



HAL
open science

Etude de la dynamique et de la physique statistique de modèles d'ADN non-linéaires à la dénaturation thermique

Sahin Buyukdagli

► **To cite this version:**

Sahin Buyukdagli. Etude de la dynamique et de la physique statistique de modèles d'ADN non-linéaires à la dénaturation thermique. Biophysique [physics.bio-ph]. Université Joseph-Fourier - Grenoble I, 2007. Français. NNT: . tel-00185960

HAL Id: tel-00185960

<https://theses.hal.science/tel-00185960>

Submitted on 7 Nov 2007

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

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

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ JOSEPH FOURIER - GRENOBLE I

Discipline : PHYSIQUE THEORIQUE

par

Sahin BUYUKDAGLI

Etude de la dynamique et de la physique statistique de modèles d'ADN non-linéaires à la dénaturation thermique

Soutenue le 3 octobre 2007, devant le jury composé de Messieurs :

R. Blossey	Rapporteur
J. Cognet	Examineur
E. Coissac	Examineur
T. Dombre	Examineur
G. Gaeta	Rapporteur
M. Joyeux	Directeur de thèse

Thèse réalisée au Laboratoire de Spectrométrie Physique, B.P. 87 - 38402 Saint Martin
d'Hères Cedex, France

TABLE DES MATIERES

Introduction	5
1. Une brève introduction à la structure et aux fonctions de l'ADN	12
2. Modèles dynamiques pour la dénaturation de l'ADN	21
2.1 Modèle de Poland Scheraga	21
2.2 Modèles de Prohofsky et de Dauxois, Peyrard et Bishop	24
2.3 Nouveau modèle basé sur des énergies d'empilement finies	27
2.4 Vers des modèles plus réalistes	40
3. Fluctuations de température de l'ADN au voisinage de T_c	51
3.1 Amplification des fluctuations de température à la dénaturation thermique	53
3.2 Effets du coefficient de dissipation γ sur les fluctuations de température	60
3.3 Fluctuations de température et bulles	61
3.4 Excitations collectives, amplification des fluctuations et comportement en $1/f$	67
3.5 Conclusions	73
4. Transition de phase de dénaturation des séquences homogènes infinies	75
4.1 Quelques notions de transitions de phase	75
4.1.1 Définitions de base	75
4.1.2 Exposants critiques	80
4.1.3 Lois d'échelle	83
4.1.4 Intégrale de Transfert pour des séquences homogènes de longueur infinie	85
4.2 Propriétés critiques de trois modèles d'ADN	88
4.2.1 Modèle de Peyrard-Bishop harmonique	89
4.2.2 Modèles anharmoniques DPB et JB	93
4.3 Conclusions	103
5. Influence de la longueur de la séquence sur la transition de phase	105
5.1 Méthode TI pour les séquences finies aux extrémités libres	107
5.2 Effets de la finitude de taille au voisinage de T_c	113
5.3 Renormalisation de l'énergie libre	121
5.4 Conclusion	127

6. Transition de dénaturation des séquences d'ADN hétérogènes	128
6.1 Méthode TI pour les séquences d'ADN hétérogènes	132
6.2 Etude de la transition de dénaturation	135
Conclusion	145
Bibliographie	148

INTRODUCTION

Une des particularités remarquables de la physique théorique est certainement sa capacité à prédire le comportement d'un système à des échelles très différentes, à partir d'une description rigoureuse de sa structure et de ses interactions avec son environnement. Les outils de la physique théorique ont ainsi permis de comprendre des phénomènes aussi divers que le confinement des quarks, les fluctuations temporelles de la bourse ou la dynamique primordiale de l'univers. La quantité de physiciens théoriciens recrutés dans les milieux en dehors de la recherche, comme la finance ou l'industrie, souligne d'ailleurs l'efficacité de ces outils.

Dans la compréhension du monde vivant, la physique apporte souvent un point de vue complémentaire très instructif à celui, primordial, de la biologie. La description de la structure statique de l'ADN proposée par Watson et Crick en 1953 [1] a par exemple ouvert la voie à l'étude théorique de la dynamique de cette molécule et, depuis 40 ans, l'ADN ne cesse d'être le sujet d'études de physique théorique et expérimentale très poussées.

Les expériences de spectroscopie d'absorption UV à 260 nm ont ainsi apporté des renseignements très importants sur la dénaturation thermique de l'ADN [2]-[4]. A travers l'enregistrement du coefficient d'absorption d'une molécule d'ADN chauffée dans un solvant, ces expériences montrent qu'à des températures particulières des portions finies de la séquence s'ouvrent brusquement, jusqu'à ce que l'on atteigne la température où le double brin se dissocie complètement. Le tracé de la fraction de paires de bases ouvertes en fonction de la température (voir par exemple la fig. 1.5 du chapitre 1) montre que chaque ouverture locale semble correspondre à une transition de phase discontinue. Ces points sont discutés un peu plus en détail dans le chapitre 1 de cette thèse, où je décris

également les processus de transcription et de réplication qui sont d'une importance vitale, mais sont aussi beaucoup plus difficiles à modéliser que la dénaturation thermique, dans la mesure où ils font intervenir à la fois la structure intrinsèque de l'ADN et des éléments extérieurs (enzymes, protéines, etc...). Pour comprendre un peu finement cette structure dynamique intrinsèque, il est donc sage et prudent de commencer par s'intéresser à la dénaturation thermique. Les premiers efforts dans cette direction privilégiaient une description du système en termes d'états discrets, comme le modèle d'Ising ou de Poland-Scheraga [5]. Cependant, une approche Hamiltonienne peut s'avérer souhaitable pour plusieurs raisons. Tout d'abord, un modèle continu est plus acceptable, dans la mesure où dans la nature les états configurationnels décrivant la séparation de deux bases sont continus, de sorte que la discrétisation de ces états ne constitue qu'une approximation, dont la vérification rigoureuse des limites est souvent négligée. De plus, la non-linéarité, une propriété incontournable des systèmes biologiques, est intégrée de façon plus naturelle dans une approche Hamiltonienne. Plus précisément, l'ADN est un système dont la dynamique est caractérisée par des vibrations de faible amplitude à des températures de l'ordre de 300 K et des fluctuations de grande amplitude lorsque l'on s'approche de la dénaturation au-dessus de 350 K. De ce fait, seul un Hamiltonien non-linéaire est capable de décrire la dynamique de l'ADN sur une telle plage de températures. Enfin, la construction d'un modèle Hamiltonien pour un système particulier permet d'étudier aussi bien sa dynamique temporelle que ses propriétés thermodynamiques. Cela est d'une importance majeure, puisque ces deux approches sont souvent complémentaires. Par exemple, la dénaturation thermique de l'ADN est une transition de phase d'équilibre que je vais étudier dans cette thèse en grande partie par les outils de la physique statistique d'équilibre. Mais lorsque l'on augmente la température d'une molécule d'ADN placée dans un solvant, la molécule est conduite vers cette dénaturation par l'amplification des bulles qui sont des états métastables, c'est à dire des structures dynamiques de longue durée de vie. Dans la suite, nous allons voir que d'autres phénomènes de nature dynamique résultent de la dynamique non-linéaire de l'ADN. En

d'autres termes, une compréhension complète du phénomène de dénaturation nécessite une analyse à la fois dynamique et thermodynamique, pour laquelle une formulation Hamiltonienne est incontestablement la plus appropriée.

Le premier modèle Hamiltonien de l'ADN basé sur des variables continues décrivant correctement la dénaturation thermique, c'est à dire les distances entre les bases appariées, a été proposé par Gao et Prohofsky en 1984 [7]. Ce modèle a été utilisé à partir de 1989 par Peyrard et Bishop [8] pour étudier la dénaturation thermique de façon plus poussée. Le modèle de Prohofsky, qui suppose une interaction harmonique entre les monomères adjacents situés sur le même brin (interaction d'empilement), conduit cependant à une dénaturation trop progressive comparée aux courbes de dénaturation expérimentales. Cela a amené Dauxois, Peyrard et Bishop à proposer en 1993 un modèle amélioré [9]. Ces auteurs ont montré que ce modèle d'ADN homogène (modèle DPB), qui intègre une interaction d'empilement anharmonique, reproduit une transition du premier ordre en très bon accord avec l'expérience. Depuis, le modèle DPB a été l'objet d'études théoriques concernant les solitons [10, 11], le transport de charge le long de l'ADN [12, 13], la dynamique des bulles [14] etc... Moyennant certaines modifications mineures, il a également été utilisé pour étudier la dénaturation des séquences hétérogènes dans le cas unidimensionnel [15] et tridimensionnel [16].

De mon point de vue, le modèle DPB présente cependant deux inconvénients. Tout d'abord, l'interaction d'empilement de ce modèle diverge lorsque les paires de bases adjacentes s'éloignent, alors que du point de vue physique, on s'attend au contraire à ce que cette interaction finisse par s'annuler. De plus, la version hétérogène de ce modèle est basée uniquement sur la variation des énergies de liaison hydrogène en fonction de la paire de bases [15] et néglige la variation des énergies d'empilement, alors que les modèles thermodynamiques semblent suggérer que ces dernières jouent un rôle fondamental [17]. Dans le chapitre 6 de cette thèse, je montrerai que cela rend discutables les résultats obtenus avec ce modèle à l'échelle de quelques paires de bases. Ces deux points nous ont poussé à proposer un nouveau modèle non-linéaire d'ADN (modèle JB). Ce modèle diffère

du modèle DPB par la forme des interactions d'empilement, qui sont finies, et par le fait qu'il s'adapte de façon plus naturelle à l'étude de la dynamique des séquences d'ADN de structure génétique arbitraire. Nous avons montré que ces modifications permettent de modéliser la dénaturation en plusieurs étapes des séquences d'ADN hétérogènes. Ce travail est présenté dans le chapitre 2.

Bien que le modèle JB, qui décrit la dynamique de l'ADN par le seul degré de liberté d'élongation de la distance entre les bases appariées, reproduise de façon très satisfaisante les aspects essentiels des courbes de dénaturation expérimentales, il considère que la molécule d'ADN n'est rien d'autre qu'une double chaîne de polymères unidimensionnelle, alors que c'est une hélice formée par la torsion de cette double chaîne. Par conséquent, la rotation des bases par rapport à l'axe de cette hélice est un degré de liberté significatif pour la dissociation du double brin. Nous avons donc amélioré le modèle JB en tenant compte des degrés de liberté de torsion. Ce modèle amélioré (modèle BSJ) est décrit à la fin du chapitre 2.

Comme je l'ai déjà noté, l'ADN a une dynamique très riche et l'accessibilité expérimentale de ce système permet des comparaisons poussées entre les études expérimentales et les analyses théoriques. Sur le plan expérimental, Nagapriya et al. [21] ont récemment observé une amplification des fluctuations de température des brins d'ADN près de la température de dénaturation. Nous avons étudié ce phénomène dans un contexte théorique à partir des modèles DPB, JB et BSJ pour des séquences homogènes et hétérogènes. Nous avons confirmé l'amplification des fluctuations de température rapportée par Nagapriya et ses collaborateurs et observé de plus un spectre de puissance en $1/f$ de ces fluctuations. De telles fluctuations en $1/f$ sont présentes dans un grand nombre de systèmes électroniques, biologiques, financiers etc... L'explication de ce phénomène proposée par Geisel et al. [22] est basée sur une structure fractale de l'espace des phases. Notre étude a montré que les fluctuations en $1/f$ de l'ADN résultent plutôt des excitations collectives des paires de bases, qui sont amplifiées près de la température critique par la divergence de la longueur de corrélation. Le chapitre 3 est consacré à la présentation de

ces résultats.

J'ai précisé que les observations expérimentales suggèrent que la transition de dénaturation de l'ADN est discontinue. Un modèle théorique censé modéliser la dissociation de l'ADN doit donc conduire à une transition du premier ordre. La façon la plus rigoureuse de vérifier l'ordre d'une transition de phase consiste à étudier le comportement de l'entropie et de la chaleur spécifique du système au voisinage de la température critique : par définition, une transition de phase discontinue implique une absorption de chaleur latente par le système, donc un saut discontinu de son entropie. Cela peut être vérifié de façon quantitative par le calcul de l'exposant critique de la chaleur spécifique, qui se comporte comme $c_V \propto t^{-\alpha}$ au voisinage de T_c . Une transition discontinue nécessite un exposant critique α plus grand que un. A partir de ce critère, j'ai montré, en utilisant la méthode de l'intégrale de transfert, que les modèles DPB et JB homogènes subissent bien une transition du premier ordre. J'ai ensuite testé les quatre lois d'échelle qui relient entre eux les six exposants critiques fondamentaux et montré que deux d'entre elles ne sont pas satisfaites pour l'ADN. En fait, l'ADN est un système critique qui, comparé à la majorité des systèmes subissant une transition de phase du premier ordre, possède des aspects inhabituels, à savoir un paramètre d'ordre (l'élongation des paires de bases) qui diverge au point critique et une dimension égale à un. La violation des deux lois résulte probablement de ces caractéristiques particulières. Ces analyses sont détaillées et interprétées en détail dans le chapitre 4.

Il est bien connu qu'une transition de phase ne se manifeste qu'à la limite thermodynamique, alors que les systèmes réels possèdent forcément un volume fini. C'est également le cas des brins d'ADN étudiés dans les expériences d'absorption UV. En fait, une transition de phase est en général lissée par la finitude de la taille du système et ce lissage se manifeste par un élargissement de la région de transition. Ce phénomène a conduit au développement de la théorie de *finite size scaling* qui vise à suivre l'évolution de la singularité critique en fonction de la taille du système. Dans le cadre de ce formalisme, j'ai étudié le comportement des observables thermodynamiques et de la température de

dissociation de l'ADN en fonction de la longueur de la séquence et vérifié la validité de certaines hypothèses de cette théorie pour la molécule d'ADN. Ce travail est discuté dans le chapitre 5.

L'étude du comportement critique de l'ADN homogène est très enrichissante dans la mesure où il implique un seul point critique dont on peut étudier les caractéristiques propres sans être gêné par les complications habituellement associées à la présence de défauts et d'hétérogénéité dans le système. Cependant, les molécules d'ADN qui jouent un rôle dans les processus génétiques sont caractérisées par une structure primaire hétérogène. Il est donc important de comprendre les effets de cette hétérogénéité sur la transition de dénaturation de l'ADN. Il est surtout crucial de vérifier si l'aspect discontinu de la transition de phase d'une séquence homogène, que j'ai quantitativement démontré dans le chapitre 4 pour deux modèles théoriques d'ADN, est maintenu dans le cas d'une séquence hétérogène. En fait, les courbes de dénaturation expérimentales, comme celle de la fig. 1.5, qui sont enregistrées pour des molécules d'ADN hétérogènes, montrent que la présence du désordre dans ce système peut conduire à plusieurs transitions de phase successives, au lieu d'un seul point critique isolé. Ceci est en accord avec l'analyse de Imry et Wortis concernant les transitions de phase dans les systèmes hétérogènes [23], selon laquelle l'apparition des domaines de cohérence est toujours favorisée pour les systèmes unidimensionnels (rapellons que la dimension effective des modèles étudiés dans cette thèse est le plus souvent égale à un). Par conséquent, la question portant sur l'ordre de la transition pour une séquence hétérogène doit être posée plus clairement. En fait, la seule question ayant un sens du point de vue physique est de comprendre la nature de la transition correspondant à chacune de ces dissociations locales. En adaptant une version généralisée de la méthode de l'intégrale de transfert au modèle JB hétérogène, j'ai étudié la dénaturation des séquences d'ADN hétérogènes dans le cadre de la physique statistique : j'ai montré que pour une séquence particulière le désordre peut effectivement conduire à plusieurs transitions de phase successives, ce qui se manifeste par plusieurs pics, ou bien un seul pic asymétrique dans le tracé de la chaleur spécifique. Le calcul des

exposants caractéristiques permettant d'estimer l'ordre des transitions en question étant actuellement en cours, je ne discute dans le chapitre 6 que les résultats déjà obtenus.

Les résultats portant sur l'ADN obtenus dans ce travail de thèse ont fait l'objet de 5 articles publiés dans les revues internationales :

– *Dynamical model based on finite stacking enthalpies for homogeneous and inhomogeneous DNA thermal denaturation*, M. Joyeux et S. Buyukdagli, Physical Review E, 72, 051902 (2005)

– *Towards more realistic dynamical models for DNA secondary structure*, S. Buyukdagli, M. Sanrey et M. Joyeux, Chemical Physics Letters, 419, 434 (2006)

– *Scaling laws at the phase transition of systems with divergent order parameter and/or internal length : the example of DNA denaturation*, S. Buyukdagli et M. Joyeux, Physical Review E, 73, 051910 (2006)

– *1/f fluctuations of DNA temperature at thermal denaturation*, M. Joyeux, S. Buyukdagli et M. Sanrey, Physical Review E, 75, 061914 (2007)

– *Theoretical investigation of finite size effects at DNA melting*, S. Buyukdagli et M. Joyeux, Physical Review E, 76, 021917 (2007)

En tout début de thèse, avant de m'intéresser à l'ADN, j'avais également travaillé sur la comparaison entre la dynamique classique et quantique d'un modèle de petite molécule. Je ne développerai pas ce travail dans ce manuscrit, parce qu'il est trop différent du reste. La référence de l'article correspondant est la suivante :

– *On the application of canonical perturbation theory up to the dissociation threshold*, S. Buyukdagli et M. Joyeux, Chemical Physics Letters, 412, 200 (2005)

1. UNE BRÈVE INTRODUCTION À LA STRUCTURE ET AUX FONCTIONS DE L'ADN

Dans la nature, les processus biologiques ayant lieu à l'échelle cellulaire sont conduits par des molécules appelées *protéines*. Les protéines sont des polymères d'une dizaine de nanomètres. Elles servent par exemple de catalyseurs pour fournir l'énergie nécessaire aux réactions biochimiques et de canaux qui alimentent les cellules.

L'ADN (acide désoxyribonucléique) est précisément le "disque dur" qui contient l'information nécessaire à la fabrication de ces molécules vitales et qui transmet notre information génétique à travers les générations. La structure statique de l'ADN admise aujourd'hui a été proposée par Watson et Crick en 1953 [1]. La molécule d'ADN est un double brin de polymère enroulé sous forme hélicoïdale (voir la figure 1.1). Les monomères de ce long polymère sont des *nucléotides*. Chaque nucléotide est formé par un groupe phosphate, un cycle de sucre à cinq atomes (un ribose) et un acide nucléique. La succession de sucres-phosphates constitue le squelette de la molécule d'ADN. Il existe deux catégories d'acides nucléiques, à savoir les *purines* et les *pyrimidines*. *Adénine* (A) et *Guanine*(G) sont des purines et *Cytosine* (C) et *Thymine* (T) des pyrimidines. Les purines sont constituées d'un cycle à cinq atomes lié à un cycle à six atomes et les pyrimidines d'un cycle à six atomes. Pour que les liaisons hydrogène se forment correctement, chaque guanine est liée à une cytosine par une triple liaison hydrogène et chaque adénine à une thymine par une double liaison. De ce fait, une liaison $G \equiv C$ est plus solide qu'une liaison $A = T$ et les deux brins sont *complémentaires* l'un de l'autre (voir la figure 1.2). Le complexe sucre-phosphate est orienté puisque deux phosphates liés à un ribose sont attachés à des atomes de carbone de natures différentes. Les deux extrémités de ce monomère orienté s'appellent 3' et 5'. Nous allons voir dans la suite à quel niveau intervient

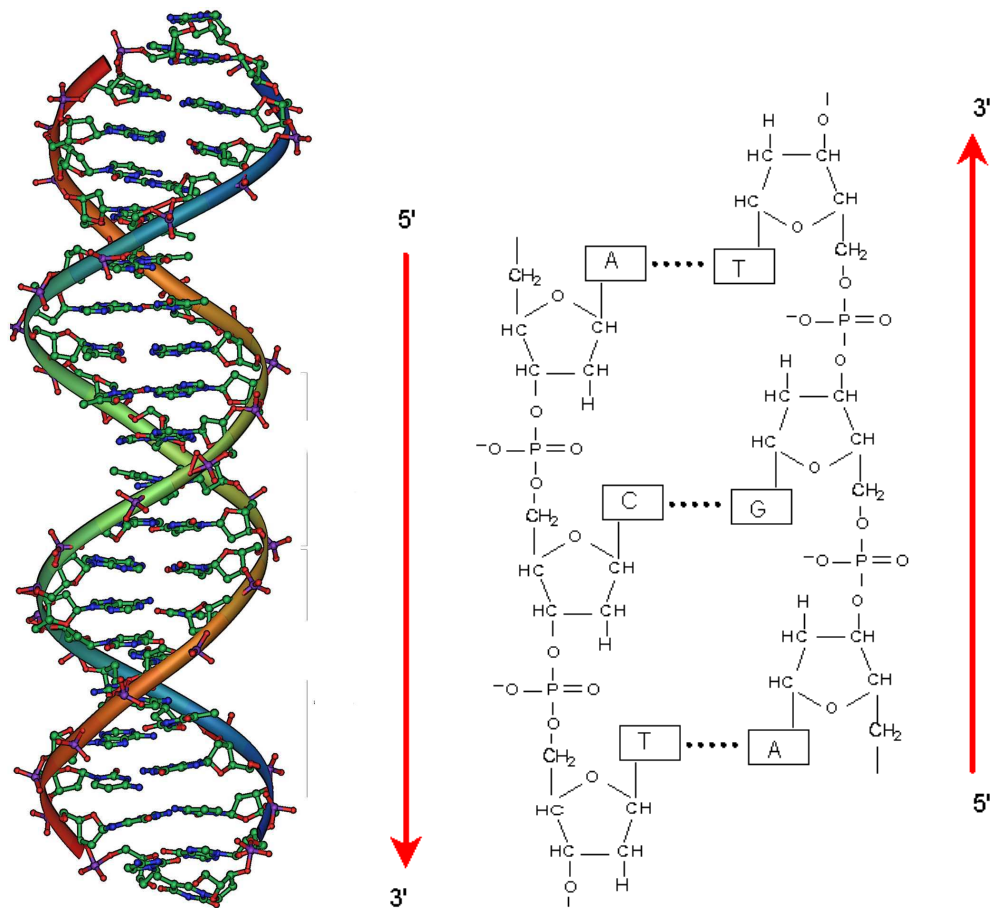


Fig. 1.1: La structure hélicoïdale de la molécule d'ADN (à gauche) et sa composition chimique (à droite). La première figure est téléchargée de Wikipedia et la deuxième du moteur de recherche d'image de yahoo.

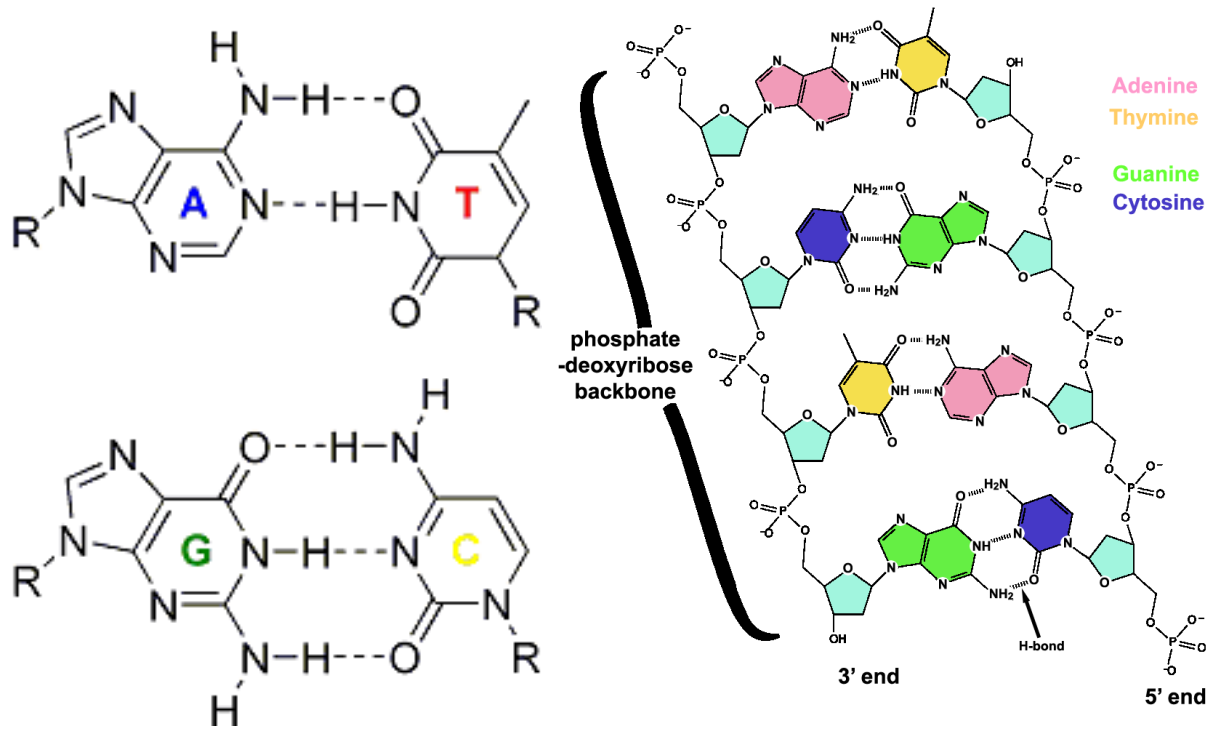


Fig. 1.2: La composition des deux types de paire d'acides nucléiques A-T et G-C (à gauche) et leur position sur le double brin (à droite). Les figures sont téléchargées du moteur de recherche d'image de yahoo.

cette direction particulière.

La succession d'acides nucléiques constitue l'information génétique et on l'appelle la *structure primaire* de l'ADN ou le *génome*. Chaque base est liée à la fois à un carbone latéral du ribose du même brin et à une base appartenant à l'autre brin. Dans la suite, nous appellerons cet agencement une *paire de bases*. Les deux brins complémentaires liés par des liaisons hydrogène constituent la *structure secondaire* de l'ADN. Il existe une catégorie conformationnelle supplémentaire que l'on appelle la *structure tertiaire* et qui désigne la forme en hélice que prennent conjointement les deux brins complémentaires. En fait, la double hélice peut adopter de nombreuses formes, dont les plus fréquemment rencontrées sont appelées A, B et Z. Ces diverses conformations diffèrent principalement par la position relative des paires de bases et l'inclinaison des plateaux par rapport à l'axe de l'hélice. Ce degré de conformation supplémentaire et d'autres structures d'ordre supérieur ne seront pas pris en compte dans cette thèse.

A l'échelle des paires de base, en plus des liaisons hydrogène, il existe une seconde

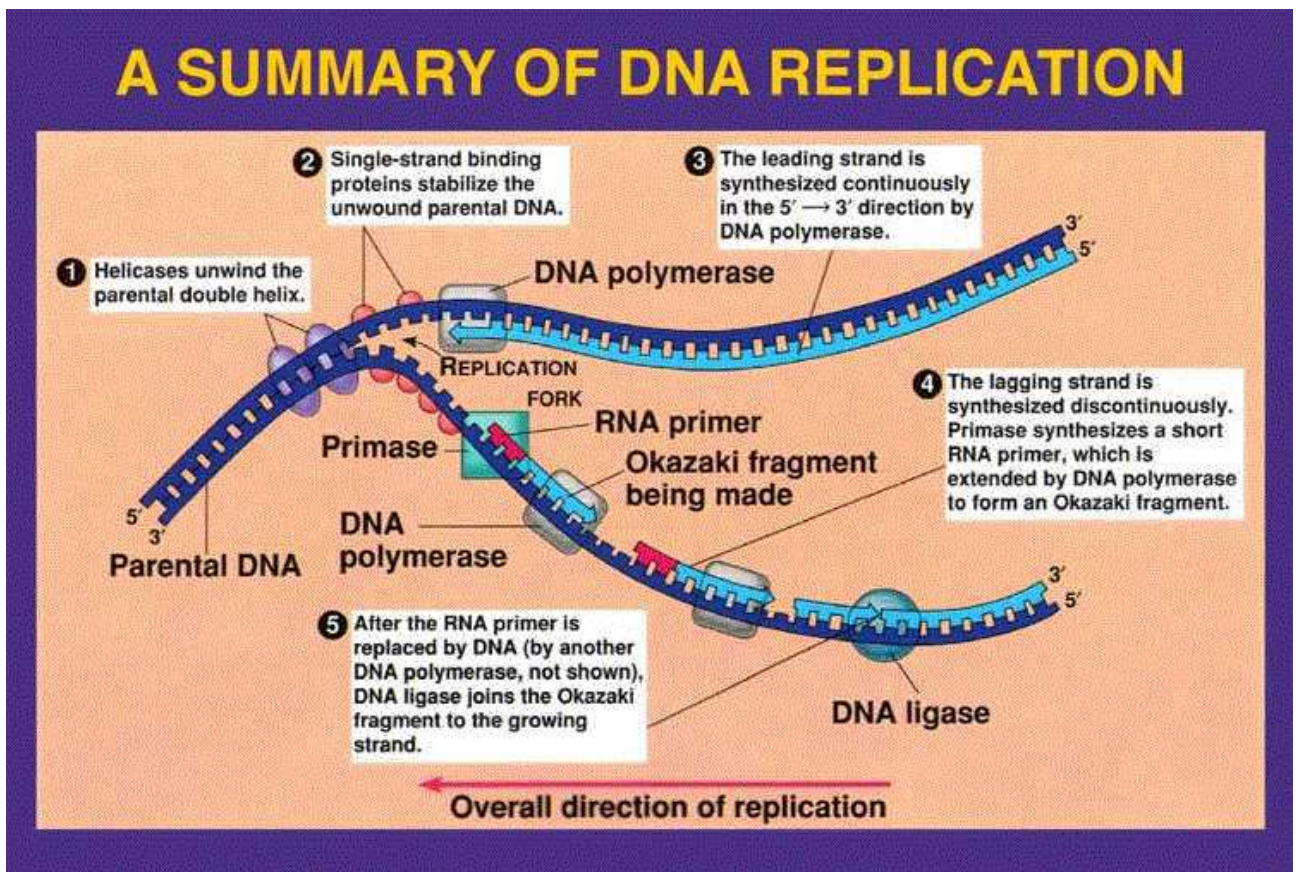


Fig. 1.3: Le processus de réplication. La figure est téléchargée du moteur de recherche d'image de yahoo.

énergie caractéristique, à savoir l'énergie d'empilement. C'est une interaction attractive due aux forces hydrophobes et au recouvrement des électrons π des paires de bases successives. En fait les plateaux successifs ont tendance à se rapprocher pour empêcher les molécules d'eau de pénétrer entre eux. C'est précisément cette tendance qui donne à la molécule sa structure tertiaire.

Il existe à l'échelle cellulaire deux processus biologiques très importants, à savoir la *réplication* et la *transcription-traduction*. Les mécanismes conduisant ces processus étant d'une sophistication impressionnante, je vais les décrire succinctement.

La réplication a lieu lors de la mitose (la division cellulaire) et est conduite par un moteur moléculaire appelé l'*ADN polymérase* (ADN-pol). L'ADN-pol se pose sur le brin principal et le lit d'un bout à l'autre dans le sens $3' \rightarrow 5'$ derrière les hélicases qui écartent le double brin. Lors de la lecture il produit une copie exacte du brin complémentaire de celui sur lequel il se déplace. A la fin du processus, les deux brins sont complètement

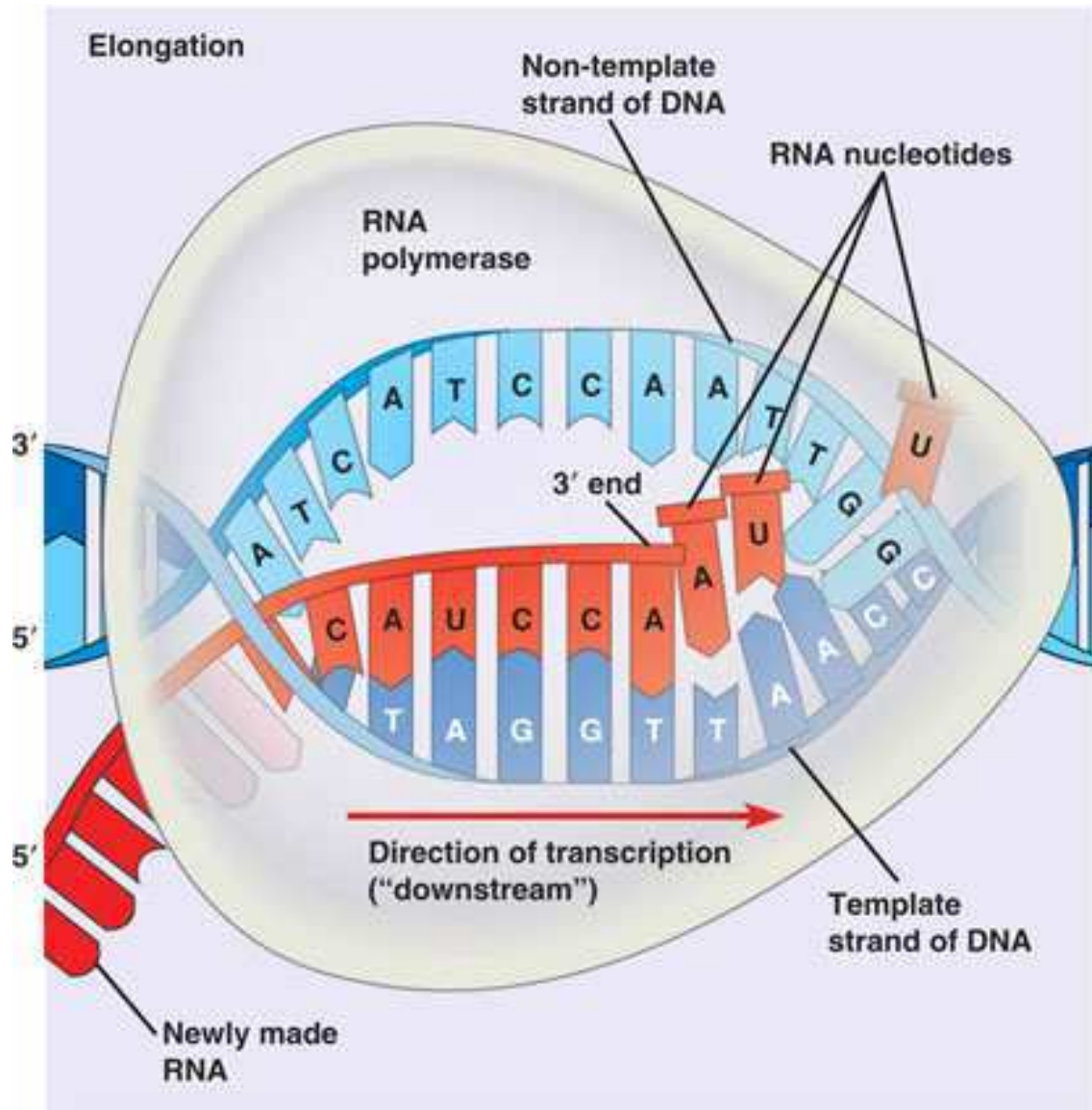


Fig. 1.4: Le processus de transcription. La figure est téléchargée du moteur de recherche d'image de yahoo.

