytes*, EPJE **9** (2002), 41.
- [ZRG03a] ———, *Anomalous Bending of a Polyelectrolyte*, Physical Review E **67** (2003), 061805.
- [ZRG03b] ———, *Probing Polyelectrolyte Elasticity Using Radial Distribution Function*, Physical Review E **67** (2003), 021803–1.
- [ZSZ97] S. R. Zhao, C. P. Sun, and W. X. Zhang, *Statistics of Wormlike Chains. (I) Properties of a Single Chain*, Journal of Chemical Physics **106** (1997), 2520.

Résumé

Le manuscrit concerne la *modélisation* de mécanismes de *régulation* de l'expression génétique. Il se divise en *deux* parties.

Une étude *statique* de la formation de *boucles* dans la molécule d'ADN par exemple observable lors de la répression, à l'étape de transcription, de l'opéron Lac. Sont pris en compte : les effets liés à la *taille* des protéines fixant la boucle, des mécanismes entraînant une *perte* de rigidité du double brin – typiquement l'accrochage d'autres protéines – ainsi que son *étirement* au moyen d'une force extérieure. Nous avons calculé *numériquement* les diverses *distributions* en extension, ainsi que le facteur de *cyclisation* du modèle de polymère *semi-flexible*, dit « du ver ». Seule la rigidité de *courbure* de l'ADN a été considérée. Nous avons adapté l'expression *analytique* obtenue par Shimada et Yamakawa en 1984.

Suit un *bilan* posant les *bases* d'une simulation *stochastique* du mécanisme permettant, à l'étape de réplication, le Contrôle du Nombre de *Copies* (CNC) du plasmide ColE1. Ce travail *numérique* devrait compléter une future caractérisation tant théorique qu'expérimentale, de la *robustesse* au bruit.

Mots-clefs : élasticité de l'ADN, polymères semi-flexibles, formation de boucles, facteur de cyclisation, plasmide ColE1, contrôle du nombre de copies.

Abstract

Modelization of mechanisms *regulating* gene expression is the main topic of this manuscript which is divided in *two* parts.

Looping of the DNA molecule is studied from the *static* point of view. *Size* of proteins clamping the loop is taken into account as well as mechanisms inducing stiffness *loss* – additional proteins binding for instance – or *stretching* of the molecule thanks to an external force. End-to-end distance *distributions* including the *cyclization* factor of the « Worm-Like Chain » (WLC) have been computed using *numerical* methods. This *semiflexible* polymer model includes only *bending* stiffness all along this work. The *analytical* expression Shimada and Yamakawa derived in 1984 is furthermore extended.

Follows a *brief* report on *numerical stochastic* algorithms which allow the dynamical simulation of the plasmid ColE1 *Copy* Number Control (CNC) mechanism. These should shortly complement further theoretical and experimental studies on *robustness* with respect to noise.

Keywords : DNA elasticity, semiflexible polymers, loops formation, cyclization factor, ColE1 plasmid, copy number control.