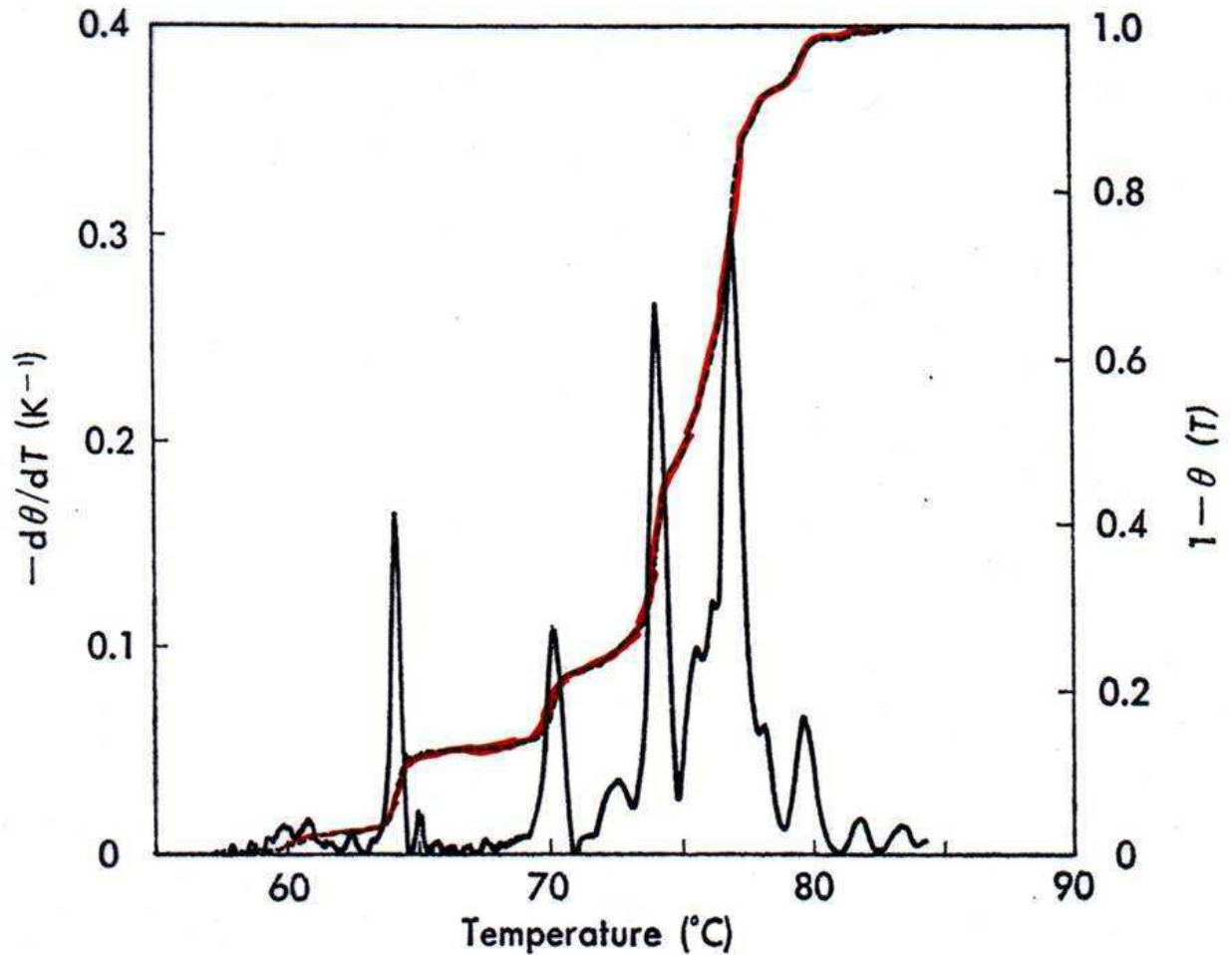


Fig. 1.5: Courbe de dénaturation thermique d'une séquence d'ADN hétérogène, le plasmide pNT1. La courbe qui croît continuellement est le coefficient d'absorption UV, qui est proportionnel à la fraction de paires de bases ouvertes. Les sauts discontinus, correspondant aux pics de la courbe de dénaturation différentielle, indiquent les dislocations locales. La courbe est prise de la réf. [24].

séparés et sont chacun inclus dans une copie exacte de la séquence originale. Lors du déplacement de l'ADN-pol sur le brin principal, une lecture du même type mais un peu plus complexe a lieu sur le brin complémentaire (voir la figure 1.3). L'ADN-pol opère avec une très grande précision. Chaque fois qu'il lit une base, il n'avance en principe que si la base complémentaire a été correctement captée dans le solvant et ajoutée à la chaîne de reproduction. Supposons qu'il vienne exceptionnellement d'inclure un mauvais nucléotide et avance d'une base. Le mauvais nucléotide qu'il a laissé derrière lui sera mal hybridé, ce qui arrête la progression de l'ADN-pol. Dans ce cas il recule d'un pas, inclut le bon nucléotide et redémarre. Le taux d'erreur de l'ADN-pol est de 10^{-9} . C'est

cette précision qui assure une transmission fidèle de l'information génétique à travers les générations. Une description un peu plus complète des mécanismes et enzymes mis en jeu durant la réplication est proposée en figure 1.3.

La transcription est le processus de lecture de l'information génétique pour la fabrication des protéines. Cela se fait par un moteur moléculaire que l'on appelle l'*ARN polymérase* (ARN-pol). Sur une molécule d'ADN, il existe des portions que l'on appelle des *gènes* et qui codent la fabrication d'une protéine particulière. Au début de chaque gène se trouve une séquence appelée le *promoteur*. Chaque fois qu'une cellule a besoin d'une protéine, l'ARN-pol est attiré par le promoteur marquant le début du gène qui correspond à la protéine en question. En partant du promoteur, l'ARN-pol avance sur le gène en écartant localement les deux brins de façon à former une bulle d'une dizaine de paires de base. C'est la *bulle de transcription* qui se refermera une fois l'ARN-pol éloigné. Au fur et à mesure qu'il lit le gène, il produit un brin complémentaire que l'on appelle l'*ARN-messenger* (ARNm). L'ARN-pol arrête la transcription dès qu'il rencontre la séquence particulière qui se trouve à la fin du gène et qui lui donne l'ordre de s'arrêter. Après une éventuelle phase d'épissage (excision des introns chez les eucaryotes), le processus de traduction démarre. L'ARNm est conduit à une grosse usine moléculaire, le *ribosome*, chargé de la fabrication de la protéine. L'ARNm entre dans le ribosome par le côté 5'. Chaque triplet de bases de l'ARNm code un acide aminé particulier de la protéine. Par conséquent la lecture de l'ARNm se fait par triplets. Chaque fois que le ribosome lit un triplet, il arrête la lecture, capte l'acide aminé correspondant dans la solution et l'ajoute à sa chaîne de production. Dès que la tête de lecture du ribosome rencontre le triplet lui donnant l'ordre de s'arrêter, la protéine est libérée dans la solution.

Au delà de ces processus vitaux, l'ADN est une molécule qui a une dynamique vibrationnelle et rotationnelle très active aux températures biologiques. De plus, les expériences effectuées avec des brins d'ADN montrent que lorsque cette molécule est posée dans une solution et chauffée jusqu'à une température caractéristique, les liaisons hydrogène se cassent, le double brin se sépare en deux et la molécule se dissocie. C'est ce

processus que l'on appelle la *dénaturation thermique* de l'ADN. Cette dénaturation peut être également réalisée en changeant les caractéristiques chimiques du solvant capables de déstabiliser les liaisons hydrogène, comme le pH ou la quantité de sel. La température de fusion ou de dénaturation de l'ADN est la température pour laquelle 50% des paires de bases de la molécule deviennent désappariées à travers la rupture des liaisons hydrogène par effet thermique. L'énergie de liaison d'une paire de bases est évidemment d'autant plus importante que le nombre de liaisons hydrogène est élevé. Par conséquent, la rupture d'une liaison $A-T$ sera plus facile que celle d'une liaison $G-C$. C'est la raison pour laquelle la température de fusion de l'ADN est une fonction de la taille et surtout de la fraction de paires de base $A-T$ (ou $G-C$) de la molécule. Comme on le verra dans la suite de ce travail, l'hétérogénéité des énergies d'empilement joue toutefois un rôle similaire et au moins aussi important que celui des liaisons hydrogène dans la détermination de la température de dénaturation d'une séquence donnée.

La transcription, qui nécessite l'accès à l'information génétique caché dans la structure liée de l'ADN, est obligatoirement accompagnée par l'ouverture locale du double brin par des hélicases. Autrement dit, la dynamique interne de l'ADN est directement impliquée lors du processus de la transcription. Par conséquent, afin de décrire de façon rigoureuse l'interaction qui a lieu entre les moteurs moléculaires et l'ADN au niveau théorique tout en tenant compte des détails microscopiques du système réel, un premier pas inévitable consiste à modéliser la dynamique structurale de l'ADN à des températures allant de la température physiologique à la température de dissociation. Une seconde raison conduisant les physiciens à s'intéresser à l'étude théorique de la dénaturation thermique est l'effort expérimental fait au cours des 30 dernières années. La dénaturation de l'ADN peut en effet être étudiée expérimentalement par spectroscopie d'absorption UV à 260 nm. Le coefficient d'absorption optique de la molécule d'ADN augmente d'un facteur 1.4 dès que le double brin se dissocie en simples brins. C'est ce que l'on appelle l'*effet hyperchrome*. Or, les expériences effectuées sur l'ADN montrent que les courbes d'absorption UV de cette molécule possèdent un certain nombre de pics

abrupts à des températures caractéristiques faisant penser à une chaîne de transitions de phase discontinues (voir la figure 1.5) . Cette accumulation de données expérimentales permet par conséquent de tester et comparer différents modèles théoriques de l'ADN.

2. MODÈLES DYNAMIQUES POUR LA DÉNATURATION DE L'ADN

2.1 *Modèle de Poland Scheraga*

Les premiers modèles de dénaturation thermique considéraient l'ADN, de façon semblable au modèle d'Ising, comme un système dans lequel une paire de bases ne peut prendre que deux états de configuration, à savoir un état fermé où les liaisons hydrogène connectant les bases sont intactes, et un état ouvert où ces liaisons sont cassées. Les ingrédients principaux de ce type de modèle discret sont l'énergie de liaison hydrogène G_i d'une paire de bases et l'énergie d'empilement $G_{i,i-1}^s$ entre les paires successives. Cela donne une description de l'ouverture d'une paire de bases i en terme d'une constante de réaction ayant la forme

$$s_i = \exp \left[-\frac{G_i + G_{i,i-1}^s}{k_B T} \right] \quad (2.1)$$

où T est la température de la solution et k_B la constante de Boltzmann. Le modèle le plus connu dans cette catégorie est le modèle de Poland-Scheraga (PS)[5]. Les divers états configurationnels de ce modèle sont donnés sur la figure 2.1. Dans le modèle PS, la fonction de partition Z_k d'une configuration k donnée est le produit de différents poids statistiques, à savoir

- un terme de stabilité de la forme 2.1 pour chaque paire de bases fermée,
- un terme de coopérativité σ_c entre un segment fermé et un segment ouvert,
- un terme $f(m)$ correspondant à l'entropie de boucle pour une bulle de longueur m .

Dans la littérature, les valeurs numériques admises pour σ_c varient entre 10^{-4} et 10^{-5} [6] tandis que pour l'entropie de boucle, on prend habituellement une loi de puissance de la forme $f(m) = (m + D)^{-c}$ où D et c sont des paramètres empiriques. En général,

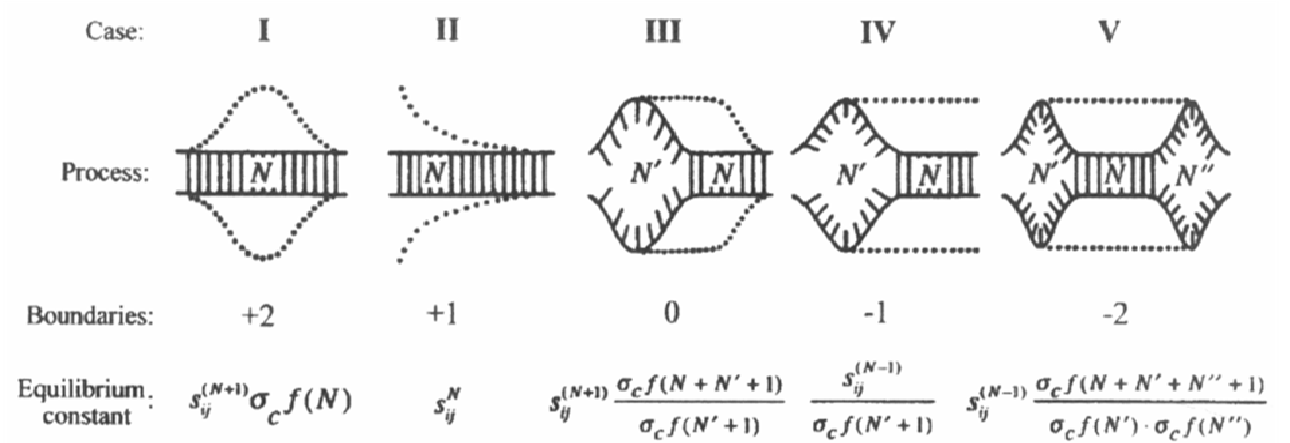


Fig. 2.1: representation de cinq processus de dissociation (parmi les six possibles) pour le modèle PS.

on considère $D = 1$ et $c = 1.7$ (transition de phase du deuxième ordre) ou $c = 2.15$ (transition de phase du premier ordre).

Divers modèles basés sur le même principe ont été construits par la suite et certains d'entre eux sont toujours employés afin de reproduire de façon numérique les courbes de dénaturation expérimentales de façon très fidèle. Le plus grand avantage du modèle PS, par rapport aux modèles dynamiques discutés dans la suite de cette thèse, est sans doute une réduction importante du temps de calcul. L'algorithme basé sur le modèle PS le plus connu est MELTSIM [17]. Avec l'algorithme MELTSIM exact, le temps de calcul d'une courbe de dénaturation pour une séquence de longueur N est en principe proportionnel à N^2 , mais ce temps devient proportionnel à N grâce à l'utilisation de la relation de la récursion de Poland [18]. En utilisant le programme MELTSIM, la courbe de dénaturation d'une séquence donnée de moins de 10000 paires de bases peut

être obtenue en quelques secondes, tandis que pour les modèles Hamiltoniens qui seront présentés dans ce travail de thèse, le temps de calcul d'un profil de dissociation, dont on extrait un seul point de la courbe de dénaturation, peut varier entre une heure et deux jours.

Modèles Hamiltoniens	Modèle PS
Information microscopique	
<ul style="list-style-type: none"> - 10 interactions d'empilement - forme de l'interaction d'empilement : gaussienne de largeur b - une enthalpie D pour les liaisons H - forme de l'interaction à un corps : Morse de largeur a - rigidité du squelette : potentiel harmonique de raideur K_b 	<ul style="list-style-type: none"> - 10 interactions d'empilement - 2 enthalpies ΔH_i pour les liaisons H - terme de coopérativité σ
Effets entropiques	
	<ul style="list-style-type: none"> - 10 entropies d'empilement ΔS_{ij} - 2 entropies ΔS_i pour les paires de bases
Information macroscopique	
	- entropie de boucle $f(m)$ pour les bulles

Tab. 2.1: Comparaison qualitative des modèles Hamiltoniens et des modèles statistiques du type Poland-Scheraga.

Les modèles Hamiltoniens diffèrent du modèle PS essentiellement par le caractère des informations que l'on y impose. Cela est résumé dans le tableau 2.1. Dans les deux modèles, on introduit pour chaque paire de bases des paramètres énergétiques correspondant aux liaisons hydrogène et aux interactions d'empilement. En revanche, les termes de coopérativité, les entropies par base et les entropies de boucle sont ajoutés au modèle PS de façon phénoménologique, tandis que dans les modèles Hamiltoniens, cette information est remplacée par une description explicite de la forme des potentiels d'interaction. Nous allons analyser dans la suite ces modèles Hamiltoniens en détail.

Le modèle PS a été utilisé avec succès dans les études théoriques concernant l'ordre de la transition de dénaturation [19] et les effets de l'hétérogénéité sur l'ordre de la transition [20].

2.2 Modèles de Prohofsky et de Dauxois, Peyrard et Bishop

Bien que les modèles phénoménologiques discrets comme le modèle PS décrivent très bien la dénaturation de l'ADN, l'information que l'on peut en tirer est assez limitée, à cause de la nature discrète des états accessibles. En particulier, ils n'apportent aucun éclaircissement quant à la dynamique temporelle de l'ADN. Par contre un modèle Hamiltonien continu peut être capable de décrire la dynamique et la physique statistique de l'ADN à tous les niveaux énergétiques s'étendant des basses températures, où la molécule effectue des mouvements de faible amplitude, jusqu'à des températures avoisinant celle de la dissociation, où la molécule effectue des vibrations et des torsions de grande amplitude. Les premiers modèles ayant ces caractéristiques ont été proposés afin d'étudier la propagation des solitons. En 1980, Englander et ses collègues ont effectué des expériences d'échange hydrogène-deuterium [25] qui leur ont permis d'observer des états d'ouverture partielle qui se déplaçaient le long de la séquence d'ADN de façon semblable aux solitons. Englander a donc proposé le premier modèle [25] ayant des solutions solitons sous la forme

$$H = \sum_n \left\{ \frac{1}{2} m r^2 \dot{\phi}_n^2 + \frac{K}{2} (\phi_n - \phi_{n-1})^2 + m g r (1 - \cos \phi_n) \right\} \quad (2.2)$$

où r est la longueur (supposée constante) du segment qui relie l'acide nucléique au squelette sucre-phosphate et ϕ est l'angle de rotation de ce segment autour du brin. Le premier terme de l'hamiltonien est l'énergie cinétique rotationnelle, le deuxième terme correspond à l'énergie élastique du squelette et le dernier terme est un potentiel attractif à un corps représentant les liaisons hydrogène. Ce modèle un peu trop simpliste a été suivi par d'autres modèles plus réalistes (le modèle de Yomosa [26], de Zhang [27] etc...) décrivant les liaisons hydrogène et l'interaction d'empilement de façon plus correcte, pour lesquels des solutions de type soliton ont également été étudiées en détail [28, 29, 30, 31]. Cependant, ces modèles ne sont pas capables de décrire la dénaturation thermique puisque le seul degré de liberté dont ils tiennent compte est la rotation d'une base autour

du brin tandis que le degré de liberté décrivant la dissociation est l'élongation relative des paires de bases. En 1984, Prohofsky et ses collaborateurs ont proposé le premier modèle d'ADN tenant compte de ce fait afin d'étudier la dénaturation par la méthode des phonons self-consistent [7] et l'existence des solitons [32]. Peyrard et Bishop ont utilisé ce modèle, qui s'écrit

$$H = \sum_n \left\{ \frac{m}{2} (\dot{u}_n^2 + \dot{v}_n^2) + \frac{k}{2} [(u_n - u_{n-1})^2 + (v_n - v_{n-1})^2] + V_M(u_n - v_n) \right\} \quad (2.3)$$

afin d'étudier la dissociation de l'ADN de façon plus poussée. Dans cet hamiltonien, les coordonnées u_n et v_n représentent les déviations par rapport à l'équilibre de la position des deux bases de la n -ième paire sur l'axe perpendiculaire à la séquence et joignant les deux brins - c'est à dire le long des barreaux si l'on considère la séquence comme une échelle. Le premier terme est l'énergie cinétique d'élongation. Le deuxième terme représente l'interaction d'empilement entre les paires de bases successives, dont l'origine physique a été expliquée dans le chapitre 1. Le modèle suppose donc une interaction aux plus proches voisins de forme harmonique. Finalement, le dernier terme, qui correspond aux liaisons hydrogène entre les bases d'une même paire, est exprimé sous la forme d'un potentiel de Morse

$$V_M(u_n - v_n) = D \left(1 - e^{-a(u_n - v_n)/\sqrt{2}} \right)^2. \quad (2.4)$$

En passant des coordonnées absolues u_n et v_n aux coordonnées des modes symétriques et antisymétriques

$$x_n = (u_n + v_n)/\sqrt{2}, \quad y_n = (u_n - v_n)/\sqrt{2}, \quad (2.5)$$

il est possible de séparer cet hamiltonien sous la forme

$$H = H_1(x_1, x_2, \dots, x_N) + H_2(y_1, y_2, \dots, y_N), \quad (2.6)$$

où

$$H_1 = \sum_n \left\{ \frac{M}{2} \dot{x}_n^2 + \frac{K}{2} (x_n - x_{n-1})^2 \right\} \quad (2.7)$$

est la partie quadratique qui décrit le déplacement du centre de masse de chaque paire de bases, alors que le deuxième terme

$$H_2 = \sum_n \left\{ \frac{M}{2} \dot{y}_n^2 + \frac{K}{2} (y_n - y_{n-1})^2 + D (1 - e^{-ay_n})^2 \right\} \quad (2.8)$$

possède l'information entière sur la dynamique dissociative de l'ADN. Cet hamiltonien possède également des solutions solitons qui ont été étudiées en détail [10, 11]. Dans ce travail de thèse nous n'avons pas abordé la question des solitons.

Le modèle (2.8) a un inconvénient majeur. Pour des valeurs réalistes des paramètres du système, à savoir $M = 300$ amu, $D = 0.03$ eV, $K = 0.06$ eV \AA^{-2} et $a = 4.5$ \AA^{-1} (on rappelle que 1 eV ≈ 96.5 kJ/mol), les courbes de dénaturation théoriques obtenues sont en désaccord avec les courbes d'absorption expérimentales, puisque les courbes théoriques montrent un processus de dissociation qui s'étend sur un intervalle de température très étendu, au lieu d'une transition abrupte. Il est important de préciser qu'une transition de phase étant un processus collectif, c'est la forme du potentiel couplant les différents degrés de liberté qui va déterminer le type de la transition. Autrement dit, le problème de ce modèle vient de la forme de l'interaction d'empilement. Une fonction quadratique est une fonction continue qui croît de façon monotone, alors que l'interaction d'empilement doit décroître rapidement lorsque les paires de bases adjacentes sont très éloignées les unes des autres au voisinage de la dénaturation. Pour corriger ce défaut du modèle, Dauxois, Peyrard et Bishop ont proposé un modèle amélioré sous la forme

$$H = \sum_n \left\{ \frac{M}{2} \dot{y}_n^2 + \frac{K}{2} (y_n - y_{n-1})^2 [1 + \rho e^{-\alpha(y_n - y_{n-1})}] + D (1 - e^{-ay_n})^2 \right\}. \quad (2.9)$$

L'avantage de cette nouvelle interaction d'empilement anharmonique est que, grâce au terme exponentiel, le terme d'interaction chute de $K(1+\rho)$ à K lorsque les bases appariées s'écartent, ce qui décroît la rigidité de la chaîne d'ADN près de la dissociation et conduit à une transition de phase abrupte du premier ordre.

Dans la suite de ce mémoire, j'appellerai "modèle PB" ou bien "modèle PB harmonique" le modèle de l'éq. (2.8) et "modèle DPB" ou bien "modèle PB anharmonique" le modèle de l'éq. (2.9).

2.3 Nouveau modèle basé sur des énergies d'empilement finies

Comme je l'ai précisé, l'un des enjeux de ce travail est de développer un modèle d'ADN continu capable de reproduire les propriétés des courbes de dénaturation expérimentales. Dans cette section, je vais introduire un nouveau modèle d'ADN [33] et montrer qu'il est possible d'obtenir une transition de phase discontinue avec une interaction d'empilement finie. Cette interaction, qui est due au recouvrement des électrons π et aux interactions hydrophobes entre les paires de bases, doit naturellement tendre vers zéro quand les bases adjacentes s'éloignent l'une de l'autre. Cela est pris en compte dans le modèle de Poland-Scheraga (PS), mais pas dans le modèle DPB puisque dans ce cas l'interaction entre les bases diverge quand les plateaux s'écartent.

Un potentiel simple, qui tient compte de la finitude de l'interaction d'empilement peut s'écrire

$$W(y_n, y_{n-1}) = \min \left[\frac{\Delta H}{2}, \frac{K}{2} (y_n - y_{n-1})^2 \right] \quad (2.10)$$

où $\Delta H/2$ est la limite supérieure de l'énergie d'empilement. C'est le premier type de potentiel que nous avons envisagé et les résultats qui seront présentés dans la suite restent

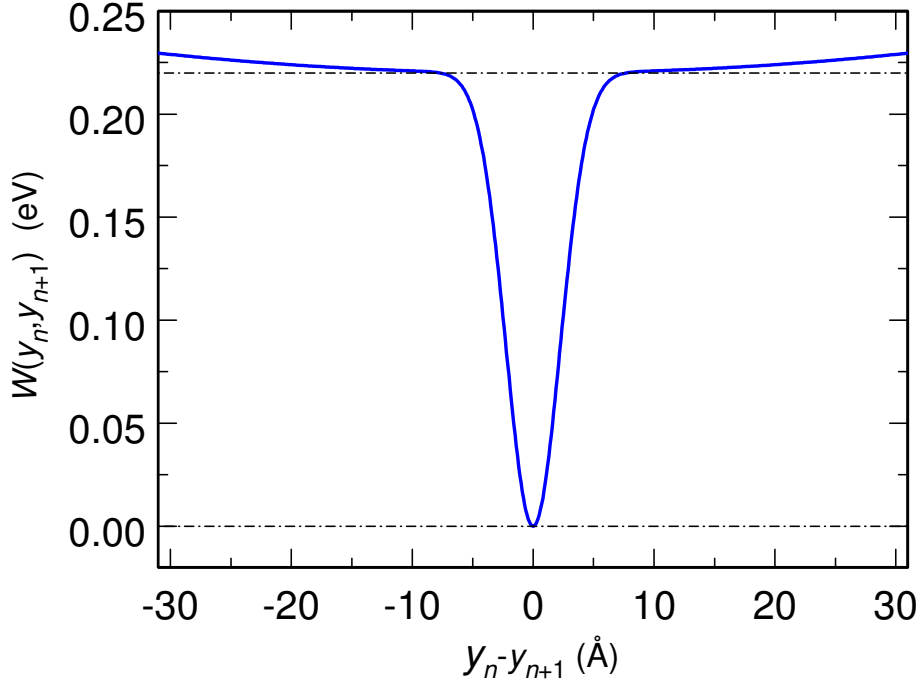


Fig. 2.2: Energie d'empilement $W(y_n - y_{n-1})$ de l'équation (2.11) en fonction de la distance $y_n - y_{n-1}$

valables avec ce potentiel. Plus précisément, on obtient des courbes de dénaturation et une entropie discontinues, ce qui est signe d'une transition de phase du premier ordre. Cela nous conduit à conclure que la prise en compte de la finitude de l'énergie d'empilement est suffisante pour obtenir une transition discontinue.

Nous avons cependant modifié ce terme d'interaction pour la suite de l'analyse. Premièrement, nous avons remplacé la fonction $\min()$ par une fonction plus physique, à savoir par un terme exponentiel. De plus, nous avons remarqué que l'utilisation de l'éq. (2.10) conduit, au voisinage de la température critique, à de grandes valeurs de y_n (ce qui est normal), mais aussi de $y_n - y_{n-1}$ (ce qui l'est beaucoup moins). Autrement dit, les brins qui doivent rester intacts après la dissociation se brisent en morceaux. Afin d'éliminer cet effet, nous avons introduit un terme harmonique qui intègre dans le modèle la rigidité du squelette phosphate-sucre. Finalement le terme d'interaction obtenu peut s'écrire

$$W(y_n, y_{n-1}) = \frac{\Delta H}{2} \left(1 - e^{-b(y_n - y_{n-1})^2} \right) + K_b (y_n - y_{n-1})^2. \quad (2.11)$$

On peut remarquer que l'interaction (2.11) dépend uniquement de la distance entre les

plateaux successifs, ce qui n'est pas le cas du modèle DPB. Précisons également que nous avons choisi le terme K_b 2000 fois plus petit que le paramètre K du modèle DPB, ce qui permet de conserver des valeurs de $(y_n - y_{n-1})$ raisonnablement basses sans changer de façon dramatique la dynamique de la dissociation.

Dans la suite de ce manuscrit, j'appellerai "modèle JB" les modèles basés sur une énergie d'empilement du type de l'éq. (2.11).

Dans cette section 2.2, les valeurs numériques choisies pour les paramètres du modèle DPB sont celles des références [8, 9], à savoir $m = 300$ amu, $D = 0.04$ eV, $a = 4.45 \text{ \AA}^{-1}$, $K = 0.04 \text{ eV \AA}^{-2}$, $\rho = 0.5$ et $\alpha = 0.35 \text{ \AA}^{-1}$. Les valeurs prises pour les paramètres de l'équation (2.11) sont $\Delta H = 0.44 \text{ eV}$, $b = 0.10 \text{ \AA}^{-2}$ et $K_b = 10^{-5} \text{ eV \AA}^{-2}$ (on rappelle que $1 \text{ eV} \approx 96.5 \text{ kJ/mol}$). Le choix particulier du paramètre b a été fait pour avoir le terme d'ordre quadratique dans le développement de l'équation (2.11) égal à celui du modèle DPB, à savoir $K/2$. La figure 2.2 montre l'évolution de $W(y_n, y_{n-1})$ en fonction de $y_n - y_{n-1}$. On peut constater que la finitude de l'interaction d'empilement conduit à un plateau bien marqué malgré l'existence du terme quadratique supplémentaire.

Sur la figure 2.3, on compare la courbe de dissociation du modèle JB avec celles des modèles PB harmonique et DPB anharmonique. Les symboles correspondent aux résultats obtenus par la méthode de l'intégrale de transfert (TI) pour les trois modèles à la limite thermodynamique, autrement dit pour une séquence de longueur infinie (voir le chapitre 4 pour les détails du calcul TI dans ce cas précis). On constate sur cette courbe que le modèle proposé présente une dissociation thermique nettement discontinue, comme le modèle DPB anharmonique. Puisqu'il existe un certain doute sur la validité de la méthode TI pour des potentiels décrivant une dynamique liée suivie par un régime libre [15], nous avons également calculé l'élongation moyenne $\langle y \rangle$ par des simulations de dynamique moléculaire (MD) pour le modèle JB, avec une séquence de 2399 paires de base et des conditions aux bords périodiques. Cela consiste à intégrer numériquement l'équation de Langevin

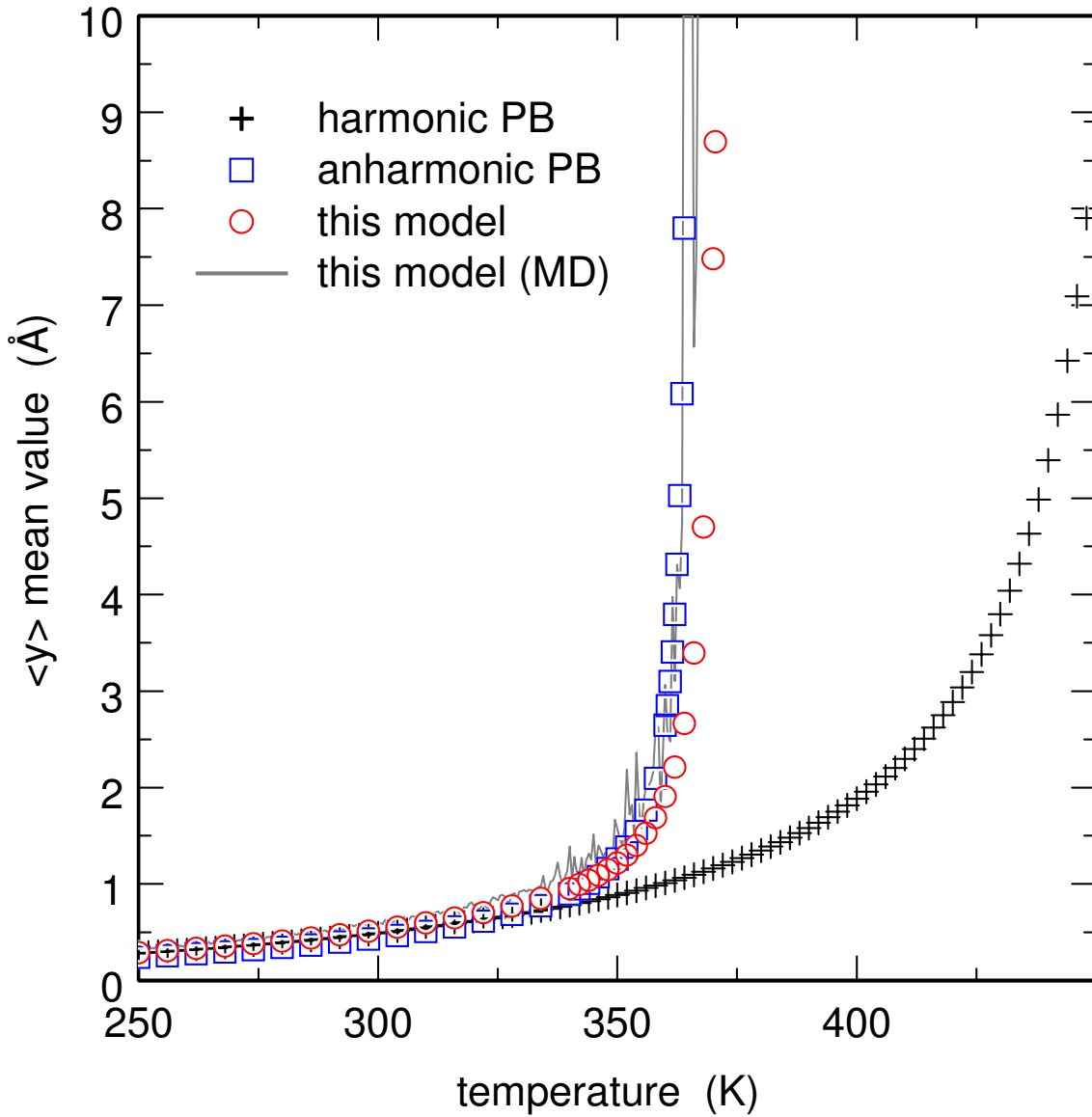


Fig. 2.3: Distance moyenne $\langle y \rangle$ entre les bases d'une même paire en fonction de la température. Les symboles montrent le résultat du calcul TI d'une séquence homogène à la limite thermodynamique pour le modèle PB harmonique (2.8) (croix), le modèle DPB anharmonique (2.9) (carrés) et le modèle JB présenté dans cette partie (2.11) (cercles). La courbe continue correspond au calcul MD pour une séquence homogène de 2399 paires de bases et le modèle JB.

$$m \frac{d^2 y_n}{dt^2} = - \frac{\partial H}{\partial y_n} - m\gamma \frac{dy_n}{dt} + \sqrt{2mk_B T} w(t). \quad (2.12)$$

Dans cette équation, $\gamma = 5 \text{ ns}^{-1}$ est le coefficient de dissipation et $w(t)$ est une fonction aléatoire gaussienne de moyenne zéro et de variance unité. Le deuxième et le troisième terme à droite de l'équation modélisent l'interaction entre la chaîne d'ADN et les particules de la solution. Cette équation a été intégrée en utilisant l'algorithme de Brünger-Brooks-Karplus [34] du second ordre,

$$\begin{aligned} \frac{m}{\tau^2} (y_n^{k+1} - 2y_n^k + y_n^{k-1}) = & - \left. \frac{\partial H}{\partial y_n} \right|_{k\tau} - \frac{m\gamma}{2\tau} (y_n^{k+1} - y_n^{k-1}) \\ & + w(n, k) \sqrt{\frac{2m\gamma k_B T}{\tau}}, \end{aligned} \quad (2.13)$$

avec un pas discret de 10 fs. L'intégration a été effectuée avec une température initiale de 0 K qui a été augmentée à un taux de 20 ns/K. Cette montée est suffisamment lente pour que la température moyenne du système, déterminée par l'expression

$$T_{kin} = \frac{2}{Nk_B} \sum_{n=1}^N \frac{\overline{p_n^2}}{2m}, \quad (2.14)$$

suive la température imposée par les chocs aléatoires. Une fois la température exigée atteinte, les moyennes des quantités dynamiques ont été prises sur des intervalles de 10 ns. Le résultat de ce calcul, qui correspond à la courbe continue de la figure 2.3, montre l'accord parfait entre les deux méthodes et donc la validité de la méthode TI.

Le point crucial est que notre modèle possède une énergie caractéristique $\Delta H/2 = 0.22 \text{ eV}$ qui est du même ordre de grandeur que les énergies d'empilement introduites dans les modèles thermodynamiques, ces modèles phénoménologiques étant capables de reproduire les courbes expérimentales de façon très fidèle. Par exemple, selon le tableau I de la réf. [17], l'énergie relachée par le système lors de l'ajout d'une paire de bases supplémentaire à la séquence originale varie entre 0.347 eV et 0.465 eV en fonction de la nature des nouvelles bases appariées. Cela signifie que le modèle (2.11) possède suf-

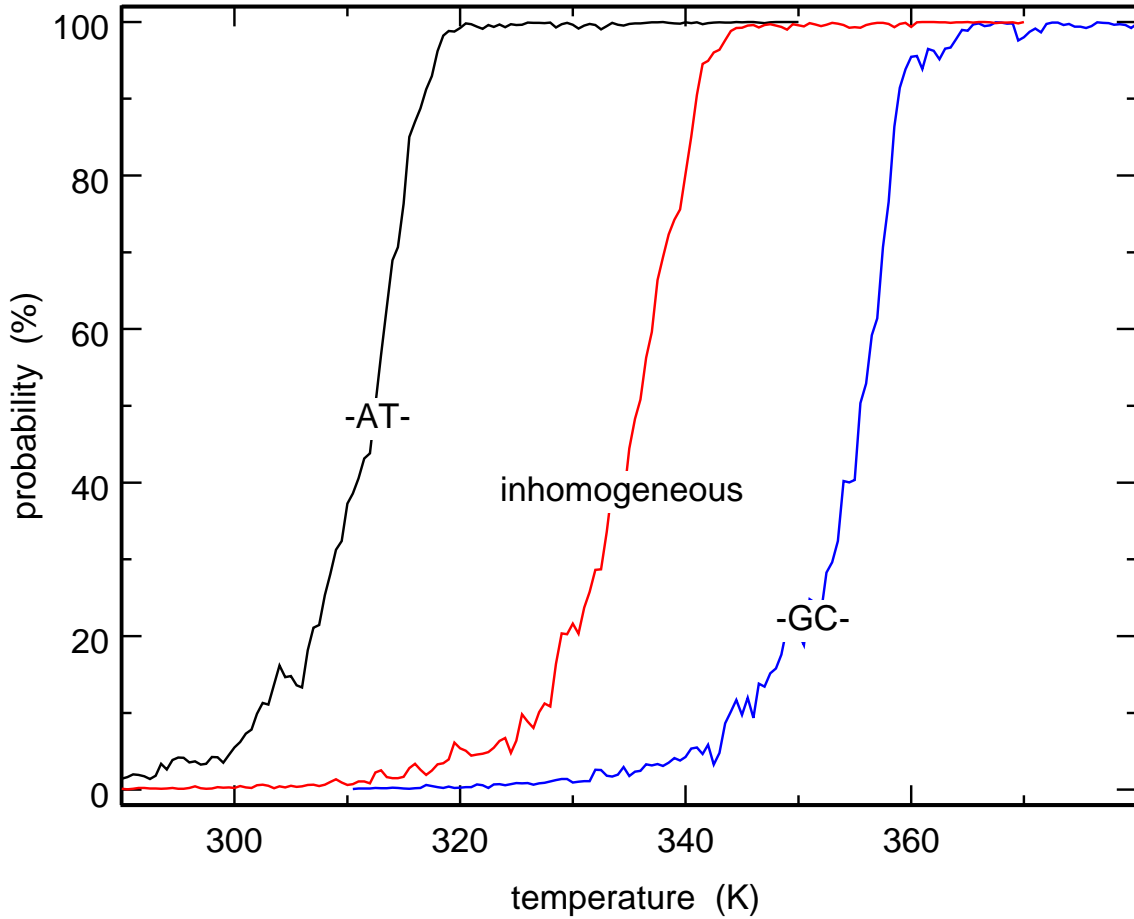


Fig. 2.4: Courbes de dénaturation moyennées sur 4 simulations MD pour le modèle JB de l'équation (2.15) avec trois séquences différentes de 2399 paires de bases, à savoir une séquence ATATAT...; une séquence hétérogène avec un pourcentage de GC de 61% et une séquence GCGCGC... . L'axe des ordonnées correspond à la probabilité $P(u_n - v_n > 15\text{\AA})$.

fisament de finesse microscopique pour que l'on puisse l'appliquer au cas des séquences hétérogènes, le hamiltonien prenant dans ce cas la forme

$$\begin{aligned}
 H = & \frac{m}{2} \sum_n (\dot{u}_n^2 + \dot{v}_n^2) + D \sum_n (1 - e^{-a(u_n - v_n)})^2 \\
 & + \frac{1}{2} \sum_n \Delta H^{(n)} (2 - e^{-b(u_n - u_{n-1})^2} - e^{-b(v_n - v_{n-1})^2}) \\
 & + K_b \sum_n [(u_n - u_{n-1})^2 + (v_n - v_{n-1})^2].
 \end{aligned} \tag{2.15}$$

Dans cette équation, les coordonnées u_n et v_n correspondent au déplacement des bases d'une même paire à partir de leur position d'équilibre et le paramètre $\Delta H^{(n)}$ est le coût

énergétique de l'extension de la séquence de la $(n - 1)$ ème à la n ème paire de bases. On peut considérer que le modèle (2.11), dont le paramètre ΔH est remplacé par $\Delta H^{(n)}$ dans l'équation (2.15), a été obtenu par la séparation approximative de cet Hamiltonien en fonction des degrés de liberté d'élongation des liaisons hydrogène $y_n = (u_n - v_n)/\sqrt{2}$ et du centre de gravité $x_n = (u_n + v_n)/\sqrt{2}$.

Dans la suite de ce manuscrit, j'appellerai également "modèle JB" le modèle de l'éq. (2.15), le contexte étant en général suffisant pour discriminer entre le modèle à N coordonnées de position de l'éq. (2.11) et celui à $2N$ coordonnées de position de l'éq. (2.15).

Les dix valeurs $\Delta H^{(n)}$ nécessaires pour modéliser une séquence d'ADN particulière ont été prises du tableau I de la réf. [17] et sont reproduites dans le tableau 2.1. Les autres paramètres du modèle sont $m = 300$ amu, $D = 0.04$ eV, $a = 3.5 \text{ \AA}^{-1}$, $b = 0.05 \text{ \AA}^{-2}$ et $K_b = 10^{-5} \text{ eV \AA}^{-2}$ (on rappelle que $1 \text{ eV} \approx 96.5 \text{ kJ/mol}$). La figure 2.4 montre les résultats des simulations de dynamique moléculaire (MD) (résolution numérique de l'équation (2.11), où y_n est remplacé par chacune des variables u_n et v_n) pour trois différentes séquences de 2399 paires de bases avec des bords libres, à savoir une séquence ATATAT \dots , une séquence GCGCGC \dots et finalement une séquence hétérogène piochée dans la littérature [35] qui contient 61% de paires de bases GC. Cette figure montre en fonction de la température la probabilité pour que la distance $u_n - v_n$ soit supérieure à 15 \AA (un seuil de 7 \AA donne un résultat très similaire). On constate que le modèle (2.15) reproduit très bien les caractéristiques des courbes de dissociation expérimentales, à savoir

- une dissociation discontinue qui se fait sur une région de température étroite,
- un écart de 40 K entre la température de dissociation d'une séquence d'ADN purement AT et celle d'une séquence purement GC,
- pour les séquences inhomogènes, une température de dissociation proportionnelle au pourcentage de paires GC.

Le modèle JB de l'équation (2.15) reproduit également le fait que pour les séquences inhomogènes la dissociation est un processus qui se fait en plusieurs étapes. Cet aspect

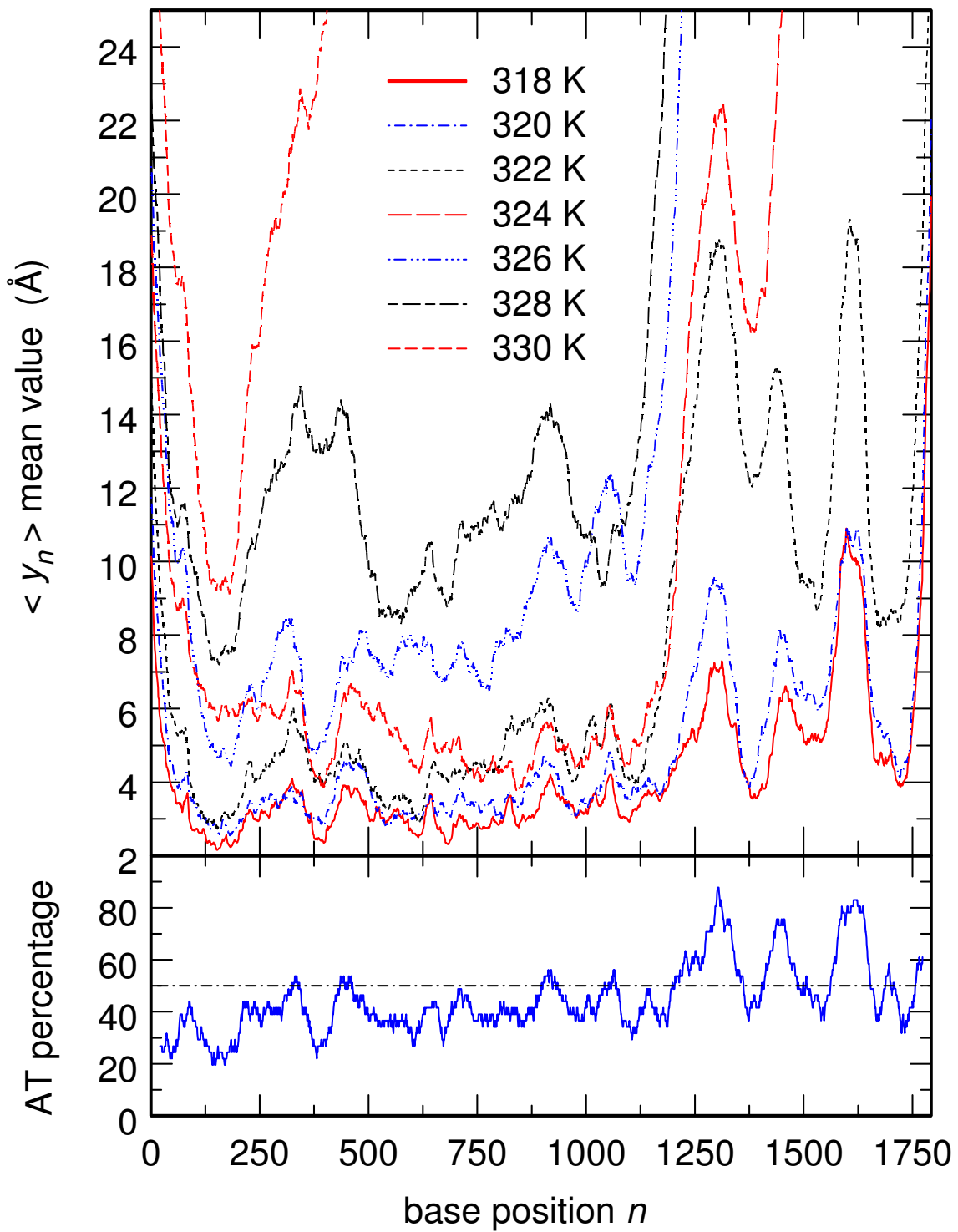
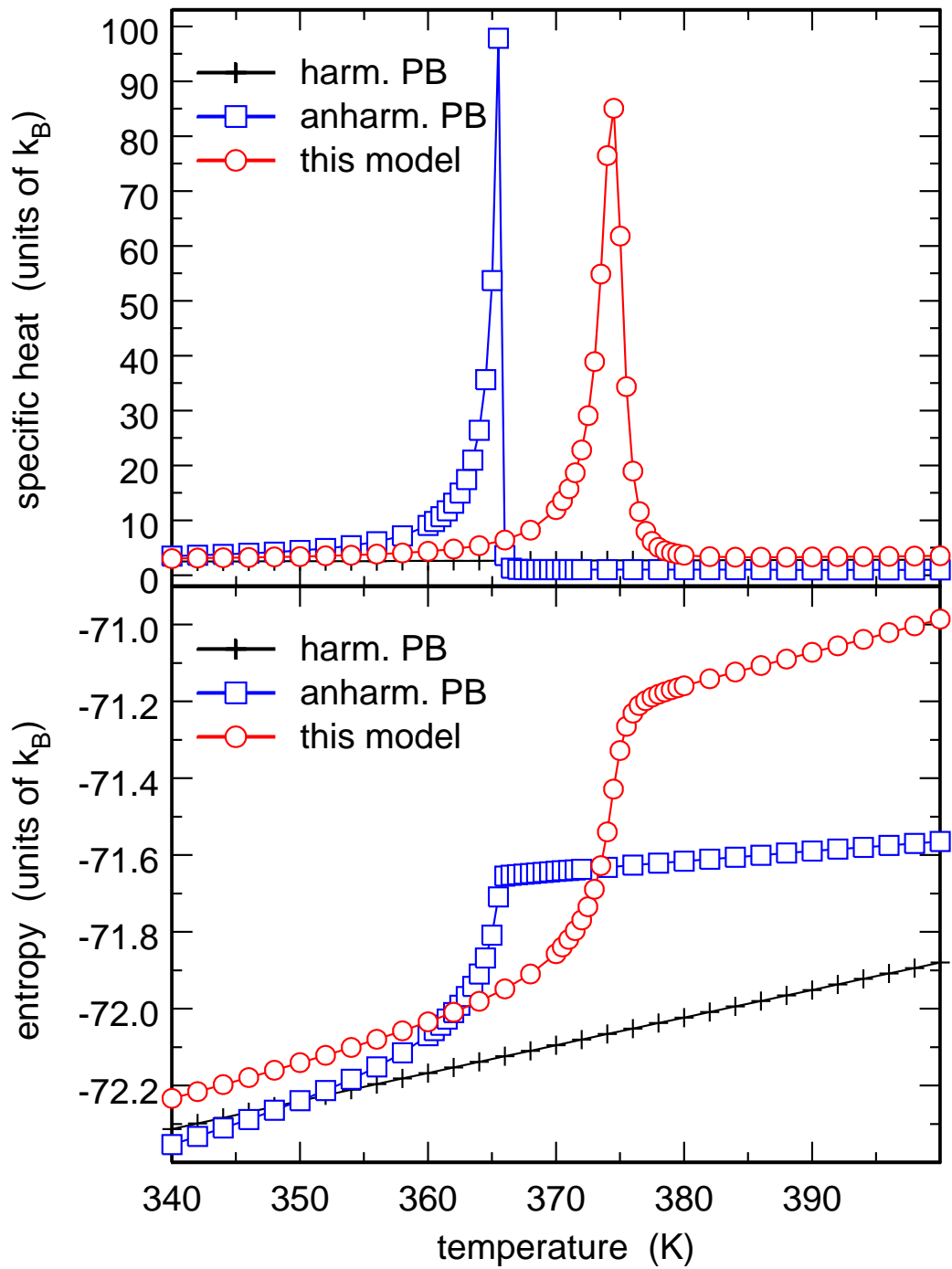


Fig. 2.5: En haut, profils de dénaturation $\langle y_n \rangle$ en fonction de la position de la paire de bases n à différentes températures pour une séquence hétérogène de 1793 paires de bases trouvée dans la littérature (NCBI, code d'entrée NM 001101). En bas, pourcentage de AT de la même séquence, moyenné sur 40 paires de bases consécutives.

Paires voisines	ΔH (eV)
A-T	0.347
T-A	0.361
A-A ou T-T	0.367
G-A ou T-C	0.396
C-A ou T-G	0.406
A-G ou C-T	0.410
A-C ou G-T	0.420
G-G ou C-C	0.449
C-G	0.465
G-C	0.465

Tab. 2.2: Energies d'empilement en fonction des paires de bases adjacentes répertoriées dans le sens $5' \rightarrow 3'$. On rappelle que $1 \text{ eV} \approx 96.5 \text{ kJ/mol}$.

qui est expérimentalement observé pour des séquences de longueurs 1000 – 10000 paires de bases peut être vérifié sur la figure 2.5. Cette figure montre pour une série de températures croissantes la moyenne $\langle y_n \rangle$ en fonction de la position de la paire de bases n dans une séquence hétérogène de 1793 paires de bases trouvée dans la littérature (NCBI, code d'entrée NM 001101). Le profil de la séquence est moyenné à chaque température sur une centaine d'ensembles de réalisations de 10 ns. Nous avons tracé au-dessous le pourcentage de AT moyenné sur 40 paires de bases successives. A grande échelle ce pourcentage augmente de $n = 1$ à $n = 1793$, avec un minimum au voisinage de $n = 150$ et trois maximums autour de $n = 1300$, $n = 1450$ et $n = 1600$. La corrélation entre ce pourcentage d'AT et le profil de la séquence est très nette. Une étude attentive du profil montre que la dénaturation de l'ADN commence par les régions centrées autour de ces trois maximums et que le double brin se dissocie pour tout $n \geq 1100$ à $T \simeq 326 \text{ K}$. Le milieu de la séquence ne se dissocie que 4 K plus haut, c'est à dire près de $T = 330 \text{ K}$. Finalement, la portion de la séquence qui se dissocie en dernier est positionnée autour du minimum $n = 150$ du pourcentage AT. Ces domaines de dissociation sont en parfait accord avec ceux obtenus par des modèles statistiques (voir la figure 2 de la référence [36]).



+

Fig. 2.6: Tracé de l'entropie (en bas) et la chaleur spécifique (en haut) par particule en fonction de la température, calculées par la méthode TI pour une séquence homogène à la limite thermodynamique. Les carrés correspondent au modèle DPB anharmonique, les croix au modèle PB harmonique et les cercles au modèle JB de l'équation (2.11).

Les figures 2.4 et 2.5 montrent sans ambiguïté que dans le cadre du modèle JB (2.15), l'hétérogénéité des valeurs de l'énergie d'empilement est suffisante pour expliquer la variation des températures de dissociation des différents domaines de la séquence en fonction du contenu génétique de celle-ci. Or les études précédentes qui visaient à modéliser la dynamique des séquences inhomogènes se contentaient de changer les paramètres du potentiel de Morse et négligeaient les variations des énergies d'empilement en fonction du type des paires de bases qui interagissent [15, 16]. Pour le moment, les paramètres du potentiel de Morse sont malheureusement déterminés de façon beaucoup moins précise que ceux de l'interaction d'empilement [15, 37]. Cependant il serait sans doute intéressant d'intégrer ce détail microscopique dans le modèle (2.11) ou (2.15) quand ces paramètres seront estimés de façon plus fiable.

Nous avons également calculé les quantités thermodynamiques extensives pour le modèle JB (2.11) par la méthode TI. La figure 2.6 montre la chaleur spécifique c_V et l'entropie par particule s pour les modèles PB harmonique et DPB anharmonique et le modèle de l'équation (2.11), dans le cas d'une séquence infinie. Le caractère discontinu de la transition de phase du modèle DPB anharmonique est révélé par la divergence de la chaleur spécifique au point critique et par la singularité de l'entropie. Bien que les courbes du modèle (2.11) soient légèrement différentes, on retrouve les mêmes aspects typiques d'une transition de phase du premier ordre, comme on peut le constater sur la figure 2.6. En augmentant la borne supérieure y_{max} de la base dans laquelle la matrice de transfert est définie, nous avons constaté que la région de température correspondant à la singularité de l'entropie s devient de plus en plus étroite et le maximum de c_V augmente linéairement avec y_{max} . Cela renforce notre conclusion selon laquelle la transition de l'ADN décrite par le modèle (2.11) correspond aussi à une transition de phase du premier ordre. L'exactitude de cette conclusion sera quantitativement prouvée dans le chapitre 4 où nous allons étudier la dissociation de l'ADN dans le cadre de la théorie des transitions de phase.

Nous avons encore une fois effectué des calculs MD afin de vérifier les résultats des

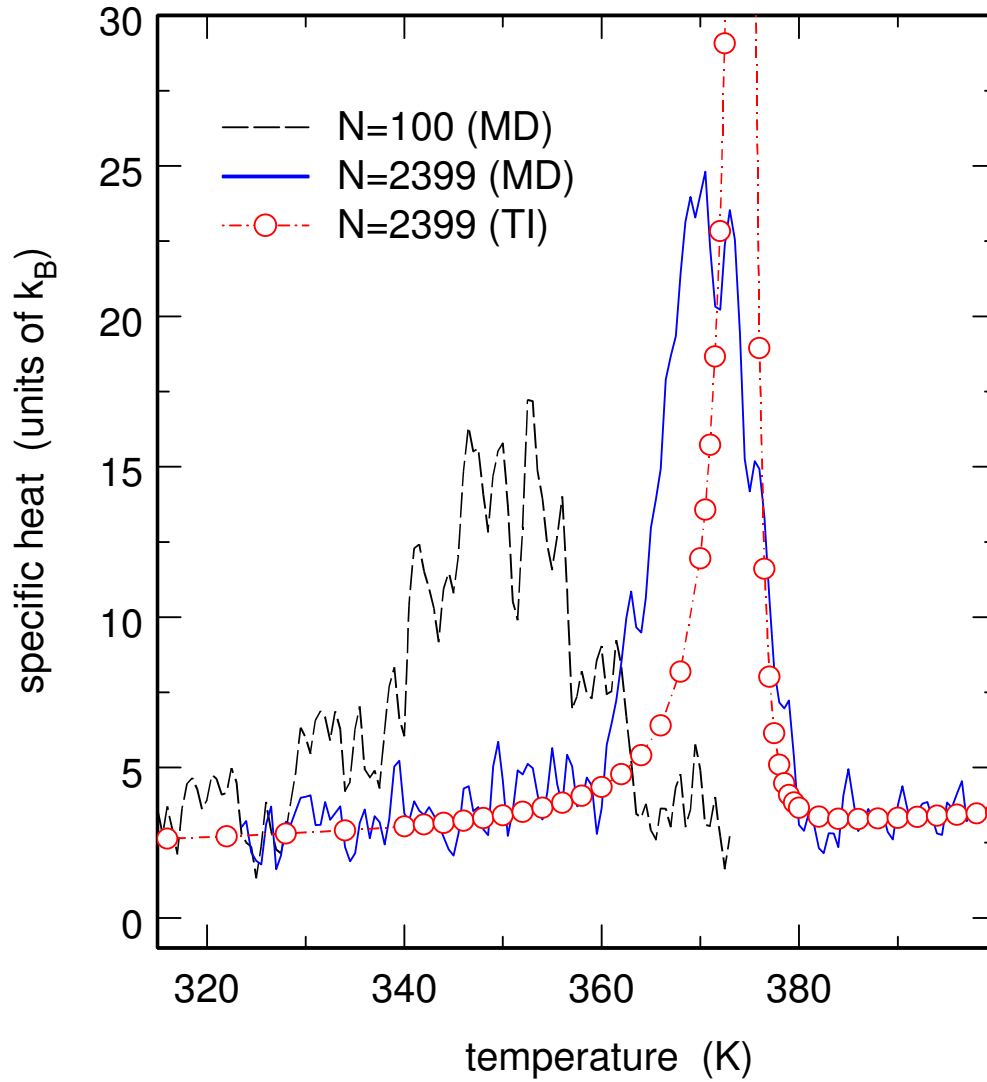


Fig. 2.7: Tracé de la chaleur spécifique correspondant au modèle JB de l'équation (2.11) calculée par la méthode MD pour une séquence de 100 paires de bases (la courbe en tirets, moyenne sur 80 simulations) et pour une séquence de 2399 paires de bases (la courbe continue, moyenne sur 4 simulations) comparée à celle calculée par la méthode TI pour le même modèle et avec une séquence de 2399 paires de bases (les cercles).

calculs TI. Lors d'une simulation MD, on peut en effet calculer c_V en utilisant la formule

$$c_V = \frac{1}{Nk_B} \frac{d\langle H \rangle}{dT}, \quad (2.16)$$

où $\langle H \rangle$ est l'énergie totale de la séquence moyennée sur le temps. c_V peut être aussi obtenue à partir de la variance de l'énergie du système,

$$c_V = \frac{1}{Nk_B T^2} (\langle H^2 \rangle - \langle H \rangle^2), \quad (2.17)$$

ce qui donne une estimation plus précise de la valeur à l'équilibre thermique. La courbe continue bleue de la figure 2.7 correspond aux résultats du calcul MD pour une séquence de 2399 paires de bases, alors que la courbe joignant les cercles a été obtenue par la méthode TI. Malgré la faible résolution du calcul MD due aux fluctuations de température, on constate qu'il y a un accord satisfaisant entre les deux méthodes.

A ce niveau, il convient de préciser que malgré les différentes expressions de $W(y_n, y_{n-1})$ utilisées dans les deux modèles d'ADN, la discontinuité de la transition de phase pour les modèles DPB et JB peut être expliquée par la même raison physique. Cette discontinuité est issue d'un effet entropique dû à une réduction de la rigidité de l'ADN quand on s'approche de la température de dissociation et est directement liée à l'anharmonicité du potentiel d'interaction.

Finalement, nous avons calculé par la méthode MD la chaleur spécifique par particule pour une séquence de 100 paires de bases avec des conditions aux bords ouvertes. Le résultat, qui correspond à la courbe en tirets de la figure 2.7, montre que le pic de c_V est déplacé de 20 K par rapport à la température de dissociation de la séquence avec 2399 paires de bases. Bien que cette différence soit trop élevée, le résultat est qualitativement en accord avec l'expérience.

Dans ce chapitre, nous avons donc présenté un nouveau modèle hamiltonien d'ADN. Ce modèle, qui intègre les énergies d'empilement finies et dépendantes de la séquence des modèles thermodynamiques, montre un bon accord avec l'expérience pour les séquences homogènes et hétérogènes. La dissociation du modèle présenté est associée à une tran-

sition de phase du premier ordre. Dans la suite de la présentation (chapitres 5 et 6), nous allons étudier les effets de taille et de désordre sur cette transition de phase afin d'obtenir plus d'informations sur le modèle et de comparer les résultats théoriques aux observations expérimentales. Ceci s'avèrera possible grâce au développement d'une méthode TI modifiée, d'application plus générale que la méthode présentée dans le chapitre 4 et plus précise que la méthode MD.

2.4 Vers des modèles plus réalistes

Dans la section 2.2, nous avons proposé un modèle hamiltonien d'ADN unidimensionnel capable de reproduire les aspects principaux des courbes de dénaturation expérimentales pour des séquences d'ADN homogènes et hétérogènes. Nous avons montré que l'hétérogénéité des séquences est correctement prise en compte par l'introduction d'énergies d'empilement dépendant de la nature des paires de base. Malgré ces résultats encourageants, on peut penser que le modèle de l'éq. (2.11) ou (2.15) est un peu simpliste. La limitation majeure de ce modèle vient du fait que chaque degré de liberté y_n représente un monomère entier, c'est à dire à la fois la base et le squelette sucre-phosphate, alors qu'il serait souhaitable que l'élongation de grande amplitude de la base et les faibles fluctuations du squelette soient traitées séparément. De plus, la dissociation de la séquence est décrite uniquement en fonction de cette coordonnée d'élongation, alors que la rotation des bases par rapport à l'axe de la double hélice joue vraisemblablement un rôle important lors de la séparation des deux brins. Le but de cette section est de proposer un modèle d'ADN pour des séquences homogènes et hétérogènes qui tient compte de ces caractéristiques supplémentaires [38]. A partir des calculs MD, je montrerai aussi que ce modèle plus sophistiqué est toujours capable de décrire correctement le processus de dénaturation thermique.

Le modèle amélioré est schématisé sur la Fig. 2.8. Le complexe sucre-phosphate de chaque monomère est représenté par une masse m et les complexes sont reliés entre eux par des "ressorts" harmoniques. Afin de réduire le nombre de degrés de liberté,

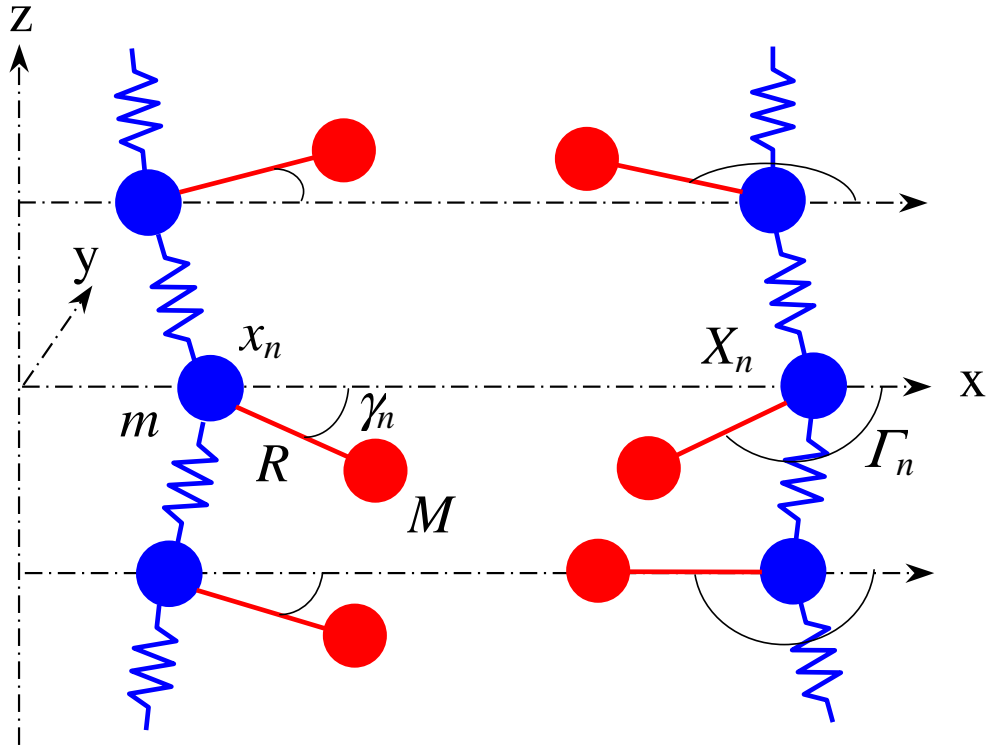


Fig. 2.8: Représentation schématique du modèle proposé à l'éq. (2.18)

les masses m ont été placées régulièrement le long de l'axe z et leur mouvement a été restreint à l'axe x qui relie les deux brins. Les coordonnées x_n et X_n indiquent la position respective de la $n^{\text{ème}}$ masse m de chaque brin. Les bases sont représentées par deux autres corps ponctuels de masse M , chacun d'eux étant rigidement lié à un groupe sucre-phosphate de masse m . La distance R reliant la base au groupe est fixée et la base est libre d'effectuer des rotations dans le plan (x, y) . Dans ce modèle la position de chaque base dépend donc de l'angle de rotation γ_n (resp. Γ_N) et de la coordonnée de translation x_n (resp. X_n). L'interaction des bases adjacentes se trouvant sur le même brin est décrite par un potentiel d'empilement semblable à celui du modèle présenté en section 2.2 et les bases d'une même paire interagissent via un potentiel de Morse. Si l'on se contente de ces termes, rien n'empêche le double brin de se plier par rapport à l'axe de l'hélice (même à très basse température) de sorte que les deux séries de masses M représentant le squelette sucre/phosphate soient quasiment superposées. Pour éviter cela, il est nécessaire d'ajouter dans le hamiltonien un terme d'interaction qui maintient les masses m et M de chacun des brins plus ou moins alignées jusqu'à la

dissociation. Ce terme a pour fonction de modéliser l'effet des deux (AT) ou trois (GC) liaisons hydrogène de chaque paire de bases qui, dans le système physique, assurent cette rigidité aux niveau des monomères et empêchent la chaîne de se plier. En notant $u_n = x_n + R \cos \gamma_n$, $v_n = R \sin \gamma_n$, $U_n = X_n + R \cos \Gamma_n$ et $V_n = R \sin \Gamma_n$ les coordonnées des masses M de chaque brin dans le plan (x, y) et par $d_n = \sqrt{(U_n - u_n)^2 + (V_n - v_n)^2}$ la distance entre les bases d'une même paire dans ce plan, le hamiltonien du modèle proposé s'écrit

$$\begin{aligned}
H &= T + V \\
T &= \frac{1}{2}(m + M) \sum_n (\dot{x}_n^2 + \dot{X}_n^2) + \frac{1}{2} \frac{mMR^2}{m + M} \sum_n (\dot{\gamma}_n^2 + \dot{\Gamma}_n^2) \\
V &= D \sum_n \{1 - \exp [a(d_o - d_n)]\}^2 \\
&\quad + K_b \sum_n (2 - \cos \gamma_n + \cos \Gamma_n) \exp [a(d_o - d_n)] \\
&\quad + \frac{1}{2} \sum_n \Delta H^{(n)} \{1 - \exp [-b(u_n - u_{n+1})^2 - b(v_n - v_{n+1})^2]\} \\
&\quad + \frac{1}{2} \sum_n \Delta H^{(n)} \{1 - \exp [-b(U_n - U_{n+1})^2 - b(V_n - V_{n+1})^2]\} \\
&\quad + K_s \sum_n [(x_n - x_{n+1})^2 + (X_n - X_{n+1})^2] .
\end{aligned} \tag{2.18}$$

J'ai choisi pour l'énergie cinétique une forme simplifiée afin de pouvoir intégrer l'équation de Langevin en utilisant un algorithme du second ordre, à savoir l'algorithme de Brunger-Brooks-Karplus déjà utilisé dans la section précédente pour le modèle unidimensionnel. Le premier terme de l'énergie potentiel est le potentiel de Morse modélisant les liaisons hydrogène qui relient les bases d'une même paire. Le second terme est l'interaction qui empêche les deux brins de se plier et le troisième et le quatrième terme représentent l'interaction d'empilement entre les bases adjacentes qui se trouvent sur le même brin. Le dernier terme de l'hamiltonien est l'énergie élastique du squelette sucre-phosphate.

Dans la suite de ce manuscrit, j'appellerai "modèle BSJ" le modèle de l'éq. (2.18).

Les valeurs numériques des énergies d'empilement $\Delta H^{(n)}$, qui varient entre 0.347 eV et 0.465 eV (on rappelle que $1 \text{ eV} \approx 96.5 \text{ kJ/mol}$), sont celles du modèle thermodynamique de la référence [17] qui sont reportées dans le tableau 2.1. Les autres paramètres du modèle présenté ont été obtenus en partie à partir des données chimiques connues : $m = 180 \text{ amu}$, $M = 130 \text{ amu}$, $d_o = 4 \text{ \AA}$, $R = 7 \text{ \AA}$ et $b = 0.05 \text{ \AA}^{-2}$. J'ai fixé les paramètres du potentiel de Morse, $D = 0.05 \text{ eV}$ et $a = 3.5 \text{ \AA}^{-1}$, en m'inspirant des modèles similaires [9, 33, 39]. Finalement les paramètres $K_b = 0.5 \text{ eV}$ et $K_s = 0.2 \text{ eV \AA}^{-2}$ ont été ajustés pour avoir un bon accord des courbes de dénaturation avec l'expérience. Pour estimer à quel point les résultats que je vais présenter sont sensibles aux variations de ces deux paramètres, j'ai étudié le déplacement de la température de dissociation en fonction de K_s et K_b . J'ai constaté qu'avec une valeur $K_b = 0.1 \text{ eV}$ (resp. $K_b = 1.0 \text{ eV}$), la température critique des séquences décroît (resp. augmente) de 40 K. J'en ai conclu que la valeur de K_b doit être prise entre 0.1 eV et 0.6 eV. D'une façon similaire, le choix $K_s = 0.1 \text{ eV \AA}^{-2}$ au lieu de $K_s = 0.2 \text{ eV \AA}^{-2}$ réduit la différence entre la température de dissociation d'une séquence GC et celle d'une séquence AT de 40 K, qui est la différence expérimentalement observée, jusqu'à 25 K. On peut en déduire qu'avec nos paramètres actuels du potentiel de Morse, la valeur K_s pour notre modèle est fixée de façon assez précise.

J'ai étudié la dynamique de la dissociation du modèle (2.18) en intégrant numériquement les équations de Langevin,

$$\mu \frac{d^2 q_n}{dt^2} = -\frac{\partial H}{\partial q_n} - \mu\gamma \frac{dq_n}{dt} + \sqrt{2\mu k_B T} w(t), \quad (2.19)$$

où μ est égal à $m+M$ pour une coordonnée d'élongation ($q_n = x_n$ ou X_n) et $mMR^2/(m+M)$ pour une coordonnée de pliage ($q_n = \gamma_n$ ou Γ_n). Les intégrations numériques ont été effectuées en utilisant l'algorithme de Brunger-Brooks-Karplus qui s'écrit pour le modèle présenté ici sous forme d'une relation de récurrence qui relie la valeur de chaque

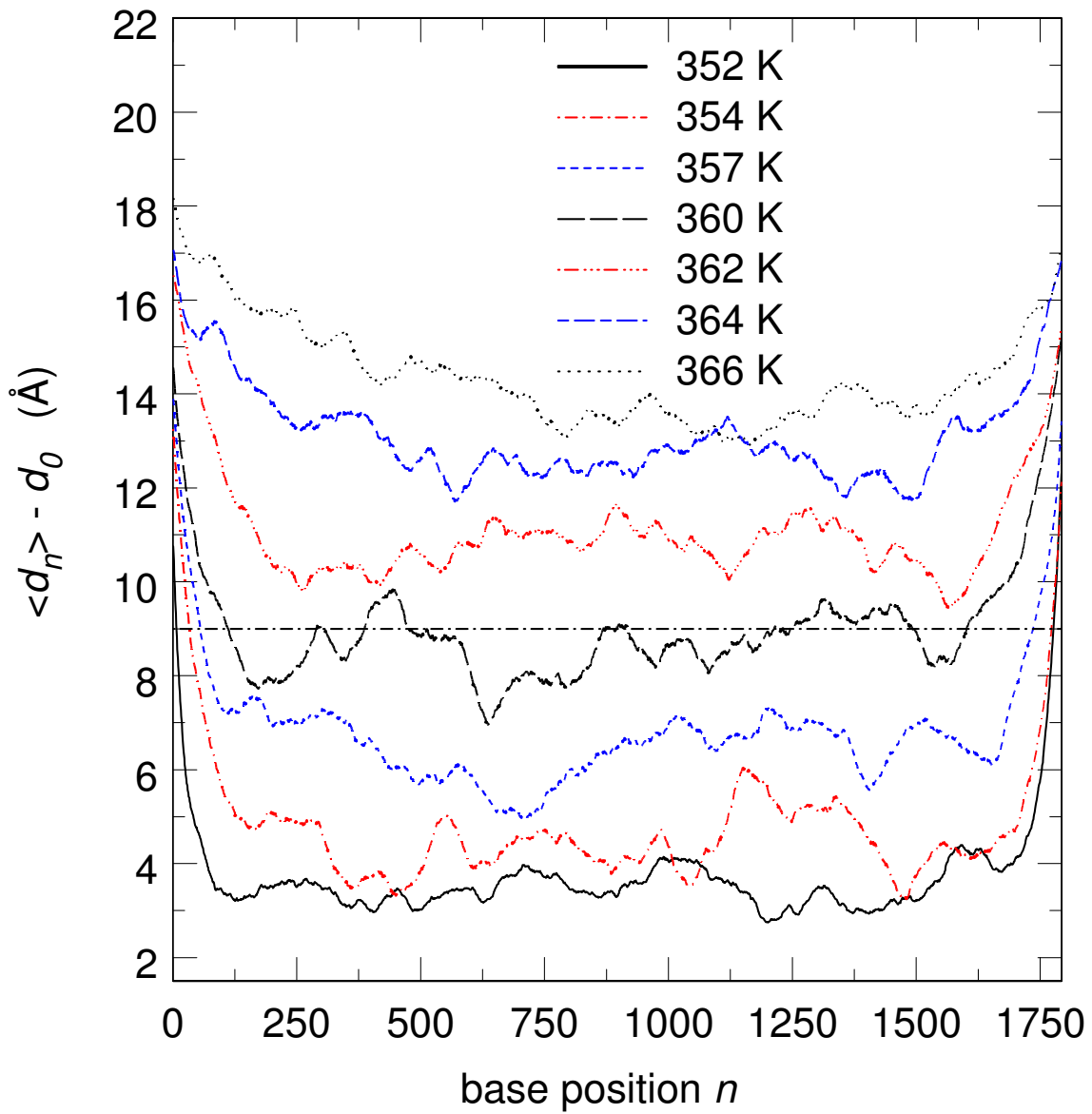


Fig. 2.9: Profils de dénaturation d'une séquence ATATAT... de 1793 paires de bases à différentes températures.

coordonnée q_n ($q_n = x_n, X_n, \gamma_n$ ou Γ_n) à des instants successifs $(k-1)\tau$, $k\tau$ et $(k+1)\tau$ selon l'expression

$$\begin{aligned} \frac{\mu}{\tau^2} (q_n^{k+1} - 2q_n^k + q_n^{k-1}) = & - \left. \frac{\partial V}{\partial q_n} \right|_{k\tau} - \frac{\mu\gamma}{2\tau} (q_n^{k+1} - q_n^{k-1}) \\ & + w(n, k) \sqrt{\frac{2\mu\gamma k_B T}{\tau}} \end{aligned} \quad (2.20)$$

où γ est le coefficient de dissipation et $w(n, k)$ est une fonction aléatoire gaussienne de valeur moyenne zéro et de variance unité. Le second et le troisième terme de l'équation (2.20) modélisent les effets du solvant sur la séquence d'ADN. L'intégration a été effectuée avec un pas de $\tau = 10$ fs, ce qui correspond à environ un centième de la période d'oscillation du potentiel de Morse à basse température. J'ai constaté que la réduction du pas τ d'un facteur dix ne changeait pas les résultats. La valeur du coefficient de dissipation considérée est $\gamma = 5 \text{ ns}^{-1}$.

Les profils de dissociation d'une séquence ATATAT... de 1793 paires de bases et ceux de la séquence hétérogène de même longueur introduite dans la section précédente sont représentés respectivement sur la figure 2.9 et 2.10 pour différentes températures. Chaque courbe représente la distance $d_n - d_o$ moyennée sur $0.15 \mu\text{s}$. On remarque que, pour une température donnée, la courbe de dissociation de la séquence AT est essentiellement plate sur toute la séquence, sauf vers les deux extrémités qui se dissocient à des températures légèrement plus basses que la partie interne de la chaîne.

Sur la figure 2.10, on constate que pour la séquence inhomogène le profil le long de la chaîne n'est en revanche pas plat mais reflète assez fidèlement le pourcentage AT de la séquence, qui figure au-dessous, comme c'était déjà le cas pour le modèle unidimensionnel présenté dans la section précédente. Autrement dit, on retrouve ici les caractéristiques de ce modèle. Plus précisément, lorsque l'on augmente la température, la dénaturation de la séquence commence par la dissociation des domaines se trouvant au voisinage immédiat des trois maximums d'AT à $n = 1300$, $n = 1450$ et $n = 1600$, alors qu'au dessus de 380 K , toutes les paires de bases avec $n \geq 1000$ sont dissociées. Puis il y a un intervalle de 3-4

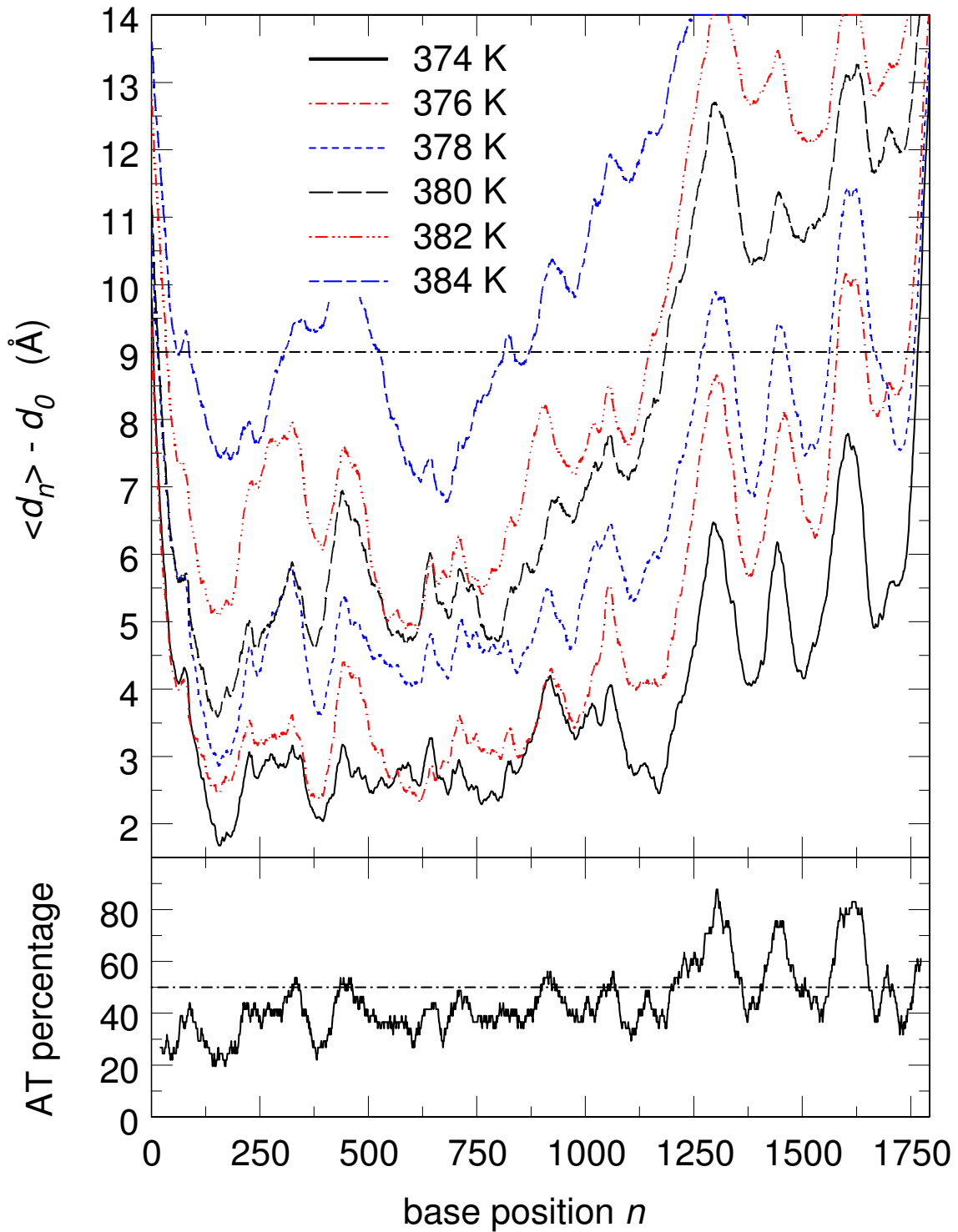


Fig. 2.10: En haut, profils de dénaturation d'une séquence hétérogène de 1793 paires de bases (NCBI, code d'entrée NM_001101). En bas, pourcentage de AT de la même séquence moyenné sur 40 paires de bases successives.

K à la fin duquel l'autre portion de la séquence se dissocie complètement. On peut noter que les domaines conduisant la dissociation de cette séquence hétérogène correspondent bien aux prédictions des modèles thermodynamiques (voir la figure 2 de la référence [36]).

Les expériences actuelles ne sont pas capables de fournir autant de détails que les figures 2.9 et 2.10. Cependant, les courbes d'extinction obtenues par la spectroscopie d'absorption à 260 ou 270 nm peuvent être directement comparées aux courbes de dénaturation théoriques qui donnent la fraction de paires de bases ouvertes en fonction de la température (voir la réf [17]). Les modèles thermodynamiques, qui sont des modèles à deux états, sont basés sur la définition des états ouverts et fermés. Cependant, pour les modèles hamiltoniens, qui ont un espace configurationnel continu, la limite qui sépare un état fermé d'un état ouvert est un peu vague. Après avoir fait des essais, j'ai conclu que le choix le plus convenable permettant de distinguer un état fermé d'un état ouvert consiste à considérer qu'une paire de bases n de la séquence est fermée si $\langle d_n - d_o \rangle \leq 9 \text{ \AA}$ tandis qu'elle est ouverte pour $\langle d_n - d_o \rangle > 9 \text{ \AA}$. Le seuil de 9 \AA correspond en fait au cas où les quatre masses du site n restent alignées tandis que la distance entre les deux brins augmente de 50%, autrement dit de 18 \AA (la valeur d'équilibre) à 27 \AA , ou de façon alternative au cas où la distance entre les deux brins est fixée à sa valeur d'équilibre de 18 \AA et les bases de la paire n font une rotation de 70° dans la même direction ou de 45° dans des directions opposées. Sur les figures 2.9-2.10, cette valeur de seuil est indiquée par une ligne horizontale. Les courbes de dénaturation obtenues par ce critère sont montrées sur la figure 2.11 pour les séquences de 1793 paires de bases des figures 2.9 et 2.10 (c'est à dire pour une séquence AT et une séquence hétérogène) et une séquence GC de la même longueur. Chaque point de la figure a été obtenu en intégrant $d_n - d_o$ pendant $0.15 \mu\text{s}$. Cette figure montre que le modèle (2.18) reproduit parfaitement les quatre aspects principaux des courbes de dénaturation expérimentales que j'ai déjà mentionnés dans la section précédente, c'est à dire

- la dénaturation des séquences homogènes se produit sur un intervalle de tempéra-

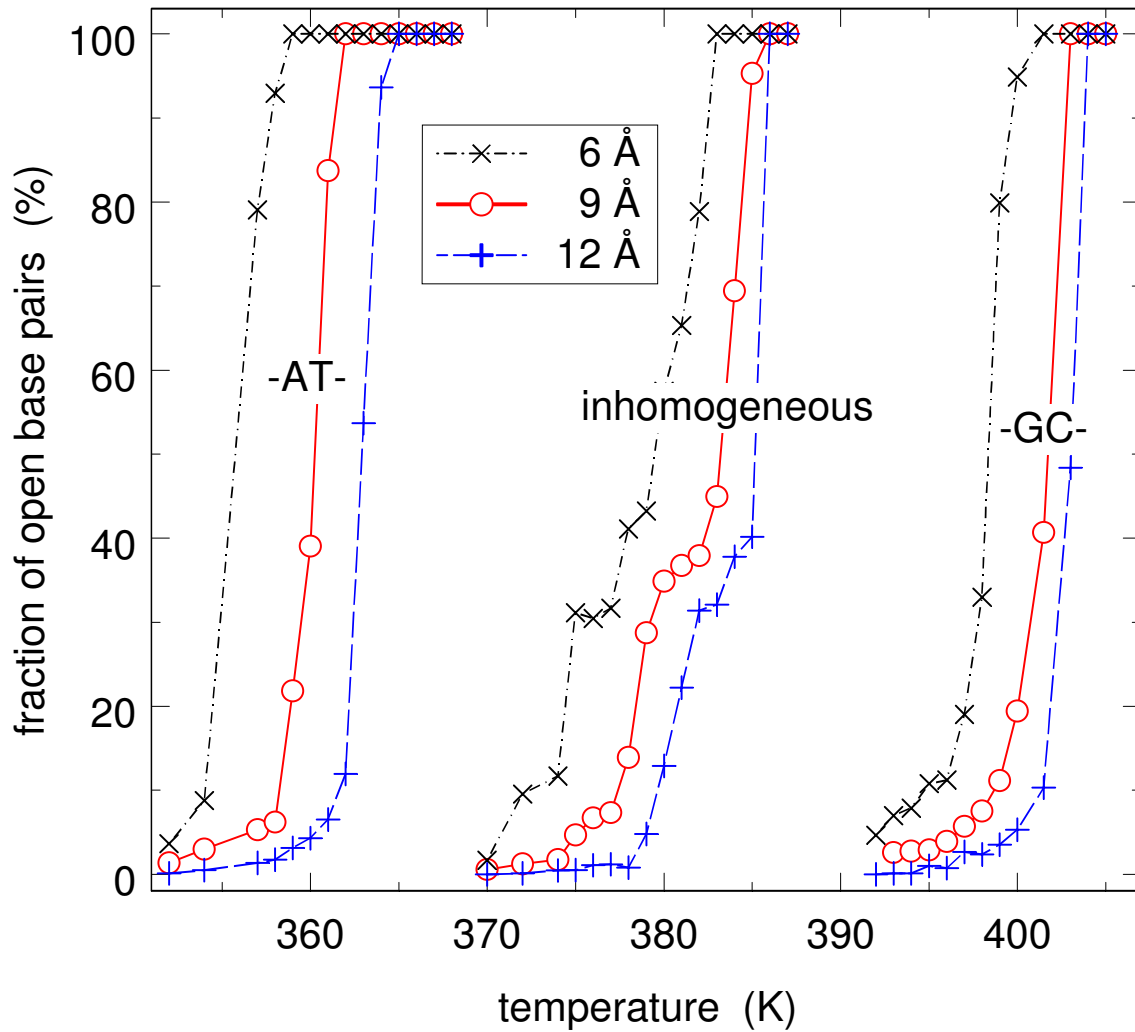


Fig. 2.11: Courbes de dénaturation pour deux séquences régulières ATATAT... et GCGCGC... de 1793 paires de bases et la séquence hétérogène de même longueur. La courbe est mesurée pour trois seuils d'ouverture différents, à savoir pour $\langle d_n - d_o \rangle > 6\text{\AA}$ (symboles X), $\langle d_n - d_o \rangle > 9\text{\AA}$ (cercles) et $\langle d_n - d_o \rangle > 12\text{\AA}$ (croix).

- tures très étroit, ce qui est le signe d'une transition de phase discontinue,
- les températures de dissociation de la séquence GC et de la séquence AT sont séparées de 40 K,
 - la température de dissociation d'une séquence hétérogène est proportionnelle au pourcentage de AT (ou de GC) de la séquence,
 - la dénaturation de la séquence hétérogène a lieu à travers une série de sauts discontinus, chaque saut correspondant à l'ouverture locale d'une portion conséquente de la séquence.

Afin de montrer que le choix de la valeur du seuil de dissociation n'est pas critique, j'ai aussi tracé les courbes de dénaturation avec un seuil de 6 Å et de 12 Å. On constate sur la figure 2.11 que la modification de cette valeur a pour effet de déplacer la température de dissociation de quelques Kelvins et de rendre les courbes plus lisses.

Avant de conclure cette section, je voudrais souligner le fait qu'il y a une légère différence entre la température où l'on considère que toutes les paires de bases sont ouvertes et celle qui correspond à la séparation définitive des deux brins. En fait dans le premier cas, la moyenne temporelle des $d_n - d_o$ doit être supérieure à 9 Å pour tout n , tandis que dans le deuxième cas, tous les $d_n - d_o$ doivent être au-dessus de ce seuil en même temps. Pour la séquence hétérogène et la séquence GCGCGC..., cet écart est de 1-2 K tandis que pour la séquence ATATAT..., il est de l'ordre de 5 K. Le dernier point tracé sur l'extrémité des courbes de dénaturation correspond précisément à la température où les deux brins se séparent pour la première fois. Il est également à noter que le critère choisi pour déterminer la fraction de paires de bases ouvertes sur la figure 2.11, à savoir $P(\langle d_n - d_o \rangle > 9\text{Å})$, est sensiblement différent de celui choisi en section 2.2 et particulièrement pour la figure 2.4, qui était simplement $P(u_n - v_n > 15\text{Å})$. Le fait de considérer les valeurs moyennes des écartements, au lieu des valeurs instantanées, semble conduire à des courbes de dénaturation plus proches des courbes expérimentales, dans la mesure où les ouvertures successives sont mieux marquées.

Dans cette section, j'ai proposé un modèle d'ADN plus élaboré que le modèle unidi-

mensionnel de la section précédente. Le nouveau modèle, qui fait intervenir les interactions d'empilement de façon similaire au modèle précédent, a la particularité de décrire séparément la dynamique du squelette sucre-phosphate et celle des bases qui sont capables de faire des rotations autour de l'axe des brins. Cette amélioration du modèle a porté à $4N$ le nombre de degrés de liberté pour une séquence ayant N paires de base, mais les propriétés principales des deux types de modèles restent très semblables et sont en bon accord avec les données expérimentales disponibles.

3. FLUCTUATIONS DE TEMPÉRATURE DE L'ADN AU VOISINAGE DE T_C

Dans un article récent [21], Nagapriya, Raychaudhuri et Chatterji ont exposé leurs résultats sur les mesures expérimentales des fluctuations de température de l'ADN près de la dissociation. Ils ont observé une forte amplification des fluctuations de basses fréquences (entre 0.01 et 1 Hz) à la température de dissociation et ont relié ce résultat à la coexistence des phases en équilibre qui se produit lors d'une transition de phase discontinue. Afin de vérifier cette observation expérimentale dans un contexte théorique, nous avons effectué des calculs de dynamique moléculaire. Bien que nos résultats aient été obtenus pour des fréquences nettement plus élevées (de 1 MHz à 10 GHz), nous avons bien observé le même accroissement des fluctuations. De plus, les calculs ont montré que le spectre de puissance des fluctuations de température suit une loi en $1/f$. Dans cette partie, nous allons exposer nos résultats et discuter l'origine possible des fluctuations en $1/f$.

Les trois modèles étudiés dans ce travail sont le modèle DPB (2.9), le modèle JB unidimensionnel de l'équation (2.11), qui prend la forme (2.15) dans le cas d'une séquence hétérogène, et le modèle BSJ représenté par l'équation (2.18). Les trois modèles en question ayant un hamiltonien séparable de la forme $H = E_c(\dot{\mathbf{q}}) + E_p(\mathbf{q})$, la température d'une séquence est simplement obtenue comme la moyenne sur la séquence de $2E_c(\dot{\mathbf{q}})/(nk_B)$, où n est le nombre de degrés de liberté du système. Deux situations différentes peuvent être envisagées dans les simulations. On peut, d'une part, étudier l'évolution des fluctuations de température des brins d'ADN posés dans un solvant, ce qui nécessite la prise en compte de l'interaction des particules du solvant avec la séquence. Cela s'effectue en intégrant numériquement l'équation de Langevin (2.12) ou (2.19) grâce à un intégrateur

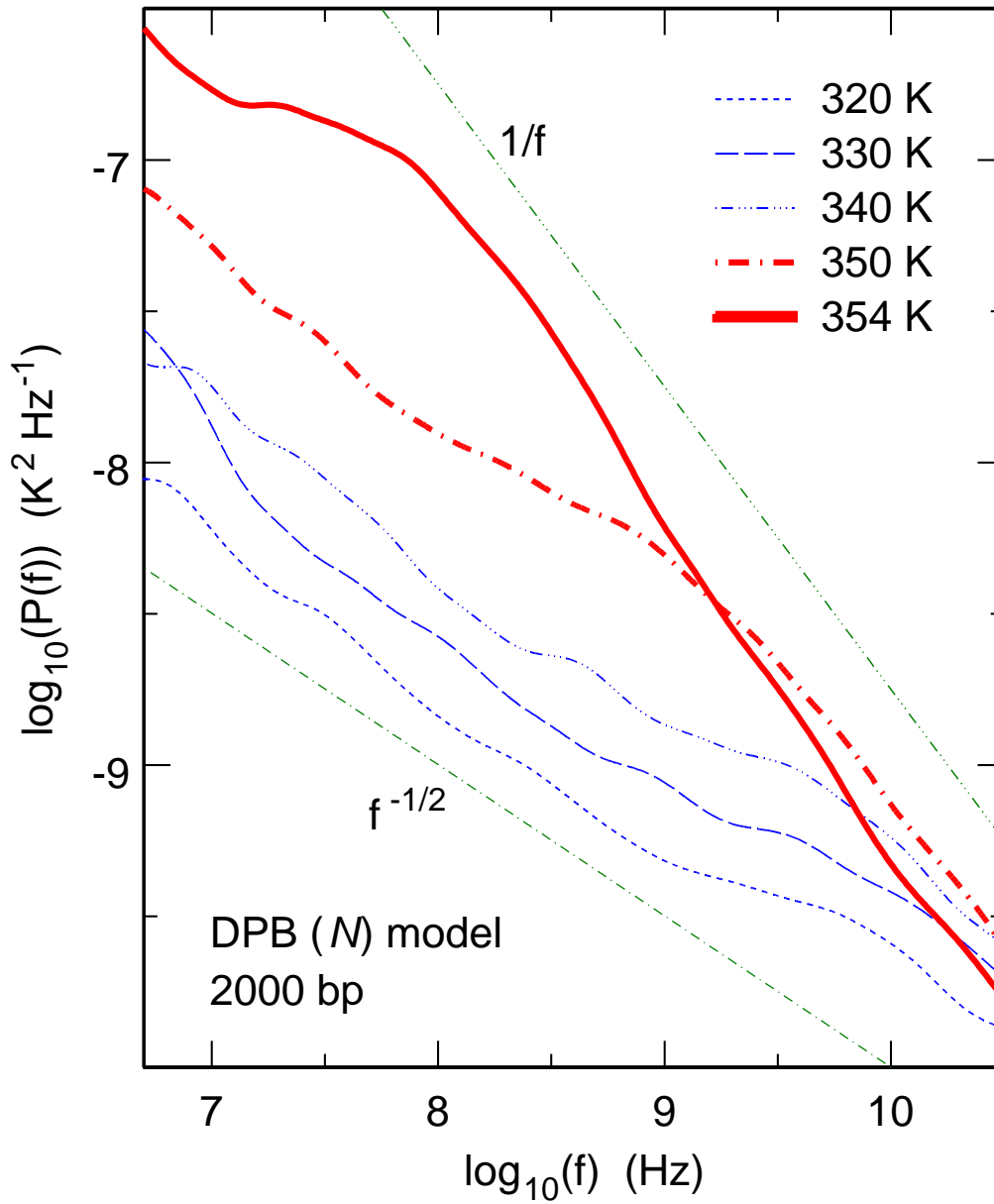


Fig. 3.1: Spectres de puissance des fluctuations de température tracés à l'échelle log-log pour une séquence homogène de 2000 paires de bases et le modèle DPB (2.9). La température de dénaturation de cette séquence est au voisinage de 354 K. Pour une séquence de N paires de bases, ce modèle possède $n = N$ coordonnées spatiales.

Brünger-Brooks-Karplus du second ordre (éqs (2.13) ou (2.20)). D'autre part, on peut considérer des brins d'ADN sans interaction avec le solvant. L'évolution de cet ensemble microcanonique est décrite par les équations de Hamilton, correspondant aux équations (2.12), (2.13), (2.19) et (2.20) amputées de leurs deux derniers termes. Le spectre de puissance est dans tous les cas obtenu par la formule

$$P(f) = \frac{2\pi}{t_{max}} \left| \int_0^{t_{max}} (T - \bar{T}) \exp(-2\pi i f t) dt \right|^2 \quad (3.1)$$

où T est la température instantanée de la séquence, t_{max} le temps d'intégration et \bar{T} la valeur moyenne de T calculée sur l'intervalle $[0, t_{max}]$. La valeur de t_{max} a été fixée à $0.5\mu s$ et la fonction $P(f)$ a été moyennée sur 10 à 20 conditions initiales différentes.

3.1 Amplification des fluctuations de température à la dénaturation thermique

Les figures 3.1, 3.2, 3.3 et 3.4 montrent les spectres de puissance obtenus par des calculs de dynamique hamiltonienne, respectivement pour une séquence homogène de 2000 paires de bases avec le modèle DPB, une séquence régulière ATATAT... de 2399 paires de bases et le modèle JB, une séquence hétérogène de 2399 paires de bases correspondant à l'inhibiteur de l'activateur du facteur d'amplification de l'hépatocyte [35] toujours avec le modèle JB, et finalement une autre séquence de 1793 paires de bases (cDNA de la β -actine humaine, code d'entrée NCB NM_001101) pour le modèle BSJ. Les simulations ont été effectuées pour des chaînes d'ADN avec des conditions aux bords libres. Les températures de dissociation pour ces quatre systèmes, c'est à dire la température à laquelle 50% des paires de bases sont ouvertes, sont respectivement 354 K, 310 K, 335 K et 385 K. Ces températures ont été déterminées à partir de courbes de dénaturation similaires à celles de la figure 2.4.

Bien que les fréquences correspondantes soient beaucoup plus élevées, ces courbes possèdent les mêmes caractéristiques que les courbes expérimentales de la référence [21] :

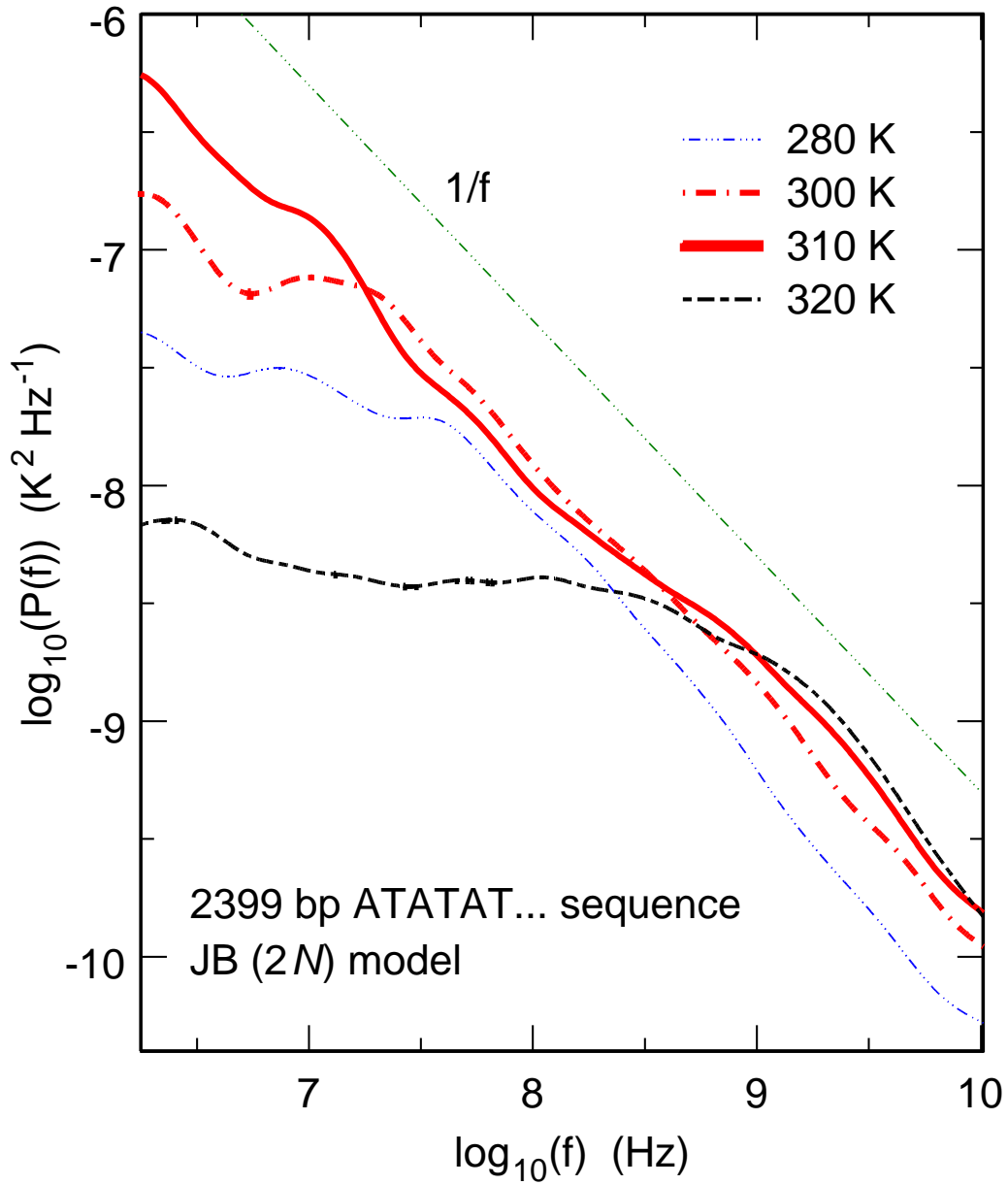


Fig. 3.2: Spectres de puissance des fluctuations de température tracés à l'échelle log-log pour le modèle JB (2.15) et une séquence ATATAT... de 2399 paires de bases. La température de dissociation de cette séquence est au voisinage de 310 K. Pour une séquence de N paires de bases, ce modèle possède $n = 2N$ coordonnées spatiales.

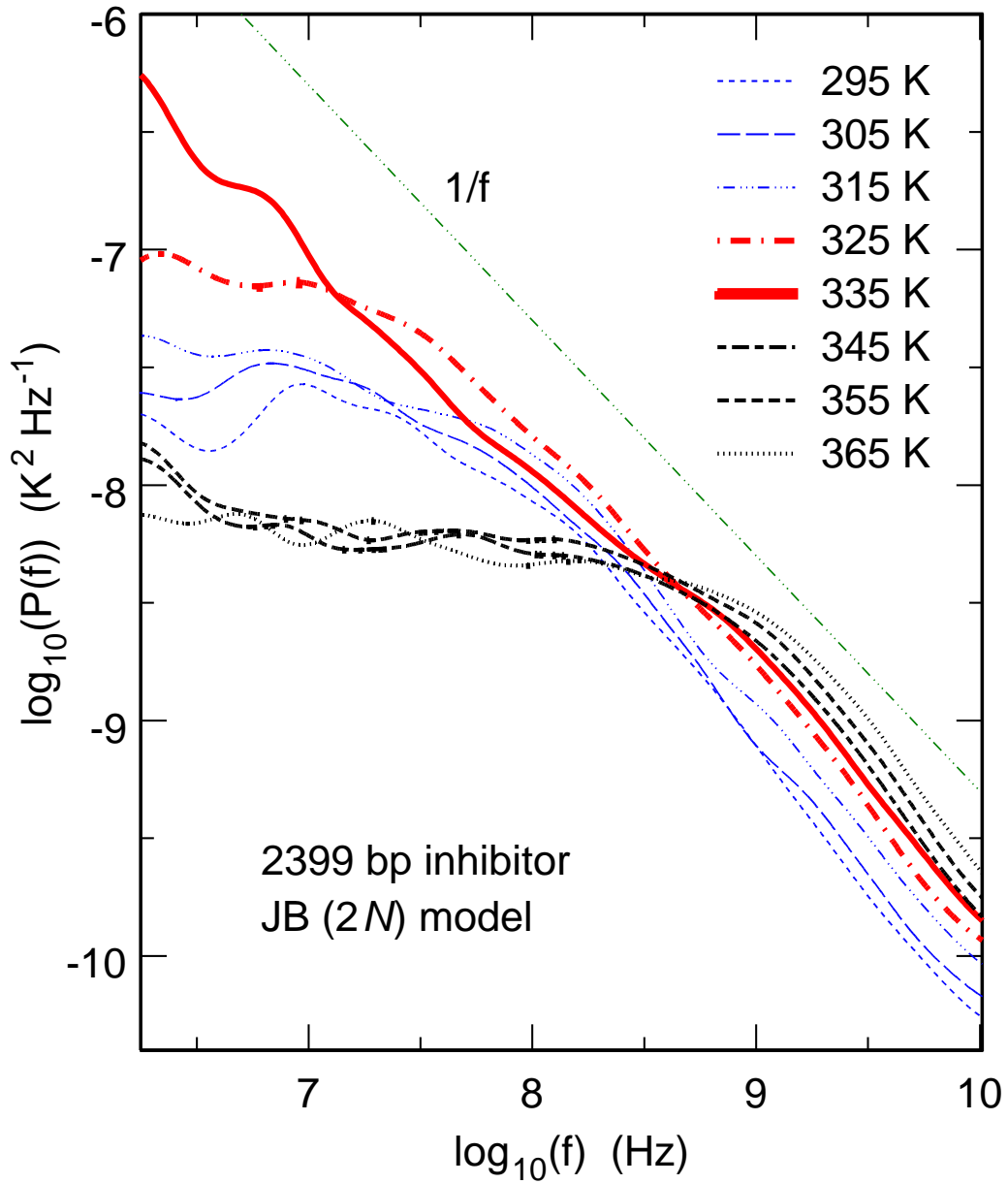


Fig. 3.3: Spectres de puissance des fluctuations de température pour le modèle JB (2.15) et la séquence inhibiteur [35] de 2399 paires de bases. La température de dissociation de cette séquence est au voisinage de 335 K. Pour une séquence de N paires de bases, ce modèle possède $n = 2N$ coordonnées spatiales.

(i) les fluctuations varient lentement avec la température en dessous de la dénaturation, (ii) elles augmentent de façon spectaculaire à la température de dissociation, particulièrement dans la région basse fréquence du spectre, (iii) elles redeviennent plus faibles dès que les deux brins se séparent. On peut noter que la comparaison des figures 3.1-3.4 montre que ce comportement ne dépend ni du modèle, ni de la séquence concernée, puisque ces résultats ont été obtenus pour trois modèles différents. Par conséquent, on peut en déduire que l'amplification des fluctuations n'est pas une propriété du modèle étudié mais qu'il s'agit vraisemblablement bien d'une caractéristique de la dénaturation thermique de l'ADN.

Comme je l'ai précisé, toutes les séquences et les modèles étudiés possèdent les caractéristiques observées sur les figures 3.1-3.4. Cela inclut aussi bien les séquences hétérogènes que homogènes, alors que Nagapriya et al ont remarqué que l'accroissement des fluctuations est beaucoup plus faible dans le cas des séquences homogènes [21]. Leur conclusion est toutefois basée sur l'observation de la dynamique d'une seule séquence régulière ATAT... de 100 paires de bases tandis que les chaînes d'ADN étudiées dans ce travail contiennent plus de 1500 paires de bases, ce choix ayant été fait afin de contourner la difficulté numérique d'imposer une température à une chaîne courte proche de la dénaturation. Il est important de noter que la dynamique des petites séquences est très sensible aux ouvertures des extrémités et peut pour cette raison différer de façon importante de la dynamique des longues séquences. Il serait donc intéressant de vérifier si des mesures expérimentales effectuées sur des séquences régulières plus longues confirment ou non les observations de Nagapriya et al.

Un autre résultat remarquable obtenu pour toutes les séquences et les trois modèles étudiés est le fait que le spectre de puissance obéit à une loi en $1/f$ sur un intervalle de fréquences très étendu. Je reviendrai sur ce point dans la section 3.4.

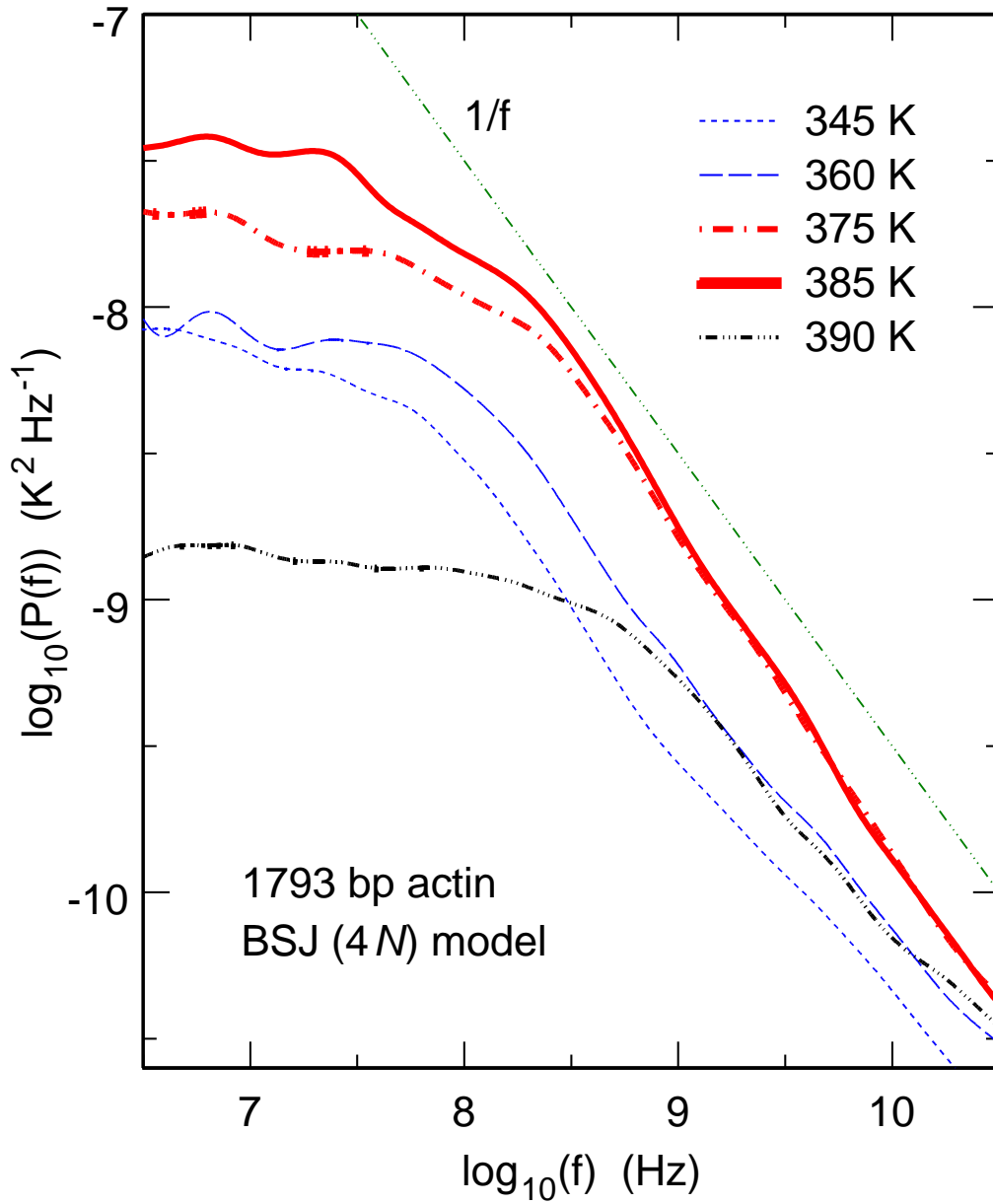


Fig. 3.4: Spectres de puissance des fluctuations de température pour le modèle BSJ (2.18) et la séquence actine (NCBI, code d'entrée NM_001101) de 1793 paires de bases. La température de dissociation de cette séquence est de 385 K. Pour une séquence de N paires de bases, ce modèle possède $n = 4N$ coordonnées spatiales.

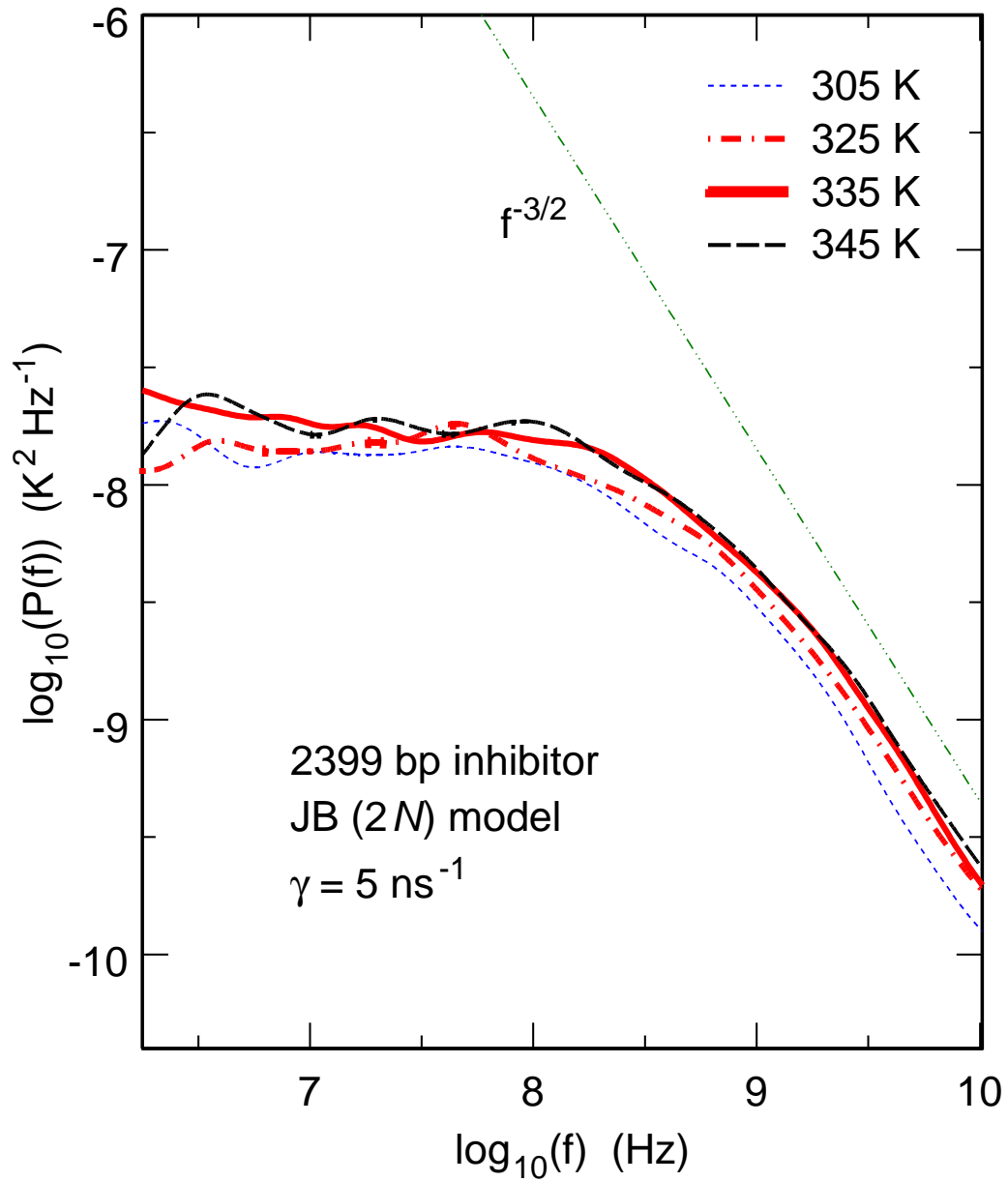


Fig. 3.5: Spectres de puissance pour le même système que la figure 3.3, mais calculés en intégrant les équations de Langevin au lieu des équations de Hamilton. La valeur du coefficient de dissipation choisie est $\gamma = 5 \text{ ns}^{-1}$.

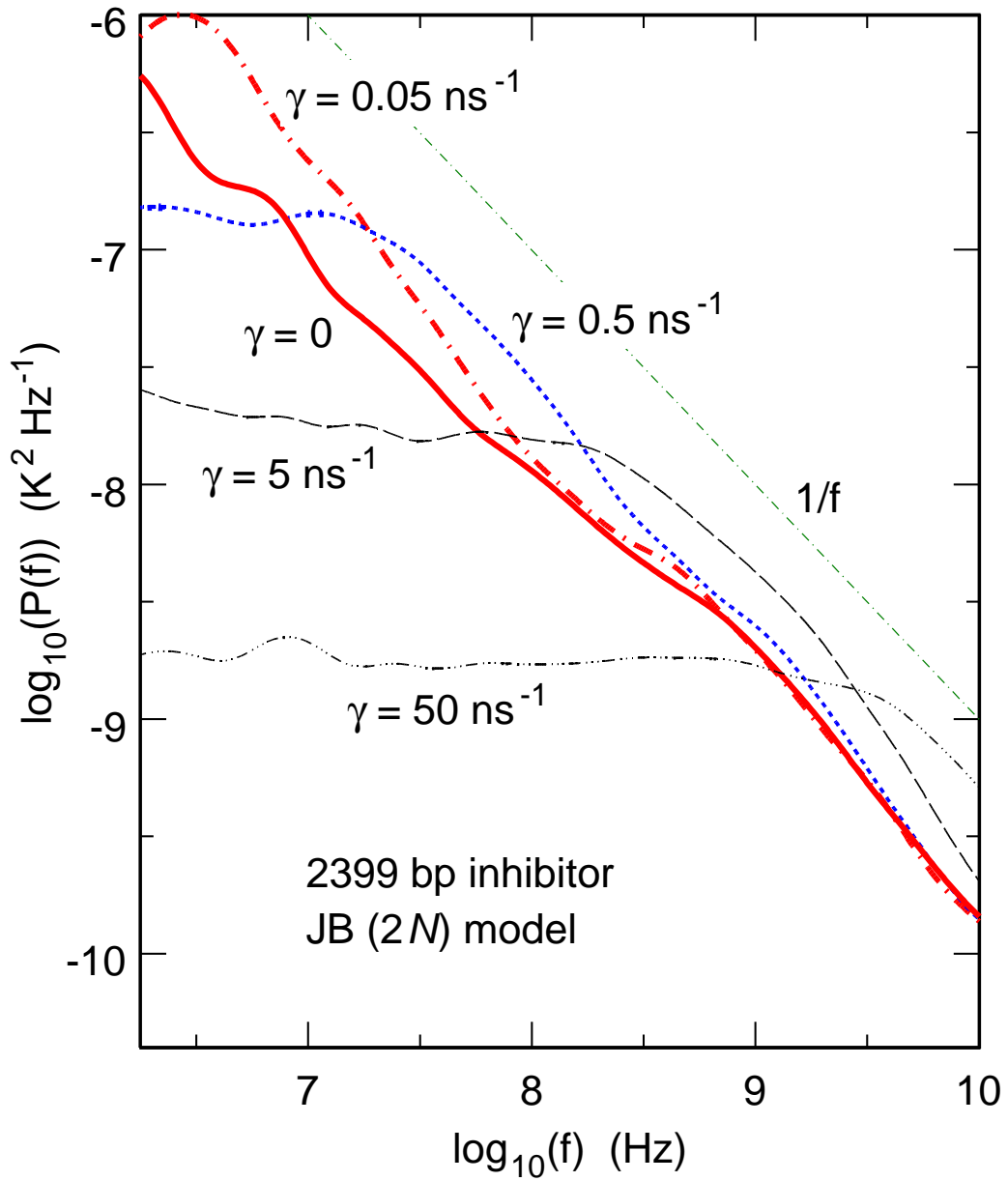


Fig. 3.6: Spectres de puissance des fluctuations de température calculés à $T=335 \text{ K}$ (la température de dissociation) pour la séquence inhibiteur de 2399 paires de bases et le modèle JB (2.15). Les courbes correspondent à différentes valeurs du coefficient de dissipation γ .

3.2 Effets du coefficient de dissipation γ sur les fluctuations de température

Comme je l'avais indiqué, les résultats présentés dans la section précédente ont été obtenus en intégrant les équations de Hamilton. Autrement dit, une fois la séquence chauffée jusqu'à la température désirée, γ a été fixé à zéro dans les équations (2.12), (2.13), (2.19) et (2.20). Le fait d'intégrer les équations de Langevin au lieu des équations de Hamilton peut donner des résultats nettement différents. Ceci est dû au fait que, à la limite thermodynamique correspondant aux chaînes de longueur infinie, les moyennes statistiques des observables dans l'ensemble micro-canonique (dynamique Hamiltonienne) et dans l'ensemble canonique (dynamique de Langevin) sont en principe les mêmes, mais leurs fluctuations dans ces deux ensembles diffèrent [41]. Par exemple, la figure 3.5 montre les spectres de puissance pour la séquence inhibiteur de 2399 paires de bases obtenus en intégrant les équations de Langevin du modèle JB pour différentes températures. Lors de cette simulation, nous avons pris une valeur $\gamma = 5 \text{ ns}^{-1}$ qui est dans l'intervalle de valeurs généralement utilisées. Contrairement aux courbes de dynamique Hamiltonienne de la figure 3.3, on constate que les courbes de la figure 3.5 ne montrent aucune variation significative à la température de dissociation. Afin d'étudier plus en détail l'influence du coefficient de dissipation, nous avons tracé en figure 3.6 les spectres de puissance pour différentes valeurs de γ à la température critique. On remarque que jusqu'à 0.05 ns^{-1} les fluctuations de température sont très faiblement influencées par la variation de γ pour des fréquences au-dessus du MHz. Cependant, les coefficients de dissipation $\gamma = 0.5 \text{ ns}^{-1}$ et $\gamma = 5 \text{ ns}^{-1}$ atténuent les oscillations de fréquences inférieures, respectivement, à 10 MHz et 100 MHz. Finalement, les spectres de puissance obtenus avec une valeur $\gamma = 50 \text{ ns}^{-1}$ diffèrent de ceux obtenus avec des valeurs de γ plus petites sur tout l'intervalle entre 10^6 Hz et 10^{10} Hz .

On en conclut que les collisions aléatoires avec les molécules du solvant peuvent faire disparaître l'amplification des fluctuations de température qui se produit dans le cas des brins d'ADN isolés. A ce sujet, on peut noter que les mesures expérimentales

de la référence [21] ont été effectuées après l'évaporation du solvant, afin d'éviter la réduction de la sensibilité expérimentale due à la masse thermique du solvant. Il est évident qu'après évaporation du solvant les brins d'ADN sont toujours en contact avec le résidu de solvant et interagissent également entre eux, mais ils sont beaucoup plus isolés que dans le solvant, ce qui peut conduire à un coefficient de dissipation effectif plus faible. Afin d'élucider cette question, il serait intéressant de répéter les mesures pour différents solvants susceptibles d'interagir plus ou moins fortement avec l'ADN.

3.3 *Fluctuations de température et bulles*

Nous savons que les bulles de grande amplitude, qui sont des ouvertures locales, se développent précisément dans l'intervalle de température où l'amplification des fluctuations devient significative. Il est donc naturel de se demander s'il existe un lien entre l'amplification des fluctuations de température et la dynamique des bulles. Plus précisément, la question qui se pose est de savoir si les fluctuations de température sont plus importantes dans les bulles que dans les régions liées de la séquence. Si c'était le cas, la formation des bulles serait directement responsable de l'amplification des fluctuations à la température critique. Afin de vérifier cette hypothèse, nous avons analysé les fluctuations de température de la séquence inhibiteur de 2399 paires de bases [35] aussi bien sur la séquence entière que sur de petites portions de la séquence. Le découpage en 11 portions distinctes est indiqué en figure 3.7, qui montre les courbes de dénaturation de la séquence en question à 325 K et 335 K (température de dénaturation). Les écarts de grande amplitude entre les paires de bases correspondent aux régions où les bulles sont formées préférentiellement. Ce sont également les régions riches en AT (voir la partie inférieure de la figure 3.7). Nous avons fragmenté cette séquence en 11 portions successives indiquées par les lettres A à K. Certaines portions (A, D, F, H et K) se dissocient à 325 K tandis que d'autres (B, E et I) sont toujours liées à cette température. A la température de dénaturation (335 K), toutes les portions à l'exception de B sont dissociées. Nous avons ensuite calculé le spectre de puissance pour chacune de ces portions. La figure 3.8

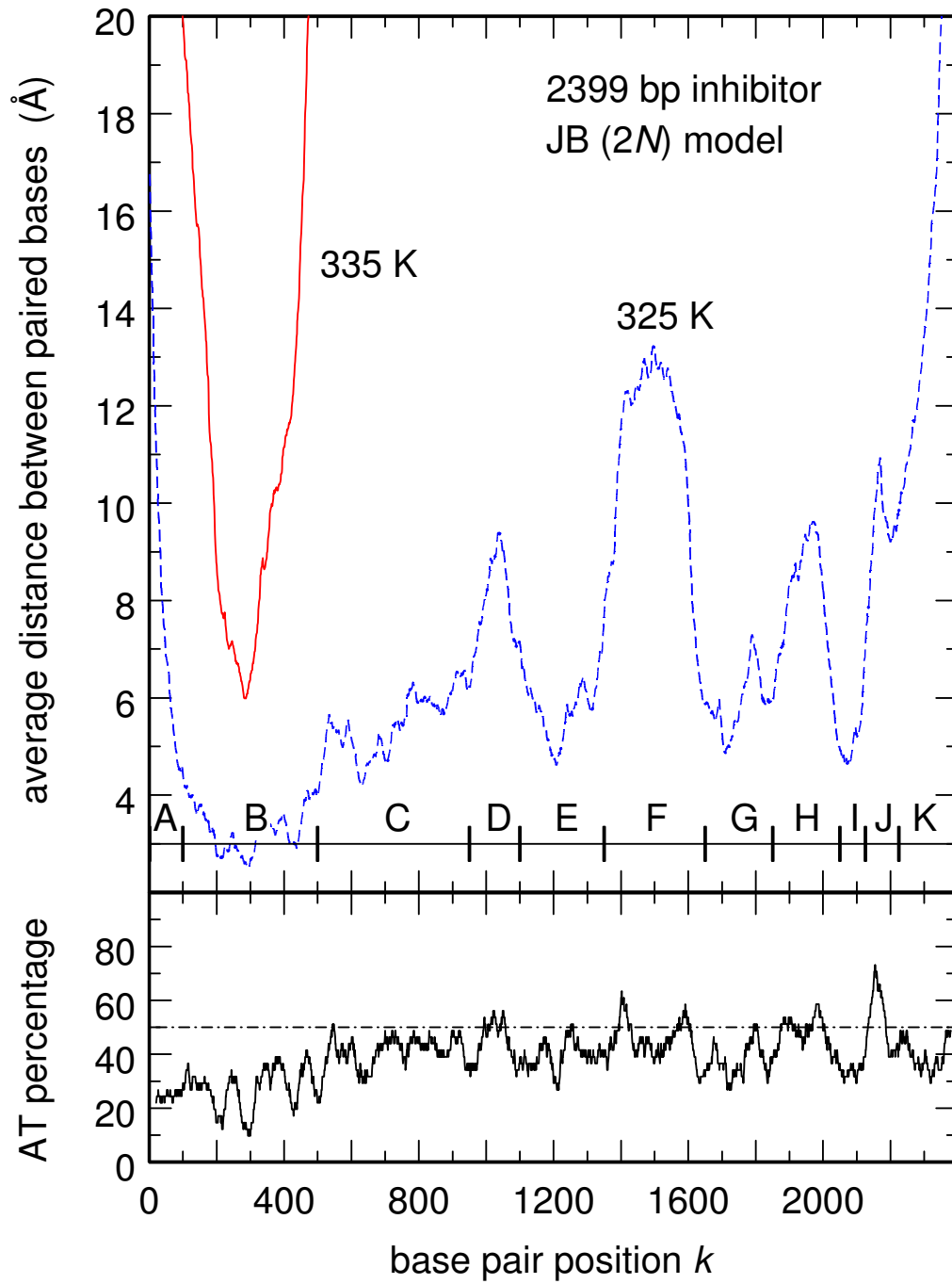


Fig. 3.7: En haut, courbes de dénaturation de la séquence inhibiteur de 2399 paires de bases pour le modèle JB (2.15) à 325 K et 335 K. En bas, pourcentage de AT de cette même séquence moyenné sur quarante paires de bases consécutives.

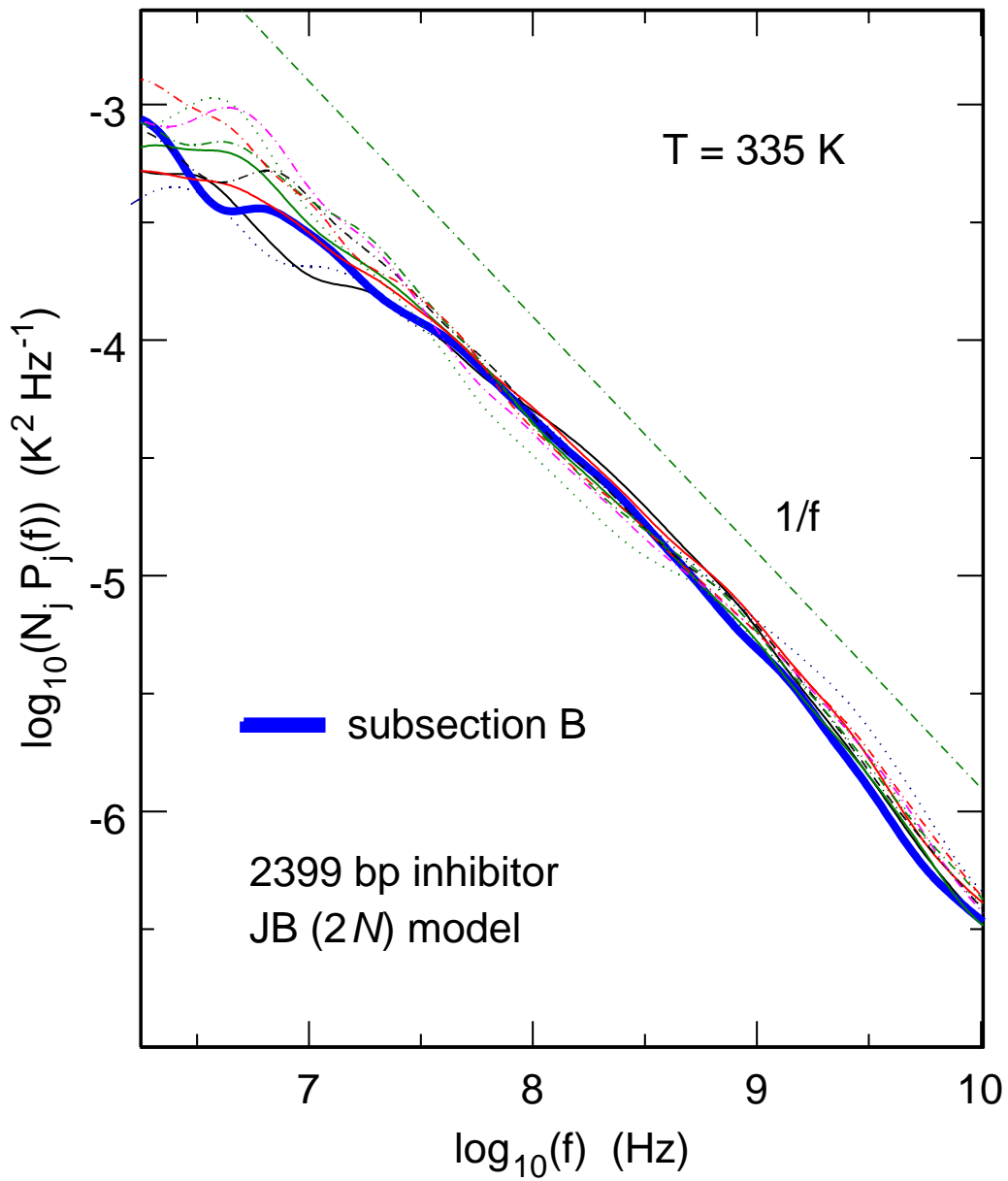


Fig. 3.8: Tracé du spectre de puissance multiplié par la longueur N_j des portions considérées, calculé à la température de dénaturation (335 K) et pour le modèle JB. Les différentes courbes correspondent aux 11 portions de la séquence inhibiteur de 2399 paires de bases.

montre les spectres de puissance à 335 K, correspondant à chaque portion, multipliés par la longueur N_j de la portion considérée. On constate que ces courbes normalisées correspondant aux différentes portions se superposent et obéissent à la loi en $1/f$, ce qui est également le cas à 325 K. Cela signifie que pour les deux températures, (i) le spectre de puissance évolue à une fréquence donnée suivant la loi en $1/N_j$, (ii) à part cette loi d'échelle, le spectre de puissance est le même pour toutes les portions et pour la séquence entière. La dernière remarque est assez évidente si l'on compare la figure 3.8 avec la courbe de la figure 3.3 correspondant à la séquence entière à 335 K, ou la courbe de la portion B avec celle des portions ouvertes.

A partir de cette analyse, on conclut donc qu'au dessus de l'échelle du MHz l'amplification des fluctuations de température à la dénaturation est uniforme le long de la séquence et non pas localisée à l'intérieur des bulles. A ce stade de la discussion, il est important de rappeler que les bulles et les portions liées d'une séquence constituent les deux phases qui coexistent à la température critique. Afin d'interpréter leurs résultats, Nagapriya et al ont proposé l'hypothèse selon laquelle l'amplification des fluctuations de température serait due à la coexistence des phases en équilibre. Il se pourrait que les "phases coexistantes" en question soient reliées à la structure tertiaire (où à des conformations d'ordre plus élevé) de l'ADN. Cependant, si ces "phases" se réfèrent effectivement à la structure secondaire, c'est à dire à la dynamique des bulles, alors les résultats que l'on vient de présenter montrent que l'amplification des fluctuations de température à la dissociation est due à un mécanisme beaucoup plus subtil que la simple existence de fluctuations plus importantes dans la phase (des bulles) qui domine au voisinage de la dissociation.

L'uniformité de l'amplification des fluctuations de température le long de la chaîne mesurée sur les spectres de puissance ne signifie cependant pas que les fluctuations sont uniformes sur la séquence (rappelons en effet que les spectres de puissance ne contiennent aucune information relative aux phases des composantes aux différentes fréquences). Cela peut être vérifié sur la figure 3.9, qui montre l'évolution temporelle de la température

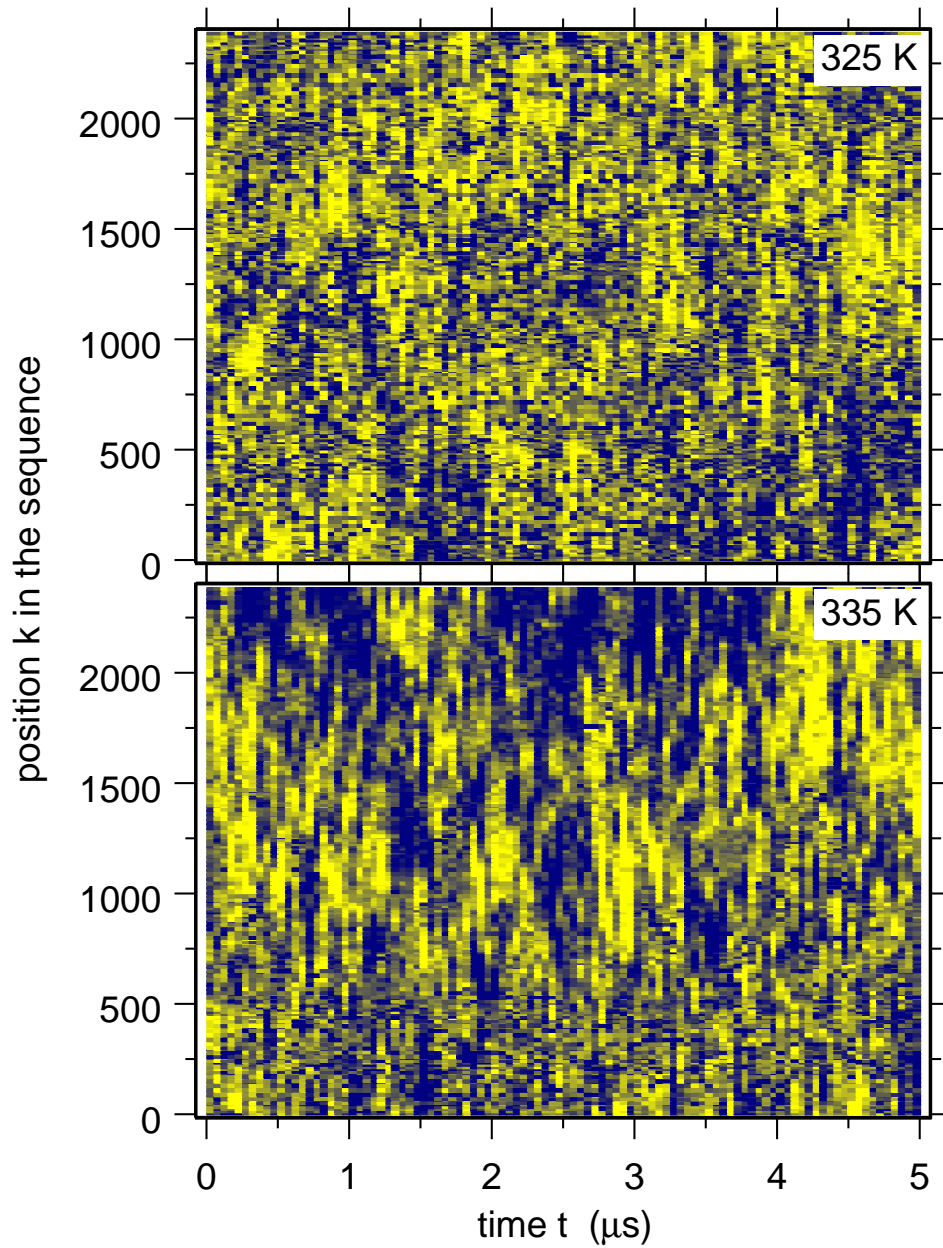


Fig. 3.9: Evolution temporelle de la température de la séquence inhibiteur de 2399 paires de bases calculée avec le modèle JB. Chaque point correspond à la température instantanée moyennée sur 10 paires de bases successives et sur un intervalle de 50 ns. Les fluctuations de la température par rapport à la moyenne s'étendent de -10 K (les points bleus/sombres) à 10 K (les points jaunes/clairs).

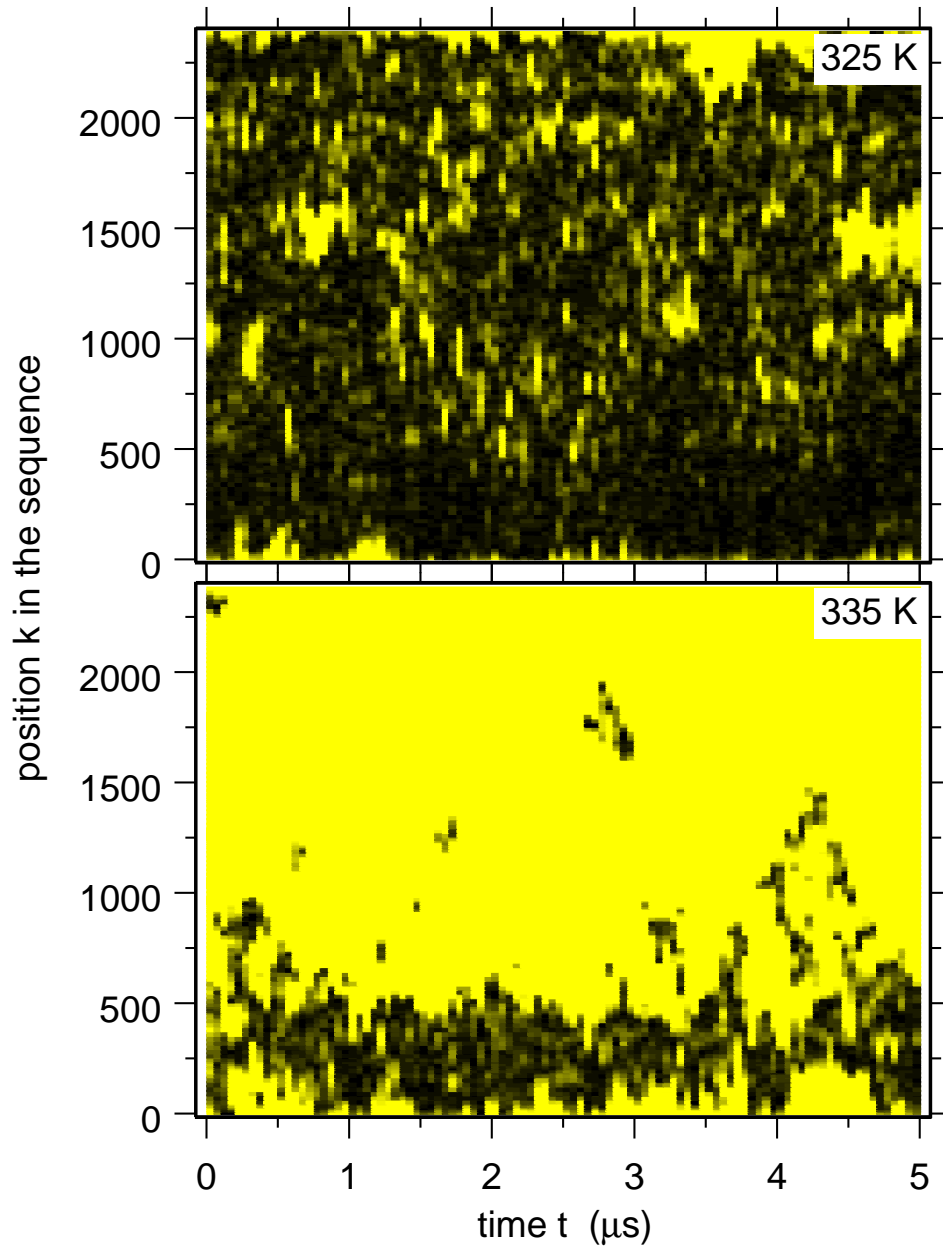


Fig. 3.10: Evolution temporelle de la distance entre les bases de chaque paire de la séquence inhibiteur de 2399 paires de bases calculée avec le modèle JB. Chaque point correspond à la distance instantanée moyennée sur 10 paires de bases successives et sur un intervalle de 50 ns. Ces distances s'étendent de 1 Å (points noirs) à 20 Å (points jaunes/clairs). Pour ce modèle, le seuil de dissociation est 12 Å.

de la séquence inhibiteur de 2399 paires de bases à 325 K et 335 K. Sur cette figure, chaque pixel correspond à la température physique locale moyennée sur 10 paires de bases successives et pendant 50 ns. Même si le diagramme à 325 K et le profil de dénaturation de la figure 3.7 semblent ne pas être corrélés, cette corrélation apparaît à 335 K, c'est à dire au voisinage de la dénaturation. A cette température, on observe une différence marquée entre la dynamique des paires de bases pour $k < 750$ et $k > 750$. Les paires de bases se trouvant dans la première portion (la partie la plus riche en GC de la séquence) possèdent des fluctuations du même type que celles observées à 325 K tandis que les paires de bases de la seconde portion (la partie la plus riche en AT de la séquence, qui se dissocie en premier) exhibent des fluctuations de durées de vie relativement longues correspondant aux excitations cohérentes de nombreuses particules. A titre de comparaison, on montre en figure 3.10 l'évolution temporelle, non plus de la température, mais de la distance entre les bases de chaque paire pour la même trajectoire que la figure 3.9.

3.4 Excitations collectives, amplification des fluctuations et comportement en $1/f$

Comme je l'ai noté dans la section 3.1, les deux résultats principaux obtenus avec tous les modèles et les séquences étudiés dans ce travail sont (i) l'amplification des fluctuations à la température de dissociation et (ii) le fait que le spectre de puissance évolue en $1/f$ sur un large intervalle de fréquences. Pour les modèles DPB et JB (voir les équations (2.9) et (2.15)) présentant uniquement des degrés de liberté d'élongation, l'échelle de fréquences sur laquelle la dépendance en $1/f$ du spectre de puissance reste valable s'élargit abruptement au voisinage immédiat de la transition. Pour le modèle DPB, cette échelle s'étend jusqu'à 100 MHz à 354 K (fig 3.1) et elle atteint 2 MHz pour le modèle JB, que ce soit pour l'inhibiteur à 335 K ou la séquence ATAT... à 310 K. Pour le modèle BSJ plus complexe, qui prend aussi en compte certaines coordonnées de pliage, cet intervalle ne change pas beaucoup en fonction de la température et s'étend jusqu'à 100 MHz pour la séquence actine de 1793 paires de bases (fig 3.4). Ces fluctuations en

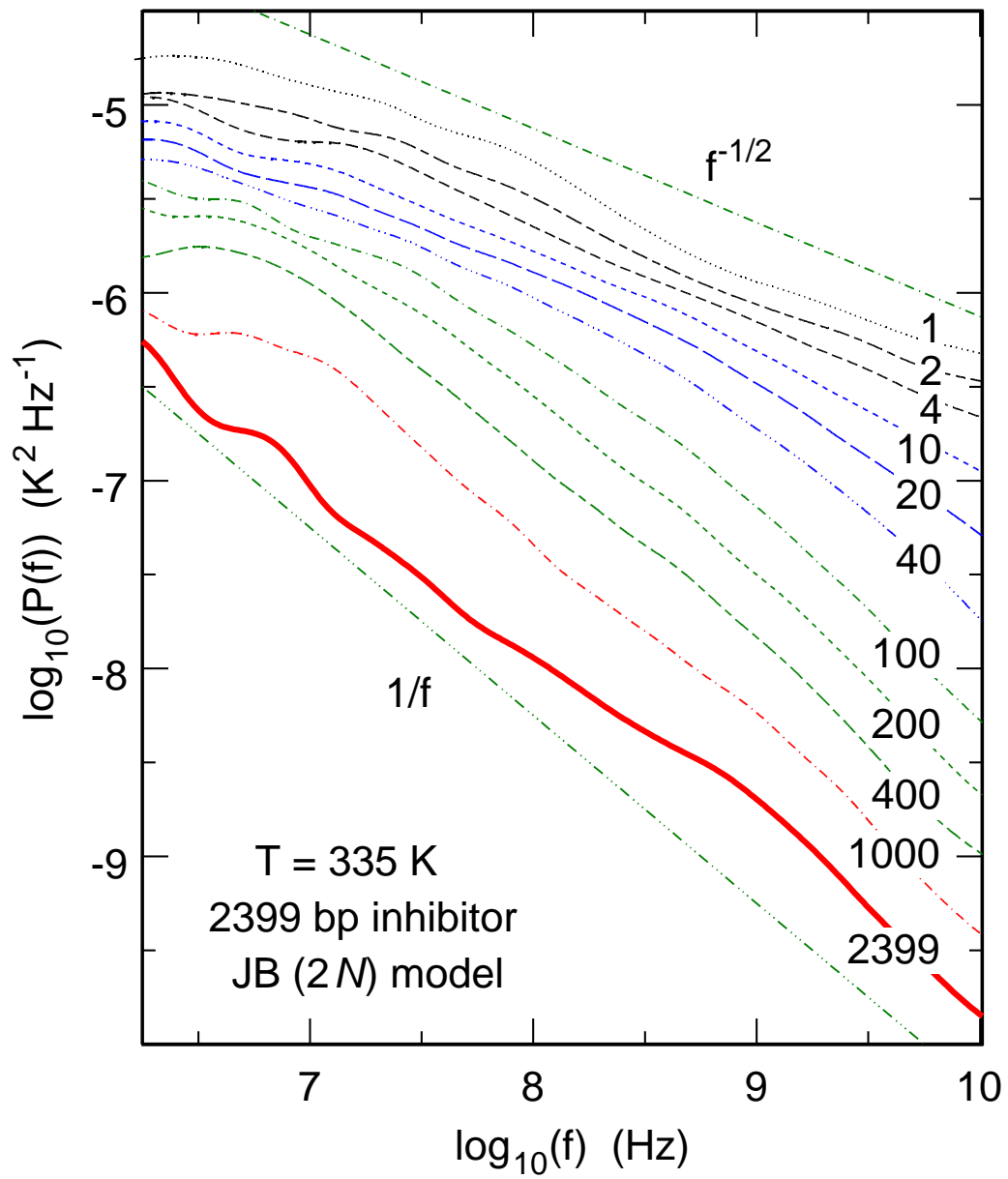


Fig. 3.11: Spectres de puissance des fluctuations de température à 335 K (la température de dissociation) pour la séquence inhibiteur de 2399 paires de bases et le modèle JB (2.15). Chaque courbe correspond à une différente longueur N_j sur laquelle la température est moyennée.

$1/f$, souvent appelées *flicker noise* ou *pink noise* à cause du fait qu'elles se situent entre le bruit blanc (*white noise*) constant et le bruit rouge (*red noise*) en $1/f^2$, ont été observées dans un grand nombre de systèmes concernant l'électronique, la physique, la biologie, la géophysique, le trafic routier, la finance etc... La recherche d'une explication à son omniprésence est l'un des plus anciens problèmes non résolus de la physique. Cependant, si une théorie universelle du bruit en $1/f$ manque toujours, la raison de son existence a néanmoins été expliquée par des arguments ad-hoc pour chaque cas réel où il a été observé. Dans les systèmes dynamiques, le scénario conduisant au bruit en $1/f$ suppose que ce phénomène existe dans le cas d'une transition stochastique [22], c'est à dire à une échelle d'énergie où les régions quasi-régulières de l'espace des phases et les régions stochastiques correspondant à une dynamique chaotique sont séparées par une hiérarchie de tores brisés (*cantori*). Le fait qu'une particule soit retenue au voisinage des tores brisés et qu'elle soit libérée avec des taux aléatoires conduit au bruit en $1/f$ en question. La vérification de cette explication est très difficile pour les systèmes à beaucoup de degrés de liberté, puisque dans ce cas la construction des sections de Poincaré comme celles des figures 2-4 de la référence [22] devient impossible. De plus, nous avons vérifié le comportement du plus grand exposant de Lyapunov λ_1 en fonction de la température. Nous avons constaté que pour tous les modèles considérés, λ_1 devient nul de façon discontinue quand les deux brins se séparent mais qu'il ne manifeste aucun autre comportement significatif au voisinage de la température critique (voir par exemple la figure 3 de la ref [42]). Ainsi que nous allons le voir dans la suite de cette section, ce que l'on peut en revanche affirmer quant à l'origine du bruit en $1/f$, c'est que pour les modèles et les séquences étudiés dans ce travail la dépendance en $1/f$ et l'amplification des fluctuations au voisinage de la dissociation constituent un phénomène collectif dans le sens où elles résultent de la *dynamique cohérente* d'une multitude de paires de bases.

Nous avons vu sur la figure 3.8 que le spectre de puissance obéit à une loi en $1/N_j$ à une fréquence donnée. Toutefois, cela n'est vrai que pour N_j supérieur à un seuil qui est de l'ordre de 100-200 pour la séquence inhibiteur de 2399 paires de bases et le modèle

JB à 335 K. On a illustré sur la figure 3.11 l'évolution du spectre de puissance pour un intervalle de N_j plus étendu ($1 \leq N_j \leq 2399$). On constate que pour les petites valeurs de N_j les fluctuations obéissent à une loi en $1/\sqrt{f}$ (c'est à dire avec une pente -1/2) plutôt qu'en $1/f$. Lorsque N_j croît, l'amplitude des fluctuations à hautes fréquence décroît plus vite que celle des fluctuations à basses fréquence jusqu'à ce que la loi en $1/f$ soit atteinte au seuil de 100-200 paires de bases. Pour les valeurs de N_j plus grandes que le seuil, les fluctuations à basses et à hautes fréquences décroissent avec la même intensité de sorte qu'on observe une dépendance en $1/f$ après avoir moyenné sur la séquence entière.

Cet aspect est étudié plus en détail sur la figure 3.12 qui montre l'évolution du carré du module de la transformée de Fourier des fluctuations de température en fonction de N_j pour différentes fréquences augmentant de façon exponentielle (2, 20, 200 MHz et 2 GHz) et à trois températures caractéristiques : 315 K (ADN double brin), 335 K (température de dénaturation) et 345 K (ADN dissocié). Il est possible d'interpréter cette figure à partir du *théorème de Wiener-Khinchin* selon lequel le spectre de puissance d'un signal est égal à la transformée de Fourier de sa fonction d'auto-corrélation. Cela signifie que le spectre de puissance des fluctuations de température obtenu à partir d'une portion N_j de la séquence en question est proportionnel à la transformée de Fourier de

$$C(N_j, \tau) = \frac{1}{N_j^2} \sum_{k=1}^{N_j} \langle p_k^2(t) p_k^2(t + \tau) \rangle + \frac{1}{N_j^2} \sum_{k=1}^{N_j} \sum_{m=1}^{N_j} \langle p_k^2(t) p_m^2(t + \tau) \rangle_{m \neq k} \quad (3.2)$$

Le deuxième terme à droite de cette équation décrit les corrélations des fluctuations correspondant aux différentes paires de bases et devient nul si les fluctuations sont indépendantes. En l'absence de corrélations, le spectre de puissance doit donc varier comme $1/N_j$ si les fluctuations de $p_k^2(t)$ sont uniformes le long de la séquence, autrement dit si elles ne dépendent pas de la position de la paire de bases k . D'autre part, si toutes les paires de bases appartenant à la même portion de la séquence évoluent comme un corps rigide, alors le second terme à droite de l'équation (3.2) devient égal à $(N_j - 1)$ fois le premier terme et le spectre de puissance ne dépend plus de N_j . Un dernier cas

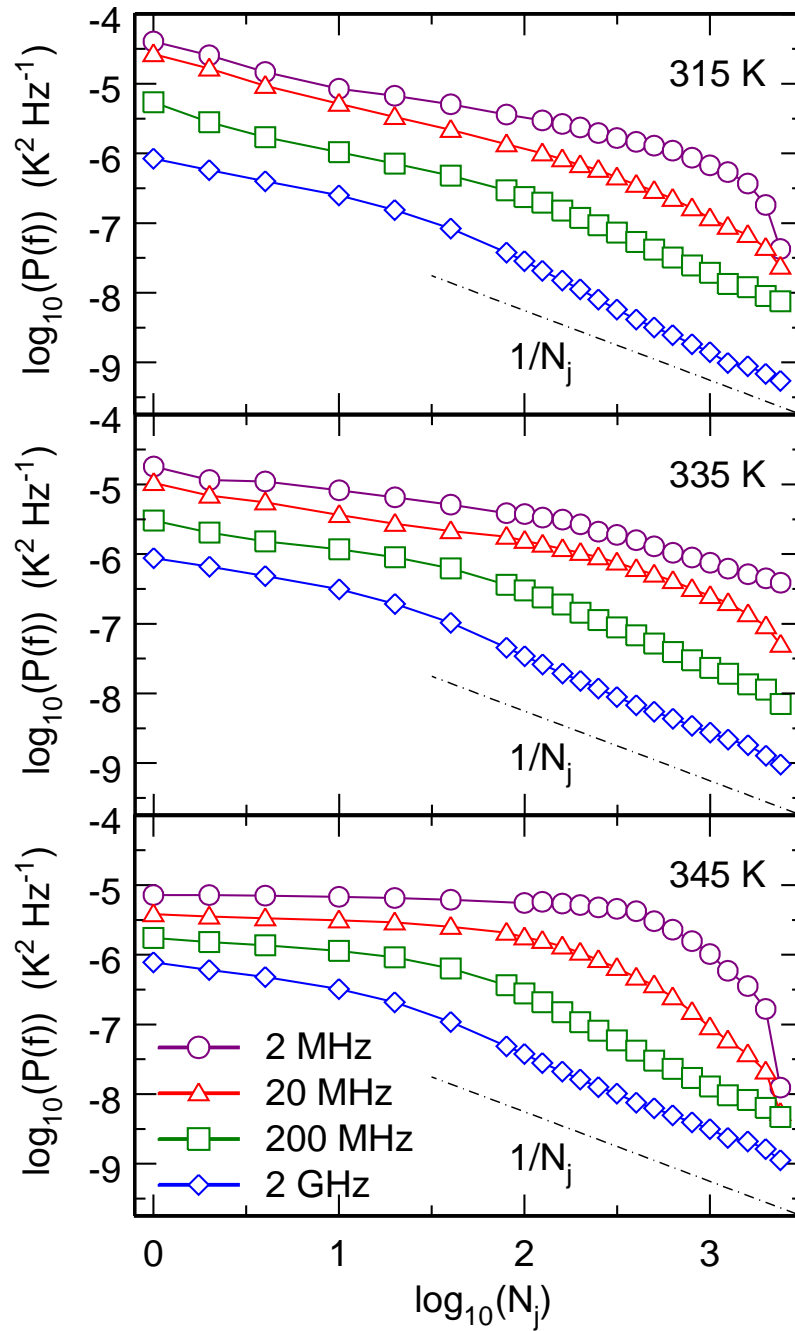


Fig. 3.12: Carré du module de la transformée de Fourier des fluctuations de température de la séquence inhibiteur de 2399 paires de bases en fonction de N_j pour des fréquences dont la valeur augmente de façon exponentielle (2, 20, 200 MHz et 2 GHz) et pour trois températures caractéristiques : 315 K (figure du haut, séquence liée), 335 K (figure du milieu, température de dissociation) et 345 K (figure du bas, molécule dissociée).

est celui des corrélations destructives conduisant à une variation du spectre de puissance plus abrupte que $1/N_j$.

La figure 3.12 montre qu'à l'échelle d'une seule paire de bases ($N_j = 1$) les fluctuations de température diminuent lorsque la température augmente. L'amplification des fluctuations de température à la dénaturation résulte donc des corrélations qui deviennent de plus en plus marquées lorsque l'on approche cette température. En particulier, on voit que pour $N_j \geq 20$, l'intensité des fluctuations à hautes fréquences ($f \geq 200$ MHz) varie comme $1/N_j$ à toutes les températures tandis que la pente est plus faible pour $N_j < 20$. On en déduit que ces fluctuations à hautes fréquences sont issues de la dynamique rapide de chaque paire de bases, dont la corrélation ne s'étend cependant pas au delà des quelques voisins les plus proches. D'autre part, aux températures 315 K et 345 K, l'intensité des fluctuations à basses fréquences ($f \leq 20$ MHz) décroît selon une pente plus aigue que $1/N_j$ pour les grandes valeurs de N_j . Cela signifie que des zones plus chaudes ou plus froides impliquant jusqu'à plusieurs centaines de paires de bases se déplacent de façon un peu aléatoire le long de la séquence, de sorte que la température de la séquence complète varie bien moins que celle de portions données de la séquence. Cela peut être vérifié sur la figure 3.9 où les régions sombres semblent diffuser de façon aléatoire le long de la séquence. Ce phénomène est semblable aux flip-flop des molécules d'eau. Finalement, à la température de dénaturation (335 K), l'intensité des fluctuations à basses fréquences ($f = 2$ MHz) varie moins vite qu'une dépendance en $1/N_j$. Cela est probablement dû au fait qu'à la dénaturation la longueur de corrélation diverge pour le modèle JB aussi bien que pour le modèle DBP [43], ce qui conduit à une dynamique collective très lente. On peut donc raisonnablement en déduire que la divergence de la longueur de corrélation à la température de dénaturation est responsable du fait que la région de validité de la dépendance en $1/f$ s'étend de façon abrupte aux basses fréquences (voir les figures de 3.1 à 3.3). Pour le modèle BSJ plus complexe de l'équation (2.18), cette région semble varier plus faiblement avec la température (voir la figure 3.4). Cela peut signifier que pour ce modèle la transition est plus discontinue que pour les deux modèles

précédents et que la région de singularité de la transition (et donc celle de la divergence de la longueur de corrélation) est trop étroite pour que l'on puisse la localiser par la dynamique de Langevin. Une étude plus détaillée est nécessaire pour éclaircir ce point.

3.5 *Conclusion*

Dans cette partie, nous avons réalisé une étude théorique des fluctuations de température de l'ADN au voisinage de la dénaturation et observé une amplification marquée de ces fluctuations à la température critique. Un phénomène très similaire a été discuté dans le travail expérimental de Nagapriya et ses collaborateurs [21], bien que dans un domaine de fréquences beaucoup plus basses. Nous avons montré que ce phénomène disparaît dans le cas d'un coefficient de dissipation dépassant quelques dizaines de ps^{-1} et insisté sur le rôle que peut jouer la nature de la solution dans les expériences réelles. Nous avons cherché une corrélation entre le développement des bulles et l'amplification des fluctuations mais nous n'avons détecté aucun effet direct des bulles sur ce phénomène. Finalement, nous avons montré qu'à l'échelle d'une seule paire de bases l'amplification des fluctuations et la dépendance en $1/f$ du spectre de puissance ne sont pas observées et que ces aspects se manifestent uniquement après avoir additionné la contribution d'un grand nombre de paires de bases. Nous en avons conclu que ces effets sont dus à des excitations collectives des paires de bases qui sont nettement facilitées au voisinage de la température critique par la divergence de la longueur de corrélation.

L'apparition de fluctuations en $1/f$ de la température (ou de façon équivalente de l'énergie potentielle totale) a été prédite pour plusieurs systèmes biophysiques [44]-[47]. L'échelle de fréquence où cette dépendance en $1/f$ est attendue peut s'étendre jusqu'à 1 GHz [44]-[47] ou même jusqu'à quelques MHz, comme on vient de le voir dans ce travail. Le travail récent de Nagapriya et al [21] est cependant l'une des rares études expérimentales s'intéressant aux fluctuations de température dans ce type de systèmes. Selon la figure 3 de la référence [21], il semble que le spectre de puissance mesuré décroît beaucoup plus rapidement qu'une simple dépendance en $1/f$ à l'échelle considérée de

0.01-1 Hz. Il est donc probable que des processus lents qui ne sont pas pris en compte dans nos modèles ont des effets à cette échelle de fréquences. Des études théoriques et expérimentales supplémentaires seront donc nécessaires afin d'expliquer le passage entre ces deux échelles de fréquence et compléter notre compréhension de la dynamique des fluctuations de température de l'ADN.

4. TRANSITION DE PHASE DE DÉNATURATION DES SÉQUENCES HOMOGÈNES INFINIES

4.1 Quelques notions de transitions de phase

4.1.1 Définitions de base

L'un des enjeux les plus importants de la physique est sans doute de pouvoir prédire les propriétés macroscopiques d'un système à partir de sa description microscopique. Cette motivation est issue du fait qu'un système réel est connu à travers l'interaction de ses composants à l'échelle microscopique mais que les appareils de mesure réagissent en général au comportement de ce système à l'échelle macroscopique. Si l'on désigne par a la distance caractéristique où le caractère discret (ou granulaire) d'un système devient important, le but de la physique statistique est précisément d'expliquer les propriétés d'un système à une échelle $d \gg a$ à partir d'une description Hamiltonienne à l'échelle $d \sim a$.

En physique statistique, un système est décrit à partir des *observables* macroscopiques comme l'entropie S , la magnétisation M , la chaleur spécifique C_V ou la compressibilité κ . Ces observables sont fonctions des *champs* externes comme la température T , la pression P ou le champ magnétique h . Dans le cas des systèmes complexes, c'est à dire des systèmes constitués de nombreux composants individuels, il se passe un phénomène collectif très intéressant lorsque l'on fait varier ces champs externes et que l'on atteint certaines valeurs précises : le système entier change brusquement d'état de symétrie. C'est ce que l'on appelle une *transition de phase*. On peut donner pour exemple l'évaporation de l'eau ou la transition du fer de l'état paramagnétique à l'état ferromagnétique à une température critique T_c . Une transition de phase a lieu plus précisément aux valeurs des

champs externes pour lesquelles une ou plusieurs observables manifestent une singularité. C'est le cas par exemple de l'évaporation d'un liquide avec l'entropie qui fait un saut discontinu

$$\Delta S = \frac{\Delta Q}{T_c} \quad (4.1)$$

où ΔQ est la chaleur latente absorbée par le système.

Les transitions de phase peuvent être classées en deux catégories dont voici les caractéristiques habituelles :

Transitions du premier ordre (discontinues)

- Au point critique a lieu la coexistence des phases. Autrement dit, à la température critique, l'état ordonné et l'état désordonné (correspondant aux états stables des deux côtés de la transition) peuvent coexister dans le même système.
- La longueur de corrélation ξ , qui mesure l'étendue des corrélations dans le système, est finie.
- Il existe une chaleur latente finie absorbée ou libérée par le système au point de transition.

Transitions du second ordre (continues)

- Il n'y a pas de mélange des phases au point critique.
- La longueur de corrélation est divergente :

$$\lim_{T \rightarrow T_c} \xi = \infty$$

- La chaleur latente est nulle.

Le mélange des phases lors d'une transition de phase discontinue est directement liée à la finitude de la chaleur latente qui intervient. En fait, puisque le système n'est pas capable d'échanger cette énergie avec son environnement de façon instantanée, seul un nombre limité de domaines subissent la transition. C'est le cas du mélange vapeur-liquide

lors de l'ébullition de l'eau. Dans la nature, la majorité des systèmes subit une transition du premier ordre. Cependant, des conditions particulières peuvent conduire à la disparition des discontinuités et transformer ainsi une transition de phase du premier ordre en une transition continue. Cela arrive par exemple lorsque l'on augmente la pression d'un mélange liquide-solide. A une valeur précise de la pression imposée, il devient impossible de distinguer les deux phases et la coexistence disparaît. C'est également le cas d'un système homogène ayant une transition abrupte qui est lissée dès que des défauts locaux ou du désordre (une distribution inhomogène des composants individuels du système) surgissent. L'évolution de la nature des transitions de phase en présence du désordre a été étudiée de façon très poussée [48, 50]. J'y reviendrai au dernier chapitre de cette thèse pour le cas de la dénaturation de l'ADN.

La grandeur qui distingue les différentes phases s'appelle le *paramètre d'ordre* m . Bien qu'il s'agisse souvent d'une dérivée première de l'énergie libre, il n'existe aucune règle générale pour choisir le bon paramètre d'ordre pour un système particulier. Le principal critère qui peut nous guider est que le paramètre d'ordre doit être une fonction qui prend des valeurs clairement distinctes dans chacune des phases. Par exemple, dans le cas de la transition ferromagnétique \rightarrow paramagnétique, m est la magnétisation par unité de volume, qui est nulle pour $T \geq T_c$ et finie pour $T < T_c$. Pour l'ébullition de l'eau, m correspond à la densité de vapeur (ou de liquide) du système. On constate que m est en général une variable intensive.

Je vais maintenant définir les observables thermodynamiques dont nous allons étudier les singularités dans la suite de ce chapitre et de cette thèse. L'ensemble statistique le plus approprié pour étudier la thermodynamique d'une séquence d'ADN dans un solvant est l'*ensemble canonique*. Dans cet ensemble, le sous-système (le double brin d'ADN) est en contact avec un réservoir de chaleur de taille infinie. De plus, nous allons coupler la molécule à un champ scalaire h qui est ici un outil purement mathématique. Dans ce cas, l'Hamiltonien prend la forme

$$H = T(\mathbf{p}) + V(\mathbf{y}) + h \sum_i f(y_i) \quad (4.2)$$

où le premier et le deuxième terme correspondent respectivement à l'énergie cinétique et potentielle, et $f(y_i)$ est la fonction qui couple le champ h à l'Hamiltonien. La fonction de partition de cet Hamiltonien, qui constitue un comptage dans l'espace des phases du nombre d'états configurationnels accessibles au système, s'écrit

$$Z(h) = \int_{\Omega} d\mathbf{p} d\mathbf{y} e^{-\beta[T(\mathbf{p})+V(\mathbf{y})]-\beta h \sum_{i=1}^N f(y_i)} \quad (4.3)$$

où $\beta = 1/(k_B T)$ est la température inverse et l'intégration est effectuée sur l'ensemble de l'espace des phases. Dans l'ensemble canonique, l'état d'équilibre thermique du système correspond au minimum de l'énergie libre F par rapport à la température, où

$$F = -k_B T \ln(Z). \quad (4.4)$$

Si l'on tente, comme c'est souvent le cas, de prendre comme paramètre d'ordre la dérivée de l'énergie libre par site, $f = F/N$, par rapport au champ externe h ,

$$m(h) = \frac{\partial f}{\partial h}, \quad (4.5)$$

on obtient, en utilisant les éq. (4.3) et (4.4)

$$m(h) = \frac{1}{Z} \int_{\Omega} d\mathbf{p} d\mathbf{y} e^{-\beta H(\mathbf{p}, \mathbf{y}) - \beta h \sum_{i=1}^N f(y_i)} \sum_{i=1}^N f(y_i) = \langle f(y) \rangle. \quad (4.6)$$

Pour des raisons "historiques" j'ai noté f à la fois la fonction de couplage au champ (éqs. (4.2), (4.3) et (4.6)) et l'énergie libre par particule (éq. (4.5)). Le contexte est en général toutefois suffisant pour éviter toute confusion.

On vérifiera dans la suite de ce chapitre que si l'on suppose que $f(y_i)$ est de la forme $f(y_i) = y_i$ ou $f(y_i) = y_i^2$, alors le paramètre $m(h)$ défini par les équations (4.5) et (4.6) présente bien la caractéristique principale d'un paramètre d'ordre, dans la mesure où il

prend des valeurs clairement différentes de part et d'autre de la température critique. Plus précisément, $m(h)$ est fini pour l'ADN double brin et infini pour l'ADN simple brin.

L'entropie et la chaleur spécifique sont définies par les dérivées successives de l'énergie libre :

$$S = -\frac{\partial F(h=0)}{\partial T} \quad (4.7)$$

$$C_V = -T \frac{\partial^2 F(h=0)}{\partial T^2}. \quad (4.8)$$

La susceptibilité, qui représente la réponse du système à une variation du champ externe, est calculée à partir de l'expression

$$\chi = \frac{\partial m}{\partial h}. \quad (4.9)$$

Le facteur de structure peut quant à lui s'exprimer sous la forme

$$S(q) = \left\langle \left| \sum_{n=1}^N (y_n - \langle y_n \rangle) e^{iqan} \right|^2 \right\rangle. \quad (4.10)$$

où a est une longueur caractéristique du système. Ici encore, il peut y avoir confusion entre la longueur a introduite dans l'équation (4.10) et le paramètre a définissant la largeur du potentiel de Morse dans l'équation (2.4). Je préciserai les choses lorsque le doute sera possible. La longueur de corrélation est alors calculée à partir du facteur de structure sous la forme

$$\xi^2 = -\frac{1}{2a^2 S(q)} \frac{d^2 S(q)}{dq^2} \Big|_{q=0}. \quad (4.11)$$

La fonction de corrélation à deux points qui mesure la corrélation des fluctuations thermiques entre différents domaines du système s'écrit enfin

$$G_{kl} = \langle y_k y_l \rangle - \langle y_k \rangle \langle y_l \rangle. \quad (4.12)$$

Notons que dans un système qui présente une symétrie translationnelle, G_{kl} est fonction uniquement de l'écart $|k - l|$. De plus, à une dimension et près de la limite thermodynamique où les effets de bord sont négligeables, on a

$$G_{kl} \propto e^{-\frac{a|k-l|}{\xi}}, \quad (4.13)$$

ce qui constitue une méthode alternative pour calculer la longueur de corrélation.

4.1.2 Exposants critiques

Au voisinage du point critique, les singularités des fonctions thermodynamiques se manifestent sous forme de divergences ou de discontinuités. De plus, les observables obéissent à des lois de puissance, fonctions des champs externes, au voisinage immédiat du point de transition. C'est le phénomène de *scaling*.

La singularité de chaque observable est caractérisée par un *exposant critique*. J'ai énuméré ci-dessous les 6 exposants critiques, α , β , γ , δ , ν , η , correspondant aux fonctions thermodynamiques auxquelles nous allons nous intéresser dans la suite de ce chapitre. En définissant la température réduite par

$$t = \left| \frac{T_c - T}{T_c} \right|, \quad (4.14)$$

on a, pour la chaleur spécifique par particule $c_V = C_V/N$

$$c_V \simeq t^{-\alpha}, \quad (4.15)$$

pour le paramètre d'ordre

$$m|_{h=0} \simeq t^\beta \quad (4.16)$$

$$m|_{t=0} \simeq h^{1/\delta}, \quad (4.17)$$

pour la susceptibilité

$$\chi \simeq t^{-\gamma}, \quad (4.18)$$

pour la longueur de corrélation

$$\xi \simeq t^{-\nu}, \quad (4.19)$$

et pour le facteur de structure

$$S(q, T_c) \simeq q^{\eta-2}. \quad (4.20)$$

L'existence des exposants critiques est intimement liée à un autre aspect remarquable des transitions de phase, à savoir l'*universalité*. En fait, des systèmes qui diffèrent dans leur structure microscopique peuvent avoir le même comportement critique. Un tel ensemble définit une *classe d'universalité*. C'est le cas d'une transition liquide-gaz qui est en général caractérisée par les mêmes exposants critiques, quelle que soit la nature du liquide en question. Il est important de préciser que les coefficients de proportionnalité des équations (4.15)-(4.20) ne sont pas universels, autrement dit l'universalité ne concerne que les exposants critiques. De plus, en ce qui concerne un système particulier, les calculs du groupe de renormalisation [51], qui ont pour but de redéfinir les paramètres du système de façon récursive afin de s'approcher du point critique, montrent que l'exposant critique caractérisant la singularité de chaque fonction thermodynamique a la même valeur des deux côtés de la transition, ce qui n'est pas toujours le cas pour le coefficient de proportionnalité. Cela peut conduire par exemple à l'asymétrie du pic de la chaleur spécifique par rapport à T_c .

Avant d'explorer les relations entre les différents exposants critiques des équations (4.15)-(4.20), je voudrais apporter une précision concernant la chaleur latente, qui sera d'une importance capitale pour la suite de l'exposé. La chaleur latente échangée par le système avec son environnement est définie comme

$$L = \int_{T_c - \epsilon}^{T_c + \epsilon} C_V dT. \quad (4.21)$$

où $\epsilon \ll T_c$ désigne le voisinage immédiat du point critique. Après avoir remplacé l'intégrande par son expression (4.8) et intégré par parties, on trouve

$$L = [S(T_c + \epsilon) - S(T_c - \epsilon)] T_c = T_c \Delta S. \quad (4.22)$$

On retrouve donc l'expression (4.1). En d'autres termes, l'existence d'une chaleur latente nécessite le saut de l'entropie au point critique. Nous allons maintenant obtenir la condition nécessaire pour avoir ce saut. En intégrant la relation

$$C_V = T \frac{\partial S}{\partial T} \quad (4.23)$$

avec la relation (4.15) des deux côtés de la transition, c'est à dire pour $T < T_c$ et $T > T_c$ et en soustrayant les valeurs obtenues, on trouve

$$\Delta S \propto \epsilon^{1-\alpha}. \quad (4.24)$$

En tenant compte de la petitesse de ϵ , il devient clair qu'un saut significatif de l'entropie (et donc l'existence d'une chaleur latente) ne se produit que si $\alpha \geq 1$. Donc

$$\begin{aligned} \alpha < 1 &\implies \text{transition de phase continue} \\ \alpha \geq 1 &\implies \text{transition de phase discontinue.} \end{aligned} \quad (4.25)$$

C'est ce critère, que je considère comme le plus fiable, qui va nous guider dans la suite de ce travail.

4.1.3 Lois d'échelle

Les fonctions thermodynamiques (4.15)-(4.20) étant reliées entre elles par des dérivations, on s'attend à ce qu'il existe des relations entre les 6 exposants critiques. Pour obtenir ces relations, un certain nombre d'hypothèses sont nécessaires. La première est l'*hypothèse d'homogénéité* selon laquelle, au voisinage du point critique, la partie singulière de l'énergie libre est une fonction homogène de la forme

$$f_{sing}(t, h) = t^{2-\alpha} g_f \left(\frac{h}{t^\Delta} \right) \quad (4.26)$$

où Δ est le "*gap exponent*". En utilisant cette relation avec les équations (4.5), (4.8) et (4.9) et en identifiant les résultats avec les lois (4.15), (4.16) et (4.18), on obtient $\Delta = \beta\delta$ et deux lois polynomiales qui relient les différents exposants critiques,

$$\text{Rushbrooke :} \quad \alpha + 2\beta + \gamma = 2 \quad (4.27)$$

$$\text{Widom :} \quad \gamma - \beta(\delta - 1) = 0. \quad (4.28)$$

Ce sont les deux premières *lois d'échelle*. Dans l'approximation de la méthode du col de la théorie de Ginzburg-Landau [51], qui consiste à développer l'énergie libre en fonction du paramètre d'ordre, on obtient une loi d'homogénéité de la forme

$$f_{sing}(t, h) = t^2 g_f \left(\frac{h}{t^{3/2}} \right) \quad (4.29)$$

Pour obtenir l'équation (4.26) à partir de l'équation (4.29), la généralisation consiste à introduire l'exposant α et un exposant Δ général au lieu du facteur $3/2$. C'est la raison pour laquelle on appelle la relation (4.26) "l'hypothèse" d'homogénéité. D'autre part, il faut préciser qu'il est possible d'obtenir la loi (4.26) par un calcul plus rigoureux basé sur le groupe de renormalisation.

L'hypothèse d'homogénéité, qui permet de relier les exposants des différentes quanti-

tés déduites à partir de l'énergie libre, ne fait aucune prédiction sur le comportement des fonctions de corrélation. On devrait noter que la divergence éventuelle de la longueur de corrélation au point critique implique celle des fonctions de réponse comme la susceptibilité (4.18) ou le facteur de structure (4.20). Pour obtenir une relation entre l'exposant ν et les autres exposants caractéristiques, l'hypothèse d'homogénéité (4.26) doit être remplacée par l'*hypothèse d'homogénéité généralisée* qui, comme son nom l'indique, est plus forte et suppose que

- la longueur de corrélation a une forme homogène,

$$\xi(t, h) = t^{-\nu} g_{\xi} \left(\frac{h}{t^{\Delta}} \right), \quad (4.30)$$

- la longueur de corrélation, qui est la seule longueur interne divergente au point critique pour la plupart des systèmes, est également la seule longueur caractéristique qui a une contribution directe à la singularité des fonctions thermodynamiques,

$$f_{sing} = L^{-d} F_{sing} \left(\frac{\xi}{L} \right). \quad (4.31)$$

La relation (4.31) s'appelle également l'hypothèse de *hyperscaling*. Finalement, à champ externe nul, l'équation (4.31) conduit aux lois d'échelle de Rushbrooke et Widom, plus une loi d'échelle supplémentaire, à savoir la loi de *Josephson*,

$$\text{Josephson :} \quad 2 - \alpha = \nu d. \quad (4.32)$$

Enfin, il est possible d'obtenir une loi reliant l'exposant η qui caractérise la dissipation spatiale des fonctions de corrélation aux autres exposants. Pour cela, on doit d'abord prendre en compte le comportement des fonctions de corrélation au point critique,

$$G(x) \propto \frac{1}{|x|^{d-2+\eta}} \quad (4.33)$$

qui est vrai en principe pour $d \geq 2$. Dans cette expression, x est la distance $k - l$ de

l'équation (4.13). En tenant compte de la relation

$$\chi \sim \int_0^\xi d^d x G(x), \quad (4.34)$$

où on a supposé que le paramètre d'ordre m reste fini à T_c , on trouve la dernière loi d'échelle, c'est à dire la loi de *Fisher*,

$$\text{Fisher :} \quad \gamma = (2 - \eta)\nu d. \quad (4.35)$$

Dans la suite de ce chapitre, je vais calculer les exposants critiques et vérifier la validité des lois d'échelle pour trois modèles non-linéaires d'ADN. Un calcul rigoureux des exposants critiques nécessitant une localisation précise de T_c , je vais utiliser la méthode la plus appropriée pour cela, c'est à dire la méthode TI que je vais maintenant présenter.

4.1.4 Intégrale de Transfert pour des séquences homogènes de longueur infinie

Pour les modèles d'ADN des éq. (2.8), (2.9) et (2.11) complétés par le couplage avec un champ de la forme $h \sum_i f(y_i)$, la fonction de partition (4.3) peut s'écrire

$$Z = \int dy_1 dy_2 \dots dy_N e^{-\beta \sum_n \{V_M(y_n) + W(y_n, y_{n-1}) + hf(y_n)\}}, \quad (4.36)$$

où les intégrales multiples sont calculées sur l'espace réel entier. Notons que l'éq. (4.36) ne prend pas en compte le terme trivial provenant de l'énergie cinétique. Des conditions aux bords périodiques sont adoptées dans ce chapitre dans l'expression (4.36) afin de simplifier les calculs. Cela correspond à lier la N -ème paire de base à la première et on note $y_0 = y_N$. Puisque j'étudie dans ce chapitre des séquences d'ADN à la limite thermodynamique $N \rightarrow \infty$, les conditions aux bords n'ont aucun effet significatif. En définissant le noyau

$$K(y_n, y_{n-1}) = \exp \left[-\beta \left\{ \frac{1}{2} V_M(y_n) + \frac{1}{2} V_M(y_{n-1}) + W(y_n, y_{n-1}) + \frac{h}{2} [f(y_n) + f(y_{n-1})] \right\} \right], \quad (4.37)$$

la fonction de partition (4.36) peut s'écrire sous la forme

$$Z = \int dy_1 dy_2 \cdots dy_N \prod_{n=1}^N K(y_n, y_{n-1}). \quad (4.38)$$

La méthode de l'intégrale de transfert (TI) [52] consiste à développer ce noyau sur une base orthonormée,

$$K(y_n, y_{n-1}) = \sum_i \lambda_i \Phi_i(y_n) \Phi_i(y_{n-1}), \quad (4.39)$$

où les $\{\Phi_i\}$ et $\{\lambda_i\}$ sont les vecteurs et valeurs propres de l'opérateur de transfert, qui satisfont

$$\int dx K(x, y) \Phi_i(x) = \lambda_i \Phi_i(y), \quad (4.40)$$

avec $\lambda_0 > \lambda_1 > \lambda_2 \cdots$ pour $T < T_c$. La résolution de l'éq. (4.40) sera discutée dans les sections suivantes. En injectant la décomposition (4.39) dans l'éq. (4.38) et en regroupant les fonctions ayant la même variable, la fonction de partition peut s'écrire comme

$$Z = \sum_{i_1, \dots, i_N} \lambda_{i_1} \cdots \lambda_{i_N} \int dy_1 \Phi_{i_1}(y_1) \Phi_{i_2}(y_1) \int dy_2 \Phi_{i_2}(y_2) \Phi_{i_3}(y_2) \cdots \int dy_N \Phi_{i_N}(y_N) \Phi_{i_1}(y_N). \quad (4.41)$$

Dans cette expression, chaque intégrale correspond à une fonction de Dirac, ce qui permet d'exprimer la fonction de partition en fonction des valeurs propres de la base,

$$Z = \sum_i \lambda_i^N. \quad (4.42)$$

La méthode TI donne par conséquent la possibilité de découpler, pour le calcul de Z , les degrés de liberté couplés de l'hamiltonien.

La valeur moyenne de l'écart de la k -ème paire de bases s'écrit

$$\langle y_k \rangle = \frac{1}{Z} \int dy_1 dy_2 \cdots y_k dy_k \cdots dy_N \prod_{n=1}^N K(y_n, y_{n-1}). \quad (4.43)$$

Il est clair, par symétrie translationnelle, que la quantité (4.43) ne dépend pas de la position k de la paire de bases. Si l'on suppose que $f(y_i) = y_i$, alors, selon l'éq. (4.6), on a $m = \langle y_k \rangle$, où m désigne $m|_{h=0}$, pour une séquence homogène de taille infinie. Dans ce cas, en définissant

$$M_{ij} = \int y dy \Phi_i(y) \Phi_j(y), \quad (4.44)$$

des calculs similaires à ceux qui ont conduit à (4.42) donnent pour le paramètre d'ordre

$$m = \frac{1}{Z} \sum_i \lambda_i^N M_{ii}, \quad (4.45)$$

et pour les moments non-diagonaux

$$\langle y_k y_l \rangle = \frac{1}{Z} \sum_{ij} \lambda_i^{N+k-l} \lambda_j^{l-k} M_{ij}^2 \quad (4.46)$$

où $l \geq k$. A la limite thermodynamique $N \rightarrow \infty$, seule la plus grande valeur propre λ_0 a une contribution significative à la fonction de partition (4.42). Cela donne pour la fonction de partition

$$Z = \lambda_0^N, \quad (4.47)$$

et par conséquent, pour l'énergie libre par particule

$$f = -k_B T \ln(\lambda_0), \quad (4.48)$$

et pour le paramètre d'ordre

$$m = M_{00}. \quad (4.49)$$

A la même limite, en utilisant les relations (4.6) et (4.46), un peu d'algèbre montre que les fonctions de corrélation (4.12) peuvent être exprimées sous la forme

$$G_{kl} = \sum_{i>0} M_{i0}^2 \left(\frac{\lambda_i}{\lambda_0} \right)^{|k-l|}, \quad (4.50)$$

qui devient, pour $|k - l| \gg 1$,

$$G_{kl} = M_{10}^2 \left(\frac{\lambda_1}{\lambda_0} \right)^{|k-l|}, \quad (4.51)$$

En faisant une identification avec l'éq. (4.13), on trouve finalement une expression importante, à savoir la longueur de corrélation,

$$\xi = \frac{1}{\ln \left(\frac{\lambda_0}{\lambda_1} \right)}. \quad (4.52)$$

Finalement, le calcul du facteur de structure [43, 53] s'effectue en injectant la relation (4.50) dans l'expression (4.10), ce qui donne

$$S(q) = \sum_{i>0} M_{i0}^2 \frac{\lambda_0^2 - \lambda_i^2}{\lambda_0^2 + \lambda_i^2 - 2\lambda_0\lambda_i \cos(qa)}. \quad (4.53)$$

4.2 Propriétés critiques de trois modèles d'ADN

Comme je l'ai noté dans la section 4.1, les lois d'échelle ont été déduites et testées pour des systèmes ayant un *paramètre d'ordre fini* au point critique, la plupart d'entre eux ayant des états ordonné et désordonné tous deux liés. Nous avons également vu que la finitude du paramètre d'ordre est importante pour la théorie des transitions de phase, puisque les modèles phénoménologiques comme celui de Landau-Ginsburg sont établis en développant la densité d'énergie au voisinage de T_c en fonction des puissances du paramètre d'ordre, qui est donc considéré comme petit et qui s'annule comme $|T - T_c|^\beta$

où $\beta > 0$. De plus, ces théories supposent que la longueur de corrélation ξ est la seule quantité ayant une contribution significative à l'énergie libre, puisque ξ constitue dans ces modèles la seule longueur interne qui diverge à T_c . Cette hypothèse peut être raisonnable pour les systèmes qui ne se dissocient pas à T_c . En revanche, elle est beaucoup moins évidente pour les systèmes dissociatifs qui ont au moins une longueur caractéristique divergente en plus de la longueur de corrélation. L'existence d'un paramètre d'ordre infini et/ou d'une longueur divergente supplémentaire peut donc faire douter de la validité de ces théories et par conséquent de celle des lois d'échelle.

L'ADN constitue précisément un cas exemplaire pour tester les lois d'échelle, puisque

- il possède un paramètre d'ordre qui devient infini à T_c ,
- il s'agit d'un système dissociatif qui, en plus de ξ , possède une longueur interne, l'écart moyen $\langle y \rangle$ entre les paires de bases, qui diverge à T_c .

Dans cette section, nous allons étudier précisément les propriétés critiques de trois modèles d'ADN, à savoir le modèle PB harmonique (2.8) et les modèles anharmoniques DPB (2.9) et JB (2.11).

4.2.1 Modèle de Peyrard-Bishop harmonique

Dans les chapitres 1 et 2, j'ai précisé que la dénaturation de l'ADN correspond à une transition de phase abrupte. Or le modèle PB que nous allons étudier dans cette partie présente une transition de phase continue du second ordre [53]. Par conséquent, ce modèle a un intérêt biologique très limité. Cependant, il possède certains aspects importants des modèles DPB et JB plus réalistes (en particulier la divergence de ξ et du paramètre d'ordre à T_c) et, comme nous allons voir, il permet une solution analytique approximative de l'équation de l'intégrale de transfert (4.40). Pour ces raisons, l'étude de ce modèle est instructif. Les résultats de cette section sont empruntés à la réf. [53], qui utilise $f(y_i) = y_i$ dans l'équation (4.36).

La résolution approximative de l'équation (4.40), dont je vais uniquement présenter les grandes lignes, s'effectue dans l'approximation des milieux continus définie par une

contrainte sur les paramètres du système (2.8), à savoir $Da^2/K \gg 1$ [53]. Dans ce régime, le facteur $e^{-K(x-y)^2}$ de l'intégrande de l'éq. (4.40) fait en sorte que la fonction $\phi_n(y)$ reste concentrée au voisinage de $y \simeq x$. Cela permet de développer $\phi_n(y)$ au voisinage de x ,

$$\phi_n(y) = \phi_n(x) + \phi'_n(x)(y-x) + \frac{1}{2}\phi''_n(x)(y-x)^2 + o((y-x)^3). \quad (4.54)$$

En injectant ce développement dans l'éq. (4.40) et en effectuant les intégrations gaussiennes qui en résultent, l'équation TI se réduit à une équation différentielle du second ordre,

$$\left[-\frac{1}{2\beta^2 K} \frac{d^2}{dy^2} + D(1 - e^{-ay})^2 \right] \phi_n(y) = e_n \phi_n(y), \quad (4.55)$$

où on a défini $e_n = \epsilon_n + \ln(2\pi/\beta K)/2\beta$ et $\lambda_n = e^{-\beta\epsilon_n}$. On reconnaît l'équation de Schrödinger d'un oscillateur de Morse de masse $M = K(\hbar\beta)^2$, dont on sait que l'état fondamental a pour énergie

$$\frac{e_0}{D} = 1 - \left(1 - \frac{ak_B T}{\sqrt{8DK}} \right)^2. \quad (4.56)$$

L'équation de l'intégrale de transfert a des solutions liées jusqu'à la dissociation de l'oscillateur de Morse à T_c . La limite $T = T_c$ correspond donc à la température au delà de laquelle le spectre discret de l'équation (4.55) disparaît complètement, c'est à dire $e_0(T_c) = D$. Cela permet d'estimer la température de dissociation sous la forme

$$T_c = \frac{2\sqrt{2KD}}{ak_B}. \quad (4.57)$$

La fonction propre de l'état fondamental s'écrit également

$$\phi_0(y) = \sqrt{\frac{a}{\Gamma(2\delta-1)}} (2\delta)^{\delta-1/2} e^{-\delta e^{-ay} - ay(\delta-1/2)}, \quad (4.58)$$

où j'ai noté, comme les auteurs de la réf. [53], $\delta = T_c/(2T)$. A partir des eqs. (4.48) et (4.56), on obtient la partie singulière de l'énergie libre,

$$f_{sing} = D(1 - |t|^2), \quad (4.59)$$

ce qui donne, après l'identification avec la relation (4.26), le premier exposant critique, à savoir

$$\alpha = 0. \quad (4.60)$$

En se basant sur le critère (4.25), on voit clairement que le modèle PB subit une transition de phase continue.

Pour la longueur de corrélation, on obtient, au voisinage de T_c ,

$$\xi^{-1} = \beta D |t|^2, \quad (4.61)$$

ce qui donne la valeur du deuxième exposant critique

$$\nu = 2. \quad (4.62)$$

L'exposant β défini dans l'éq. (4.16) s'obtient en calculant (4.49) avec la fonction propre (4.58). La partie singulière de ce calcul donne

$$m = \frac{1}{at}. \quad (4.63)$$

On trouve donc

$$\beta = -1. \quad (4.64)$$

La relation entre la susceptibilité (4.18) et le facteur de structure (4.20) peut être trouvée à partir de (4.6). Cela donne

$$\chi = S(0)/k_B T. \quad (4.65)$$

Par conséquent, le calcul de $S(q)$ nous conduit à travers les relations, (4.18), (4.20) et (4.65), aux deux exposants critiques γ et η . Dans ce calcul, dont je ne vais pas présenter les étapes puisque son intérêt est limité pour ce travail, les solutions de l'équation de Schrödinger (4.55) au-dessus du seuil interviennent. Après avoir fait la sommation de l'équation (4.53), les auteurs de la réf. [53] trouvent

$$S(q) = \lambda^2 \xi G(q\xi). \quad (4.66)$$

où

$$\lambda \propto m \propto \sqrt{\langle y_k^2 \rangle} \propto |t|^{-1} \quad (4.67)$$

et

$$G(q) = \frac{2}{x^2} \left[1 - \frac{1}{\cosh(\operatorname{arcsinh}(x))} \right], \quad (4.68)$$

ce qui donne

$$\chi \propto |t|^{-4} \quad (4.69)$$

$$S(q, T_c) \propto \frac{1}{q^{2-\eta}}. \quad (4.70)$$

Par conséquent, on obtient

$$\gamma = 4 \quad (4.71)$$

$$\eta = 0. \quad (4.72)$$

Dans la même référence, il est noté que ces exposants critiques satisfont les lois d'échelle de Rushbrooke (4.27), Fisher (4.35) et Josephson (4.32), ce que l'on peut vérifier facile-

ment (comme l'exposant critique δ décrivant le comportement de $m(t=0, h)$ n'a pas été calculé, la validité de la loi de Widom n'a pas été testée). J'apporterai un petit bémol à la vérification de la loi de Fisher dans la partie suivante.

4.2.2 Modèles anharmoniques DPB et JB

Pour les modèles anharmoniques DPB (2.9) et JB (2.11), la résolution analytique, même approximative, de l'éq. TI (4.40) est impossible. Je l'ai donc résolue numériquement afin d'étudier la transition de phase de ces deux modèles.

Les paramètres du modèle DPB sont ceux de la réf. [43], à savoir $D = 0.03$ eV, $a = 4.5 \text{ \AA}^{-1}$, $K = 0.06 \text{ eV \AA}^{-2}$, $\rho = 1$ et $\alpha = 0.35 \text{ \AA}^{-1}$ et ceux du modèle JB sont $D = 0.04$ eV, $a = 4.45 \text{ \AA}^{-1}$, $\Delta H = 0.44 \text{ eV}$, $b = 0.10 \text{ \AA}^{-2}$ et $K_b = 10^{-5} \text{ eV \AA}^{-2}$ (on rappelle que $1 \text{ eV} \approx 96.5 \text{ kJ/mol}$).

Afin de tester la validité des lois d'échelle sur un plus grand nombre de systèmes, je supposerai dans la suite de cette section que la fonction $f(y_i)$ dans les équations (4.2) et (4.36)-(4.37) est définie soit par $f(y_i) = y_i$, soit par $f(y_i) = y_i^2$. D'après l'équation (4.6), cela conduit à des paramètres d'ordre $m = \langle y_i \rangle$ et $m = \langle y_i^2 \rangle$, respectivement. La résolution numérique de l'éq. (4.40) visant à trouver les valeurs et vecteurs propres de la décomposition (4.39) consiste à diagonaliser le noyau (4.37) sur une grille discrète. Puisque l'on a un système dont la dynamique est décrite par des degrés de liberté continus et pour lequel le nombre de configurations accessibles est infini, la taille de cette matrice, dont chaque élément correspond à un état configurationnel, devrait être également infinie. Comme il est impossible de diagonaliser une telle matrice, l'approximation habituelle consiste à la couper au bon endroit et choisir ses éléments de façon optimale. J'y reviendrai plus tard.

Après la diagonalisation, les valeurs et vecteurs propres obtenus sont utilisés pour calculer les observables thermodynamiques à partir des relations (4.48)-(4.53). Les quantités faisant intervenir des dérivées par rapport à la température T ou au champ h , comme l'entropie par particule $s = S/N$ et la chaleur spécifique par particule $c_V = C_V/N$ ou

la susceptibilité (4.9), ont été calculées à partir de formules de différences finies. Plus précisément, ces trois quantités ont été obtenues à partir des formules

$$s(T) = -\frac{f(T + \epsilon_T) - f(T - \epsilon_T)}{2\epsilon_T} \quad (4.73)$$

$$c_V(T) = -T \frac{f(T + \epsilon_T) + f(T - \epsilon_T) - 2f(T)}{\epsilon_T^2} \quad (4.74)$$

$$\chi(T) = \frac{m(h + \epsilon_h) - m(h - \epsilon_h)}{2\epsilon_h} \quad (4.75)$$

où j'ai pris $\epsilon_T = 10^{-4}$ K et $\epsilon_h = 10^{-12}$ eV Å⁻¹ pour $f(y_i) = y_i$ ou $\epsilon_h = 10^{-15}$ eV Å⁻² pour $f(y_i) = y_i^2$.

Les exposants critiques ont été calculés en traçant les quantités thermodynamiques des équations (4.15)-(4.20) sur une échelle log-log et en mesurant la pente des droites sur la plage de températures où ces lois de puissance sont satisfaites. Ces lois étant attendues au voisinage de la température critique, la région de validité de chaque loi de puissance est bien sûr limitée lorsque l'on s'éloigne de T_c vers les basses températures. Il existe également un artefact numérique qui masque ces lois de puissance lorsque l'on est trop près de T_c . En effet, on étudie un système dissociatif qui, près de la température de dissociation, possède des états configurationnels presque libres, correspondant aux grandes valeurs de y_i dans l'expression (4.3), dont la contribution à la fonction de partition n'est pas négligeable. Tronquer la matrice TI revient précisément à négliger les états correspondant aux valeurs de y_i supérieures à une certaine limite. Cela signifie que la contrainte de finitude de la matrice TI limite la précision numérique de la méthode TI lorsque l'on se trouve trop près de T_c . En d'autres termes, il est obligatoire de prendre des matrices de plus en plus grandes pour élargir la région sur laquelle les lois de puissances sont valables et augmenter la précision avec laquelle les valeurs des exposants critiques sont estimées. Les auteurs de la réf [43] ont déjà calculé quelques exposants critiques pour le modèle DPB anharmonique et ont remarqué que la divergence de certaines observables thermo-

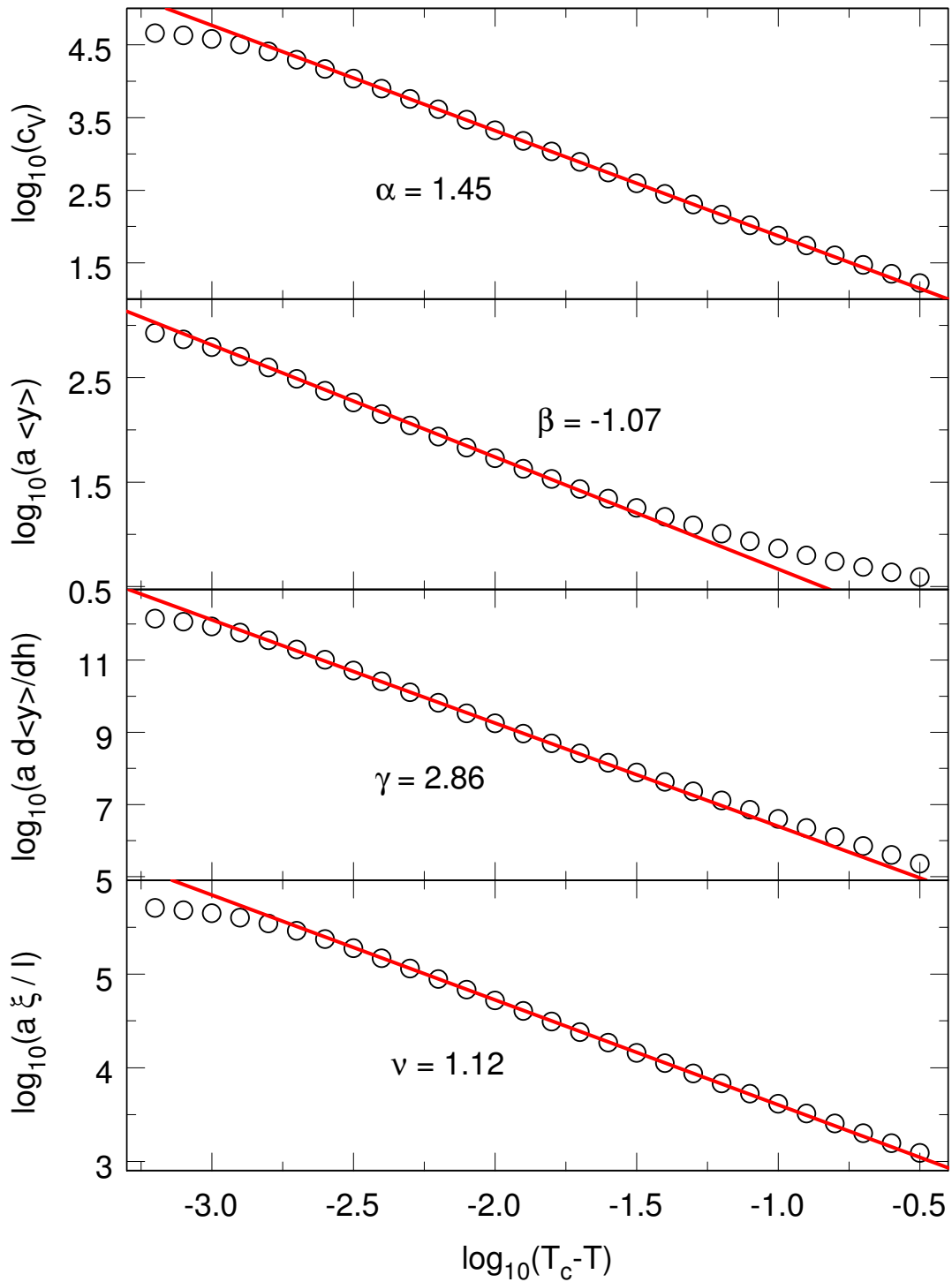


Fig. 4.1: Courbes log-log permettant de déterminer les exposants critiques α , β , γ , ν pour le modèle DPB, avec un couplage au champ $f(y_k) = y_k$.

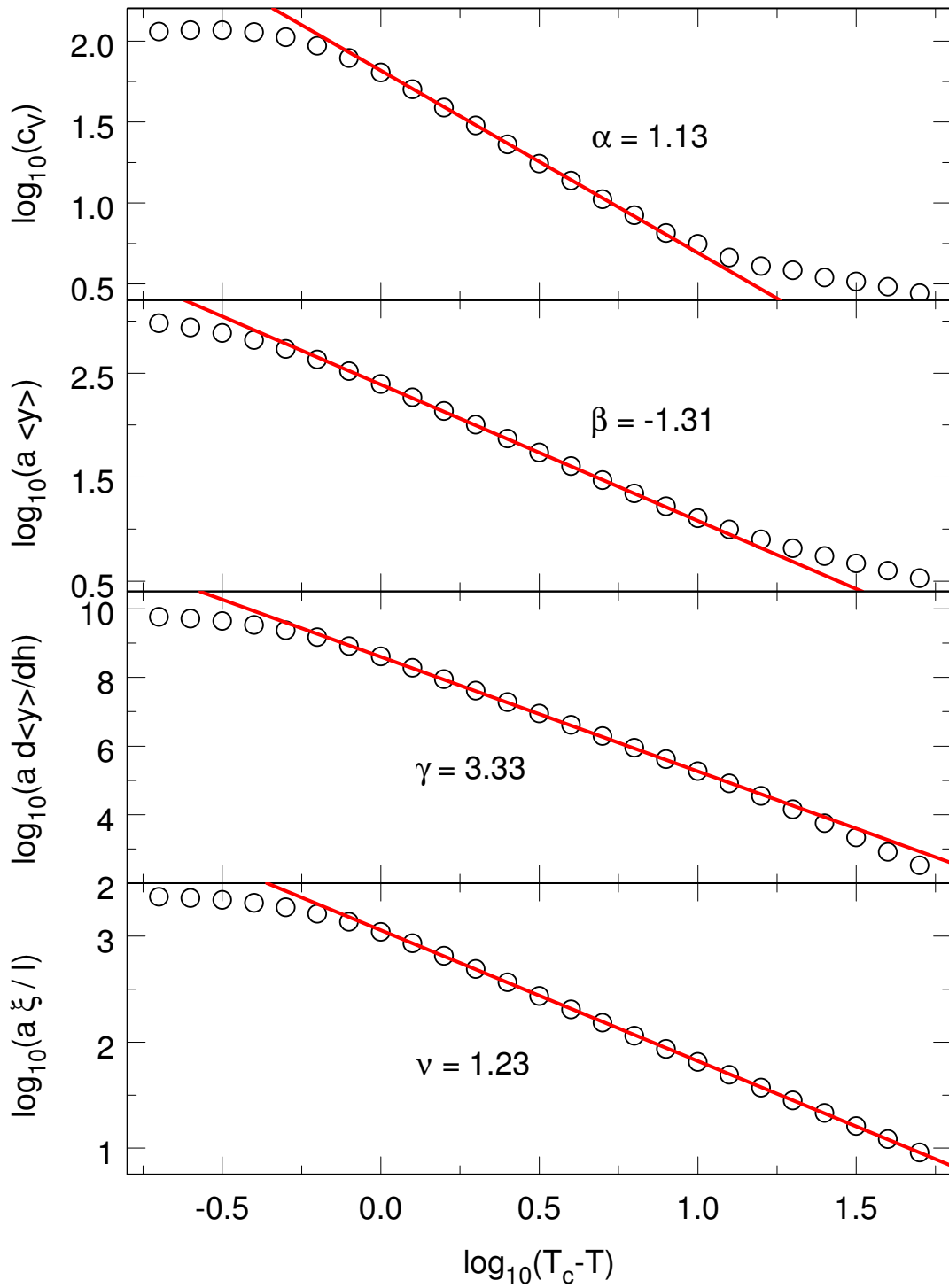


Fig. 4.2: Courbes log-log permettant de déterminer les exposants critiques α , β , γ , ν pour le modèle JB, avec un couplage au champ $f(y_k) = y_k$.

dynamiques, qui sont censées tendre vers l'infini de façon abrupte pour ce modèle, était masquée au voisinage immédiat de T_c par un effet dont ils n'ont pas précisé la nature. Cela les a conduits à donner des estimations arrondies des exposants, qui diffèrent parfois des valeurs exactes d'un facteur 2. Nos calculs ont montré que la raison les empêchant d'avoir des divergences correctes est l'effet purement numérique de la méthode TI que je viens de mentionner. Afin d'obtenir la meilleure précision possible pour le calcul des exposants critiques, j'ai effectué la diagonalisation de l'opérateur de transfert (4.37) sur un réseau discret de taille 4201, avec la valeur du degré de liberté d'élongation qui variait soit de $y = -200/a$ à $y = 4000/a$ par pas de $1/a$, soit de $y = -200/a$ à $y = 5067/a$ avec des incréments croissant de façon exponentielle de $0.2/a$ à $4/a$ (notons que a désigne ici le paramètre du potentiel de Morse). On constate que les deux types de grille conduisent essentiellement aux mêmes résultats. En me basant sur différents essais numériques, j'ai conclu que nous sommes capables de calculer les exposants critiques avec une incertitude maximale de 5%. Pour les deux modèles en question, les températures critiques ont été prises au maximum de la longueur de corrélation ξ à champ h nul. Les températures critiques, déterminées par le calcul TI sur le réseau exponentiel de longueur 4201, sont $T_c = 280.29$ K pour le modèle DPB et $T_c = 368.15$ K pour le modèle JB.

Comme je l'ai indiqué un peu plus haut, les valeurs des exposants critiques ont été obtenues en traçant les observables thermodynamiques sur une échelle log-log et en mesurant la pente des droites sur la région où la loi de puissance est valable pour chaque quantité. Une partie des résultats, correspondant aux exposants α , β , γ et ν , est illustrée sur les figures 4.1 et 4.2, respectivement pour le modèle DPB et JB, et la contrainte externe $f(y_k) = y_k$. Le calcul des exposants δ (4.17) et η (4.20) a été effectué de la même façon, à la seule différence que les axes des abscisses correspondent dans ce cas respectivement aux variables h et q alors que la température est fixée à T_c . Comme je l'ai déjà précisé, les exposants ont été calculés pour chaque modèle avec deux contraintes externes. Les valeurs de β , γ et δ dépendent explicitement du choix de cette contrainte, tandis que α et ν ne dépendent pas de ce choix, ce qui est normal puisque ces

exposants sont calculés à champ nul. Les valeurs des exposants critiques pour les deux modèles et les deux contraintes sont rapportées dans le tableau 4.1. A cette étape de la discussion, il est important de noter que l'exposant critique de la chaleur spécifique, α , est significativement supérieur à 1 pour les modèles DPB et JB. Cela montre que le processus de dénaturation des deux modèles correspond à une transition de phase du *premier ordre*. On remarque ici une particularité très importante de la dénaturation de l'ADN. Au début de ce chapitre, nous avons vu que les transitions de phase caractérisées par une longueur de corrélation divergente sont en général celles du second ordre, tandis que, dans le cas de la dissociation thermique de l'ADN, il s'agit d'une transition de phase du premier ordre ayant une longueur de corrélation divergente. Cette remarque a déjà été faite dans les travaux concernant le modèle de Poland-Scheraga [20]. On peut finalement noter que comme le paramètre d'ordre m diverge au point critique ($t = 0, h = 0$), les exposants critiques β et δ sont négatifs, alors qu'ils sont positifs pour la grande majorité des systèmes.

A partir de ces exposants calculés, j'ai vérifié les lois d'échelle, ce qui est illustré dans le tableau 4.2. Comme je l'ai déjà noté, les deux premières lois d'échelle, à savoir les lois de Rushbrooke (4.27) et Widom (4.28), sont déduites de l'hypothèse d'homogénéité (4.26). Pour cette raison, j'ai d'abord testé cette hypothèse pour les modèles DPB et JB. Sur la figure 5.1, j'ai tracé, pour trois valeurs du champ externe variant entre $h = 10^{-6} D$ et $h = 10^{-4} D$, l'énergie libre réduite $f_{sing}/(T_c - T)^{2-\alpha}$ en fonction du logarithme de la variable $h/(T - T_c)^{\beta\delta}$. La partie singulière de l'énergie libre est calculée à partir de la relation

$$f_{sing}(T, h) = f_{sing}(T, h) - f_{sing}(T_c, 0). \quad (4.76)$$

Le fait que les points correspondant aux différentes valeurs de h sont situés sur la même ligne montre que la loi d'homogénéité est satisfaite. Par conséquent, il est naturel que les lois de Rushbrooke (4.27) et Widom (4.28) soient vérifiées pour les deux modèles en question. Cela est illustré plus explicitement dans le tableau 4.2 qui montre, pour les

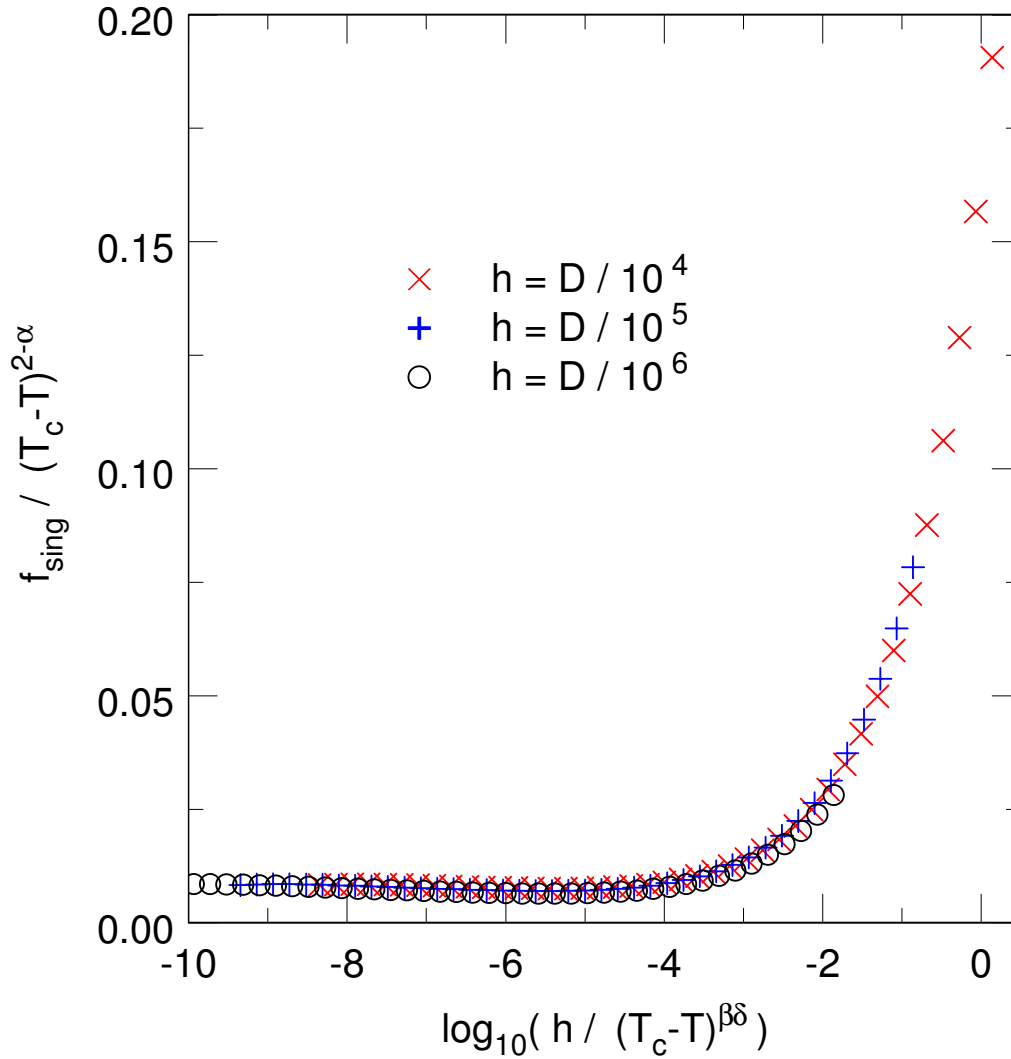


Fig. 4.3: Tracé de $f_{\text{sing}}/(T_c - T)^{2-\alpha}$ en fonction du logarithme de $h/(T_c - T)^{\beta\delta}$ pour le modèle JB, avec un couplage au champ $f(y_k) = y_k$. La courbe est tracée pour trois valeurs de h variant de $10^{-4} D$ à $10^{-6} D$. f_{sing} et h sont exprimés en unité de D . La superposition des courbes correspondant aux différentes valeurs de h montre la validité de l'hypothèse d'homogénéité de l'équation (4.26) pour le modèle JB.

polynômes $\alpha+2\beta+\gamma$ et $\gamma-\beta(\delta-1)$, les valeurs prédites par les lois d'échelles (colonne 2) et celles obtenues à partir des exposants critiques calculés pour les deux modèles (colonnes 3 à 6). Le même tableau montre l'estimation des incertitudes obtenue en supposant des erreurs additives de 5% sur chacun des exposants. On voit que les valeurs prédites par les lois de Rushbrooke et de Widom se situent clairement à l'intérieur de la région d'incertitude.

Nous avons également vu que la loi de Josephson (4.32) est déduite en supposant que la condition (4.31) est satisfaite, autrement dit dans le cas où la longueur de corrélation ξ est la seule longueur interne ayant une contribution significative à la singularité de l'énergie libre. En fait, dans la littérature, on peut rencontrer une version plus générale de cette loi, à savoir l'inégalité de Josephson [54, 55] qui s'écrit

$$\alpha + \nu d \geq 2. \quad (4.77)$$

Cette inégalité n'est pas basée sur l'hypothèse d'homogénéité généralisée et doit toujours être vérifiée. Il est intéressant de noter que selon le tableau 4.2, les exposants critiques calculés satisfont l'inégalité (4.77) mais pas la loi de Josephson (4.32). Le même tableau montre que pour les deux modèles la différence entre la valeur de $\alpha + \nu d$ calculée par les exposants critiques et celle prédite par la loi de Josephson (2), est supérieure à trois fois l'incertitude estimée. Cela montre que, contrairement à de nombreux systèmes subissant une transition de phase, l'hypothèse d'homogénéité généralisée n'est pas satisfaite pour la dénaturation de l'ADN décrite par ces deux modèles réalistes. Comme je l'ai déjà noté, cela est prévisible pour les systèmes dissociatifs, puisque ces systèmes possèdent en plus de ξ une longueur divergente supplémentaire, à savoir le degré de liberté dissociatif (dont la moyenne thermique est le paramètre d'ordre pour l'ADN). Dans ce cas, il devient clair que l'hypothèse selon laquelle les observables thermodynamiques sont uniquement fonctions du rapport de ξ sur la taille L du système n'est plus justifiée. En utilisant les données du tableau 4.2, j'ai essayé, sans succès, de trouver une fonction qui pourrait remplacer ξ (cette quantité devrait avoir la dimension d'une longueur et un exposant

caractéristique égal à $2 - \alpha$). Le fait que la loi de Josephson soit vérifiée pour le modèle PB harmonique (section 4.2.1) est probablement dû au fait que, dans ce cas précis, la longueur de corrélation d'exposant caractéristique $\nu = 2$ diverge beaucoup plus vite que le paramètre d'ordre d'exposant $\beta = -1$, ce qui peut permettre que ξ soit effectivement la longueur caractéristique ayant la plus grande contribution à la singularité des observables thermodynamiques. Il est bien sûr nécessaire de tester cette hypothèse sur d'autres systèmes ayant des caractéristiques similaires avant de conclure définitivement sur ce point.

Finalement, le tableau 4.2 montre que la loi de Fisher (4.35) est loin d'être satisfaite. Ceci est dû au fait que cette loi est obtenue en supposant une décroissance en loi de puissance de la fonction de corrélation (4.33), ce qui est correct uniquement pour $d \geq 2$, ainsi que je l'ai précisé dans la partie consacrée aux lois d'échelle de ce chapitre. En fait, à une dimension, l'éq. (4.13) montre que la fonction de corrélation obéit à une loi exponentielle qui devient constante à T_c . Autrement dit, pour $d = 1$ la fonction G_{ij} au point critique ne dépend plus de la distance $|k - l|$, ce que l'on peut voir explicitement à travers les équations (4.50) et (4.51). En tenant compte de ce fait et en calculant de nouveau l'intégrale (4.34) avec $G(x) \sim \langle y^2 \rangle - \langle y \rangle^2 \sim \langle y^2 \rangle$, on trouve une nouvelle loi polynomiale,

$$\gamma = \mu + \nu, \quad (4.78)$$

où on a défini un nouvel exposant critique μ qui caractérise la divergence de la moyenne thermique du carré de la contrainte externe,

$$\langle f(y)^2 \rangle \propto t^{-\mu}. \quad (4.79)$$

Le tableau 4.2 montre que cette nouvelle loi, alternative à la loi de Fisher, est très bien vérifiée.

	modèle DPB		modèle JB	
	$f(y) = y$	$f(y) = y^2$	$f(y) = y$	$f(y) = y^2$
α	1.45	1.45	1.13	1.13
β	-1.07	-1.72	-1.31	-2.11
γ	2.86	4.00	3.33	4.82
δ	-1.66	-1.39	-1.58	-1.35
η	0.01	0.01	0.02	0.02
ν	1.12	1.12	1.23	1.23
μ	1.72	2.98	2.11	3.52

Tab. 4.1: Exposants critiques $\alpha, \beta, \gamma, \delta, \eta$ et ν pour les modèles DPB et JB, avec des contraintes externes $f(y) = y$ et $f(y) = y^2$. L'exposant μ caractérise la divergence de $\langle f(y)^2 \rangle$ au voisinage de T_c .

	Prédiction	modèle DPB		modèle JB		
		$f(y) = y$	$f(y) = y^2$	$f(y) = y$	$f(y) = y^2$	
Rushbrooke :	$\alpha + 2\beta + \gamma$	2	2.17 ± 0.32	2.01 ± 0.44	1.84 ± 0.35	1.73 ± 0.51
Widom :	$\gamma - \beta(\delta - 1)$	0	0.01 ± 0.37	-0.11 ± 0.53	-0.05 ± 0.44	-0.14 ± 0.63
Josephson :	$\alpha + \nu d$	2	2.57 ± 0.13	2.57 ± 0.13	2.36 ± 0.12	2.36 ± 0.12
Fisher :	$\gamma - \nu(2 - \eta)$	0	0.63 ± 0.25	1.77 ± 0.31	0.89 ± 0.29	2.86 ± 0.36
	$\gamma - (\mu + \nu)$	0	0.02 ± 0.28	-0.10 ± 0.40	-0.01 ± 0.33	0.07 ± 0.48

Tab. 4.2: Valeurs de $\alpha + 2\beta + \gamma$, $\gamma - \beta(\delta - 1)$, $\alpha + \nu d$, $\gamma - \nu(2 - \eta)$ et $\gamma - (\mu + \nu)$, prédites par les lois d'échelle et obtenues à partir des exposants critiques présentés dans le tableau 4.1 pour les modèles DPB et JB et des contraintes externes $f(y) = y$ et $f(y) = y^2$. Les plages de variation ont été obtenues en supposant des incertitudes additives de 5% sur chacun des exposants critiques rapportés dans le tableau 4.1.

Avant de conclure cette section, je voudrais faire une remarque. Comme je l'ai déjà noté, les auteurs de la réf. [53] précisent à la page 888 de cet article que les exposants critiques (4.60), (4.62), (4.64), (4.71), (4.72) satisfont la loi de Fisher (4.35), c'est à dire $\gamma = (2 - \eta)\nu$. D'autre part, la relation (4.66) qu'ils ont obtenue indique une loi différente, à savoir $\gamma = -2\beta + \nu$ ou $\gamma = \mu + \nu$. En fait, puisque pour le modèle PB on a $\mu = -2\beta$, il est impossible de distinguer entre ces deux lois. En d'autres termes, la solution analytique du modèle PB conduit également à une loi d'échelle de la forme (4.78) au lieu de la loi de Fisher (4.35). Le fait que la loi de Fisher soit vérifiée uniquement par un modèle sur les trois montre que cela est vraisemblablement simplement une coïncidence, ce que l'on peut voir clairement en notant que $(2 - \eta)\nu = \gamma = \mu + \nu$ pour le modèle PB.

4.3 Conclusion

Dans ce chapitre, nous nous sommes familiarisés avec les notions de base des phénomènes critiques. Nous nous sommes servis de ces acquis pour étudier les caractéristiques de la transition de dénaturation de trois modèles d'ADN. Nous avons calculé les exposants critiques et vérifié les lois d'échelle. Ma motivation pour cette étude provenait de l'existence de certaines propriétés de l'ADN qui sont inhabituelles pour un système subissant une transition de phase, à savoir une dimension effective inférieure à 2, une transition de phase discontinue avec une longueur de corrélation infinie et un paramètre d'ordre également divergent à T_c . Nous avons montré que l'hypothèse d'homogénéité est respectée pour ce système malgré la divergence du paramètre d'ordre. Cela conduit évidemment à la validité des lois de Rushbrooke et Widom, ce que l'on a vérifié explicitement. Nous avons remarqué que la loi de Josephson n'est pas satisfaite, ce qui résulte probablement de la divergence du paramètre d'ordre, puisque cela met en doute l'hypothèse de hyperscaling selon laquelle la longueur de corrélation est la seule longueur interne responsable de la singularité des observables thermodynamiques au point critique. Finalement, nous avons noté que la loi de Fisher est également loin d'être satisfaite pour le système concerné. Nous avons montré que cela est dû à la dimensionnalité

particulière ($d = 1$) de ce système, et obtenu une loi alternative très bien vérifiée. Bien que des études supplémentaires soient nécessaires pour vérifier nos conclusions, ce travail montre que les lois d'échelle doivent être considérées avec précaution pour des systèmes possédant des caractéristiques inhabituelles.

5. INFLUENCE DE LA LONGUEUR DE LA SÉQUENCE SUR LA TRANSITION DE PHASE

Dans le chapitre 4, nous avons vu que la singularité des fonctions thermodynamiques caractérisant un point critique provient de la singularité de l'énergie libre qui n'est pas une fonction analytique de la température. En fait, cette non-analyticité de l'énergie libre résulte des interactions entre le très grand nombre de particules constituant le système. Par conséquent, les systèmes de petite taille ne subissent pas de transition de phase. Dans ces systèmes, la divergence des observables à la limite thermodynamique est remplacée par un pic de hauteur finie s'étendant sur une plage de température appelée *rounding region*. D'autre part, la limite thermodynamique est une idéalisation théorique, puisque dans la vie réelle et notamment les expériences, les systèmes subissant une transition de phase ont forcément un volume fini. Il est donc important de suivre l'évolution des singularités thermodynamiques en fonction de la taille du système et de comprendre à quelle taille on a le droit de supposer que la limite thermodynamique est atteinte pour le système en question.

La transition de dénaturation de l'ADN est un processus pour lequel la question des effets de taille sur la transition se pose naturellement puisque des séquences d'ADN de tailles très différentes, de quelques paires de bases à des dizaines de milliers, peuvent être synthétisées à la demande pour des études expérimentales. L'étude théorique des effets de taille sur des séquences d'ADN homogènes, c'est à dire dont chaque brin est constitué d'une séquence $NNN \dots$ uniforme, s'avère particulièrement fructueuse, car une séquence homogène subit une seule transition de phase dont il est facile d'étudier les caractéristiques. Par contre, pour une séquence d'ADN hétérogène, ces caractéristiques sont en partie masquées par le contenu génétique de la séquence en question qui, tout

comme la longueur de la séquence, influe sur la température de dissociation et le caractère plus ou moins abrupt de la dénaturation. Je reviendrai dans le chapitre 6 sur l'influence de l'hétérogénéité de la séquence sur la dénaturation.

Il a été établi depuis longtemps que l'élargissement de la région de transition résultant de la réduction de la longueur de la séquence est accompagné par la diminution de la température de dénaturation (voir par exemple la figure 5 de la réf [56]). Un comportement similaire a depuis été observé pour des systèmes dissociatifs plus complexes, comme les polylactides [57] et les hybrides d'ADN et d'oxy-peptides [58]. Les formules semi-empiriques utilisées par les calculateurs en ligne dédiés à la prédiction des propriétés des oligonucléotides supposent d'ailleurs en général que la température de dissociation d'une séquence d'ADN homogène diminue en $820/N$ si $13 < N < 50$ et en $500/N$ pour $N > 50$. Bien que des dépendances légèrement différentes aient parfois été ajustées sur les résultats des expériences [59], la forme en $1/N$ semble faire l'unanimité pour les grandes valeurs de N .

Les raisons pour lesquelles l'évolution de la température de dissociation en fonction de la taille de la séquence a été étudiée de façon aussi approfondie sont certainement les suivantes : (i) la connaissance de cette température est d'une importance majeure pour les expériences, (ii) la température de dissociation peut être localisée facilement par les techniques couramment utilisées pour étudier la dénaturation de l'ADN, comme par exemple, la spectroscopie d'absorption UV à 260 nm [2]-[4], le dichroïsme circulaire, la calorimétrie différentielle [60, 61], la spectroscopie Raman [61] ou les tests de mobilité électrophorétiques [62, 63]. De plus, les expérimentateurs ont également étudié comment les observables thermodynamiques dépendent de la longueur de la séquence grâce à la calorimétrie différentielle et la spectroscopie d'absorption UV (voir par exemple les réf. [60, 64, 65]) et analysé des aspects plus spécifiques comme la persistance des états intermédiaires (c'est à dire des états métastables correspondant aux bulles) jusqu'à des tailles de séquence très petites, à savoir $13 < N < 48$ [66, 67]. Je voudrais cependant préciser que la validité des modèles étudiés dans ce chapitre n'est pas prouvée pour des

séquences aussi courtes [68, 69].

Le but du travail présenté dans ce chapitre est de compléter les résultats expérimentaux mentionnés ci-dessus en proposant une description théorique de la dissociation des séquences homogènes d'ADN possédant un nombre fini N de paires de bases. Je discuterai bien sûr l'évolution de la température de dénaturation en fonction de N , mais j'analyserai aussi le comportement des fonctions thermodynamiques intensives comme l'entropie et la chaleur spécifique par particule, l'écart moyen des paires de bases $\langle y \rangle$ ou la longueur de corrélation ξ . Les résultats ont été obtenus pour les modèles DPB (2.9) et JB (2.11). Les paramètres utilisés dans ce travail sont ceux des réf. [9, 40], c'est à dire $D = 0.03$ eV, $a = 4.5 \text{ \AA}^{-1}$, $\alpha = 0.35 \text{ \AA}^{-1}$, $K = 0.06$ eV \AA^{-2} , $\rho = 1$ pour le modèle DPB et $D = 0.04$ eV, $a = 4.45 \text{ \AA}^{-1}$, $\Delta H = 0.44$ eV, $K_b = 10^{-5}$ eV \AA^{-2} et $b = 0.10 \text{ \AA}^{-2}$ pour le modèle JB (on rappelle que $1 \text{ eV} \approx 96.5 \text{ kJ/mol}$). $m = 300$ uma pour les deux modèles.

Je commencerai par présenter une version modifiée de la méthode TI, qui tient compte du fait qu'une séquence finie a des extrémités qui sont nécessairement libres. M'appuyant sur le travail de Zhang et al [15], j'ai établi les formules permettant d'obtenir l'écart moyen entre les bases appariées et la longueur de corrélation pour de telles séquences. Je discuterai ensuite en détail l'influence de N sur l'ensemble des variables intensives. Enfin, je décrirai comment on peut utiliser la version TI modifiée pour tester la théorie de *finite size scaling* développée par Fisher et Barber [70]. Cette théorie, qui a été établie afin d'étudier les effets de la finitude du volume sur les phénomènes critiques, a également été utilisée comme méthode alternative pour déterminer les exposants critiques. Parmi les dizaines de systèmes critiques étudiés dans le cadre de cette théorie, on peut citer par exemple les modèles d'Ising à cinq [71, 72] et six dimensions [73] et les modèles de percolation [74, 75].

5.1 Méthode TI pour les séquences finies aux extrémités libres

Dans cette section, je vais présenter une version modifiée de la méthode TI, dont la version standard a été décrite dans le chapitre 4. Cette nouvelle version, initialement

proposée par Zhang et ses collaborateurs [15], tient compte du fait que les deux extrémités de la chaîne d'ADN sont libres, contrairement au choix du chapitre 4 où les deux bords de la séquence étaient liés. Dans ce cas, la fonction de partition s'écrit (comparer à l'expression (4.38)) :

$$Z = \int dy_1 dy_2 \cdots dy_N e^{-\beta V_M(y_1)/2} K(y_2, y_1) K(y_3, y_2) \cdots K(y_N, y_{N-1}) e^{-\beta V_M(y_N)/2}. \quad (5.1)$$

Pour calculer cette expression, définissons, comme dans la réf [15], le vecteur suivant,

$$a_i = \int dy e^{-\beta V_M(y)/2} \Phi_i(y). \quad (5.2)$$

En substituant le développement (4.39) du noyau défini par (4.37) dans l'expression (5.1), on obtient pour la fonction de partition

$$Z = \sum_i a_i^2 \lambda_i^{N-1}. \quad (5.3)$$

Comme je l'ai expliqué dans le chapitre 4, ce résultat permet de calculer, à partir des éqs. (4.4), (4.7) et (4.8), les fonctions thermodynamiques extensives, plus précisément l'énergie libre, l'entropie et la chaleur spécifique.

Le calcul des quantités intensives, comme l'écart moyen des paires de bases, défini par l'expression

$$\langle y \rangle = \frac{1}{N} \sum_{n=1}^N \langle y_n \rangle, \quad (5.4)$$

ou la longueur de corrélation de l'éq. (4.11) est plus complexe. Pour calculer $\langle y \rangle$, on doit d'abord réduire l'expression de l'écart moyen d'une paire de bases,

$$\langle y_n \rangle = \frac{1}{Z} \int dy_1 dy_2 \cdots dy_N y_n e^{-\beta V_M(y_1)/2} K(y_2, y_1) K(y_3, y_2) \cdots K(y_N, y_{N-1}) e^{-\beta V_M(y_N)/2} \quad (5.5)$$

qui est l'équivalent de l'éq. (4.43) pour une chaîne fermée, à une forme dépendant uniquement des vecteurs et valeurs propres de l'équation TI (4.40). En injectant le développement (4.39) dans l'éq. (5.5) et en définissant

$$\begin{aligned} b_i &= \int dy e^{-\beta V_M(y)/2} \Phi_i(y) y \\ Y_{ij}^{(1)} &= \int dy \Phi_i(y) y \Phi_j(y), \end{aligned} \quad (5.6)$$

on obtient

$$\langle y_1 \rangle = \langle y_N \rangle = \frac{1}{Z} \sum_i a_i b_i \lambda_i^{N-1} \quad (5.7)$$

et

$$\langle y_n \rangle = \frac{1}{Z} \sum_{i,j} a_i a_j Y_{ij}^{(1)} \lambda_i^{n-1} \lambda_j^{N-n} \quad (5.8)$$

pour $n \neq 1, N$. En substituant les éqs. (5.7), (5.8) dans l'expression (5.4) et en effectuant les sommations géométriques qui en résultent, on trouve finalement

$$\langle y \rangle = \frac{2}{ZN} \sum_i a_i b_i \lambda_i^{N-1} + \frac{1}{ZN} \sum_{i,j} a_i a_j Y_{ij}^{(1)} \lambda_i^{-1} \lambda_j^N \frac{r_{ij}^2 - r_{ij}^N}{1 - r_{ij}} \quad (5.9)$$

où $r_{ij} = \lambda_i / \lambda_j$.

Le calcul de la longueur de corrélation, définie par l'expression (4.11), est plus fastidieux. Pour cela, j'ai dû exprimer le facteur de structure (4.10) sous une forme pratique pour une évaluation numérique. Afin de séparer explicitement la contribution des paires de bases aux extrémités, j'ai d'abord écrit l'éq. (4.10) sous la forme

$$S(q) = \sum_{n=1}^N \sum_{m=1}^N \langle \delta y_n \delta y_m \rangle e^{iqa(n-m)}, \quad (5.10)$$

où $\delta y_n = y_n - \langle y_n \rangle$. En isolant les valeurs moyennes des extrémités, on obtient

$$\begin{aligned}
S(q) &= \langle \delta y_1^2 \rangle + \langle \delta y_N^2 \rangle + 2 \langle \delta y_1 \delta y_N \rangle \cos[qa(N-1)] \\
&+ S_1 e^{iqa} + S_1^* e^{-iqa} + S_N e^{iNqa} + S_N^* e^{-iNqa} + S_{mid},
\end{aligned} \tag{5.11}$$

où

$$\begin{aligned}
S_1 &= \sum_{m=2}^{N-1} \langle \delta y_1 \delta y_m \rangle e^{-iqam} \\
S_N &= \sum_{m=2}^{N-1} \langle \delta y_m \delta y_N \rangle e^{-iqam} \\
S_{mid} &= \sum_{m=2}^{N-1} \sum_{n=2}^{N-1} \langle \delta y_n \delta y_m \rangle e^{iqa(n-m)}.
\end{aligned} \tag{5.12}$$

En définissant

$$\begin{aligned}
c_i &= \int dy e^{-\beta V_M(y)/2} y^2 \Phi_i(y) \\
Y_{ij}^{(2)} &= \int dy \Phi_i(y) y^2 \Phi_j(y),
\end{aligned} \tag{5.13}$$

on obtient ensuite les relations suivantes :

$$\begin{aligned}
\langle y_1^2 \rangle &= \langle y_N^2 \rangle = \frac{1}{Z} \sum_i a_i c_i \lambda_i^{N-1} \\
\langle y_1 y_N \rangle &= \frac{1}{Z} \sum_i b_i^2 \lambda_i^{N-1} \\
\langle y_1 y_m \rangle &= \frac{1}{Z} \sum_{ij} a_j b_i \lambda_i^{m-1} \lambda_j^{N-m} Y_{ij}^{(1)}
\end{aligned}$$

$$\begin{aligned}
\langle y_m y_N \rangle &= \frac{1}{Z} \sum_{ij} a_i b_j \lambda_i^{m-1} \lambda_j^{N-m} Y_{ij}^{(1)} \\
\langle y_n^2 \rangle &= \frac{1}{Z} \sum_{ij} a_i a_j \lambda_i^{n-1} \lambda_j^{N-n} Y_{ij}^{(2)} \\
\langle y_n y_m \rangle &= \frac{1}{Z} \sum_{ijk} a_i a_k Y_{ij}^{(1)} Y_{jk}^{(1)} \cdot \begin{cases} \lambda_i^{n-1} \lambda_j^{m-n} \lambda_k^{N-m} & \text{si } m > n \\ \lambda_i^{m-1} \lambda_j^{n-m} \lambda_k^{N-n} & \text{si } m < n . \end{cases}
\end{aligned} \tag{5.14}$$

En remplaçant ces expressions dans les éqs. (5.10), (5.11) et en effectuant les sommations géométriques qui en résultent, on trouve pour le facteur de structure $S(q)$

$$\begin{aligned}
S(q) &= \langle \delta y_1^2 \rangle + \langle \delta y_N^2 \rangle + 2 \langle \delta y_1 \delta y_N \rangle \cos[qa(N-1)] + \sum_{ij} H_{ij} \frac{f_{ij}(2) - f_{ij}(N)}{1 - f_{ij}(1)} \\
&+ \sum_{i,j} \frac{1}{\cosh(\alpha_{ij}) - \cos(qa)} \{ [D_{ij} f_{ij}(1) + C_{ij} f_{ij}(N)] \cos[(N-2)qa] - D_{ij} f_{ij}(N-1) - C_{ij} f_{ij}(2) \\
&- [D_{ij} f_{ij}(2) + C_{ij} f_{ij}(N-1)] \cos[(N-1)qa] + [D_{ij} f_{ij}(N) + C_{ij} f_{ij}(1)] \cos[qa] \} \\
&+ \sum_{ijk} \frac{2M_{ijk}}{(1-g(1,1))(1+f_{ij}(2)-2f_{ij}(1)\cos(qa))(1+f_{jk}(2)-2f_{jk}(1)\cos(qa))} \\
&\times \{ -g(3,3) - g(4,4) - g(2,4) + g(N,N) + g(N,N+2) + g(N+1,N+1) \\
&+ [g(2,3) - g(N,N+1) + 2g(3,4) - g(N+1,N+2) + g(4,3) - g(N+1,N) \\
&- g(N,N+1)] \cos(qa) \\
&+ [g(N+1,N+1) - g(3,3)] \cos(2qa) + [g(2,N+1) - g(3,N+2)] \cos[(N-3)qa] \\
&+ [g(4,N+2) - g(2,N)] \cos[(N-2)qa] + [g(3,N) - g(4,N+1)] \cos[(N-1)qa] \} \\
&+ \left| \sum_{ij} M_{ij} \frac{e^{2(ika-\alpha_{i,j})} - e^{N(ika-\alpha_{i,j})}}{1 - e^{ika-\alpha_{i,j}}} \right|^2,
\end{aligned} \tag{5.15}$$

avec

$$\begin{aligned}
\alpha_{ij} &= -\ln(r_{ij}) \\
M_{ij} &= \frac{1}{Z} \lambda_i^{-1} \lambda_j^N a_i a_j Y_{ij}^{(1)} \\
H_{ij} &= \frac{1}{Z} \lambda_i^{-1} \lambda_j^N a_i a_j Y_{ij}^{(2)} \\
T_{ij} &= \frac{1}{Z} \lambda_i^{-1} \lambda_j^N b_i a_j Y_{ij}^{(1)} \\
G_{ij} &= \frac{1}{Z} \lambda_i^{-1} \lambda_j^N a_i b_j Y_{ij}^{(1)} \\
C_{ij} &= T_{ij} - \langle y_1 \rangle M_{ij} \\
D_{ij} &= G_{ij} - \langle y_1 \rangle M_{ij} \\
f_{ij}(n) &= e^{-n\alpha_{ij}} \\
g(n, m) &= e^{-n\alpha_{ij} - m\alpha_{jk}} .
\end{aligned} \tag{5.16}$$

Finalement, d'après l'éq. (4.11) cette expression doit être dérivée deux fois par rapport au vecteur d'onde q pour obtenir la longueur de corrélation ξ .

Les valeurs et vecteurs propres de l'intégrale de transfert (4.40) ont été obtenus par la méthode décrite dans le chapitre 4. Dans la plupart des calculs, l'opérateur de transfert a été diagonalisé sur une grille régulière constituée par 4201 valeurs de y comprises entre $y_{min} = -200/a$ et $y_{max} = 4000/a$ par pas de $1/a$ (a désigne ici le paramètre du potentiel de Morse). Les intégrations numériques ont été effectuées sur la même grille. Comme je l'ai montré dans le chapitre 4, il est crucial de définir des grilles suffisamment étendues pour que la convergence numérique des observables thermodynamiques soit assurée, particulièrement près du point critique T_c . Afin de vérifier cette convergence, j'ai effectué quelques calculs sur une grille plus étendue constituée de 6201 valeurs de y entre $y_{min} = -200/a$ et $y_{max} = 6000/a$. J'y reviendrai plus loin.

Comme je l'ai noté dans le chapitre 4, la seule contribution significative à la fonction de partition à la limite thermodynamique $N \rightarrow \infty$ vient de la plus grande valeur propre λ_0 et il est raisonnable à cette limite de négliger les valeurs propres d'indice supérieur. Dans ce chapitre, nous allons cependant étudier aussi bien les petites séquences d'ADN

que les séquences longues où on s'approche de la limite $N \rightarrow \infty$. Pour cette raison, le plus grand nombre possible de valeurs et vecteurs propres doit être pris en compte dans le calcul des observables thermodynamiques. A partir des tests numériques, on peut conclure que la prise en compte des 400 premières valeurs propres est cependant largement suffisante pour la convergence des quantités calculées ici.

5.2 Effets de la finitude de taille au voisinage de T_c

Ainsi que je l'ai rappelé au début de ce chapitre, un système de taille finie ne subit pas de transition de phase. A la température correspondant au maximum de la chaleur spécifique, l'énergie libre est une fonction analytique et, par conséquent, toutes les fonctions thermodynamiques ont un comportement régulier. Soit L la taille d'un système possédant un comportement critique à la limite thermodynamique $L \rightarrow \infty$. Les effets de finitude de taille se manifestent dans ce système en $e^{-L/\xi}$. En d'autres termes, ces effets deviennent significatifs lorsque $\xi \sim L$ et lissent la singularité du point critique. Un exemple simple de ce phénomène se manifestant dans le modèle d'Ising est décrit dans la réf. [76]. Pour un système d'Ising de taille infinie, le paramètre d'ordre saute de façon discontinue de $-M_{cr}$ à $+M_{cr}$ lorsque le champ magnétique externe H passe par la valeur critique $H = 0$. En revanche, si la taille du système est finie, alors la transition a lieu sur une région étendue de taille $\Delta H \simeq k_B T / (M_{cr} L^d)$ avec une pente également finie, de l'ordre de $\sim M_L^2 L^d / (k_B T)$, où d est la dimension du système et M_L est la valeur la plus probable de la magnétisation.

Pour les modèles d'ADN unidimensionnels étudiés dans ce travail, la taille L du système est égale au nombre N de paires de bases multiplié par la distance entre les sites successifs. Puisque ce dernier paramètre ne joue aucun rôle dans la dynamique de ces modèles, j'utiliserai dans la suite de ce chapitre les grandeurs N et L sans distinction.

Pour une séquence de longueur N , la première tâche consiste à déterminer la température critique, que l'on notera $T_c(N)$. J'ai choisi pour $T_c(N)$ la température correspondant au maximum de la chaleur spécifique, qui est plus prononcé que le maximum de ξ et per-

met par conséquent une localisation plus précise de $T_c(N)$. Deux observations confirment que la température critique ainsi calculée est correcte. Premièrement, $T_c(N)$ évolue en $1/N$, une loi prédite aussi bien par la théorie de *finite-size scaling* que par les expériences. En haut de la figure 5.1, on voit la température réduite $1 - T_c(N)/T_c$ en fonction de N , où T_c désigne $T_c(\infty)$. Avec des noyaux TI définis sur une grille de taille 4201, s'étendant de $y_{min} = -200/a$ à $y_{max} = 4000/a$, T_c est égale à 281.40 K pour le modèle DPB et 367.63 K pour le modèle JB. Les symboles vides correspondent aux résultats de la méthode TI, tandis que les lignes droites montrent les lois de puissance ajustées à partir des résultats TI. Les pentes de ces lois de puissance, $N^{-1.00}$ pour le modèle DPB et $N^{-1.05}$ pour le modèle JB, sont en parfait accord avec le comportement en $1/N$ prédit théoriquement et observé expérimentalement. Notons cependant que les coefficients de proportionnalité de ces lois de puissance sont trop grands d'un facteur 4 pour le modèle DPB et 6.5 pour le modèle JB par rapport aux résultats expérimentaux. On pourrait évidemment changer légèrement les paramètres des modèles afin de réduire le coefficient de proportionnalité de cette loi, mais je ne me suis pas penché sur cette question. La validité de la méthode de localisation de $T_c(N)$ est également vérifiée, comme nous le verrons plus loin, par le fait que les courbes des observables intensives c_V , ξ et $\langle y \rangle$ tracées pour différentes longueurs de séquence N en fonction de la température réduite

$$t(N) = \frac{T - T_c(N)}{T_c(N)}. \quad (5.17)$$

sont superposées loin de la température critique.

Afin de montrer explicitement les effets de finitude de taille sur la dénaturation de l'ADN, j'ai d'abord calculé l'entropie par particule, $s = S/N$, pour une chaîne de longueur infinie et une séquence de petite taille, pour les modèles DPB et JB. Comme nous l'avons vu dans le chapitre 4, la discontinuité de la transition est caractérisée à la limite thermodynamique par la singularité de l'entropie, qui fait un saut brusque à la température critique. Sur la figure 5.2, on voit clairement la singularité en question pour la séquence de longueur infinie, tandis que pour la séquence finie de taille $N = 100$, on

retrouve une fonction continue de la température. En d'autres termes, la discontinuité disparaît.

En intégrant l'éq. (4.15), on peut voir qu'à la limite thermodynamique le saut discontinu (4.24) de l'entropie aux deux côtés du point critique doit être forcément accompagné par une divergence de cette même quantité à T_c . La divergence de l'entropie est de la forme $s \propto t^{1-\alpha}$, où α est l'exposant caractéristique de c_V , qui est légèrement mais significativement plus grand que 1 (voir le tableau 4.1). L'absence sur la figure 5.2 d'un quelconque pic correspondant à la divergence de s est expliquée par le fait que, dû aux petites valeurs de l'exposant α et du préfacteur, cette observable diverge très lentement, de sorte que sur l'intervalle où la méthode TI est valable s n'augmente que d'une valeur nettement plus petite que celle du saut (4.24).

J'ai ensuite calculé la chaleur spécifique par particule, c_V , en fonction de la température pour sept longueurs de séquence variant de $N = 100$ à $N = \infty$. Les résultats sont présentés sur la figure 5.3 pour les modèles DPB et JB. Sur cette figure, on remarque que la disparition de la singularité de c_V se manifeste par une diminution de sa valeur maximale lorsque N décroît, mais aussi par un élargissement de la région de température $|t(N)|$ sur lequel c_V augmente rapidement. Cela est particulièrement clair pour le modèle DPB, qui subit une transition très brusque, c'est à dire une transition qui à la limite thermodynamique n'est visible qu'aux très petites valeurs de $|t| = |t(\infty)|$ [40]. En examinant attentivement la figure 5.3, on remarque que les deux modèles d'ADN conduisent à des résultats très similaires jusqu'à $N \simeq 1000$, mais que la nature plus abrupte de la transition de phase du modèle DPB se manifeste pour les séquences plus longues.

J'ai également calculé pour le modèle JB la longueur de corrélation en fonction de la température pour différentes valeurs de N . Le calcul a été effectué à partir de l'équation (4.11) et l'expression de $S(q)$ de l'équation (5.15). Pour un système de taille finie, on s'attend à ce que la valeur de ξ à $T_c(N)$ soit du même ordre de grandeur que la taille du système et donc qu'elle augmente proportionnellement à N . Ce comportement peut être vérifié sur la figure 5.1 qui montre l'évolution de ξ (en unité de la distance entre les sites

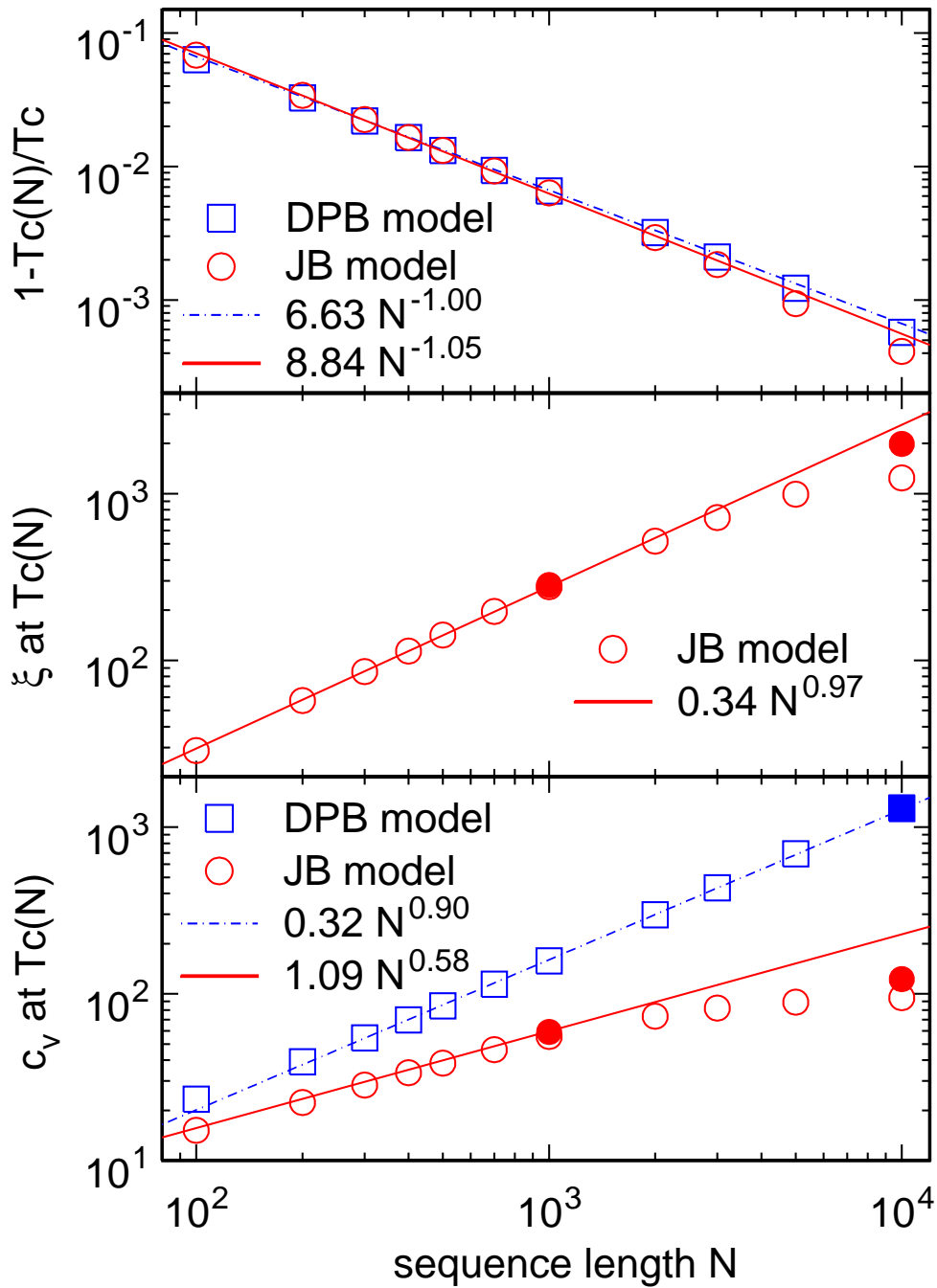


Fig. 5.1: Courbes log-log, en fonction de la longueur N de la séquence, de la température critique réduite, $1 - T_c(N)/T_c(\infty)$ (en haut), de la longueur de corrélation ξ à $T_c(N)$ (au milieu), et de la chaleur spécifique c_v à $T_c(N)$ (en bas). ξ est calculée en unités de la distance entre deux sites successifs et c_v en unités de k_B . Les carrés et les cercles correspondent aux résultats obtenus par la méthode TI respectivement pour le modèle DPB et JB. Les symboles vides désignent les résultats obtenus avec des matrices de taille 4201 et les symboles pleins des matrices de taille 6201 (voir la fin de la section 5.2). Les lignes droites montrent l'ajustement des résultats calculés par des lois de puissance.

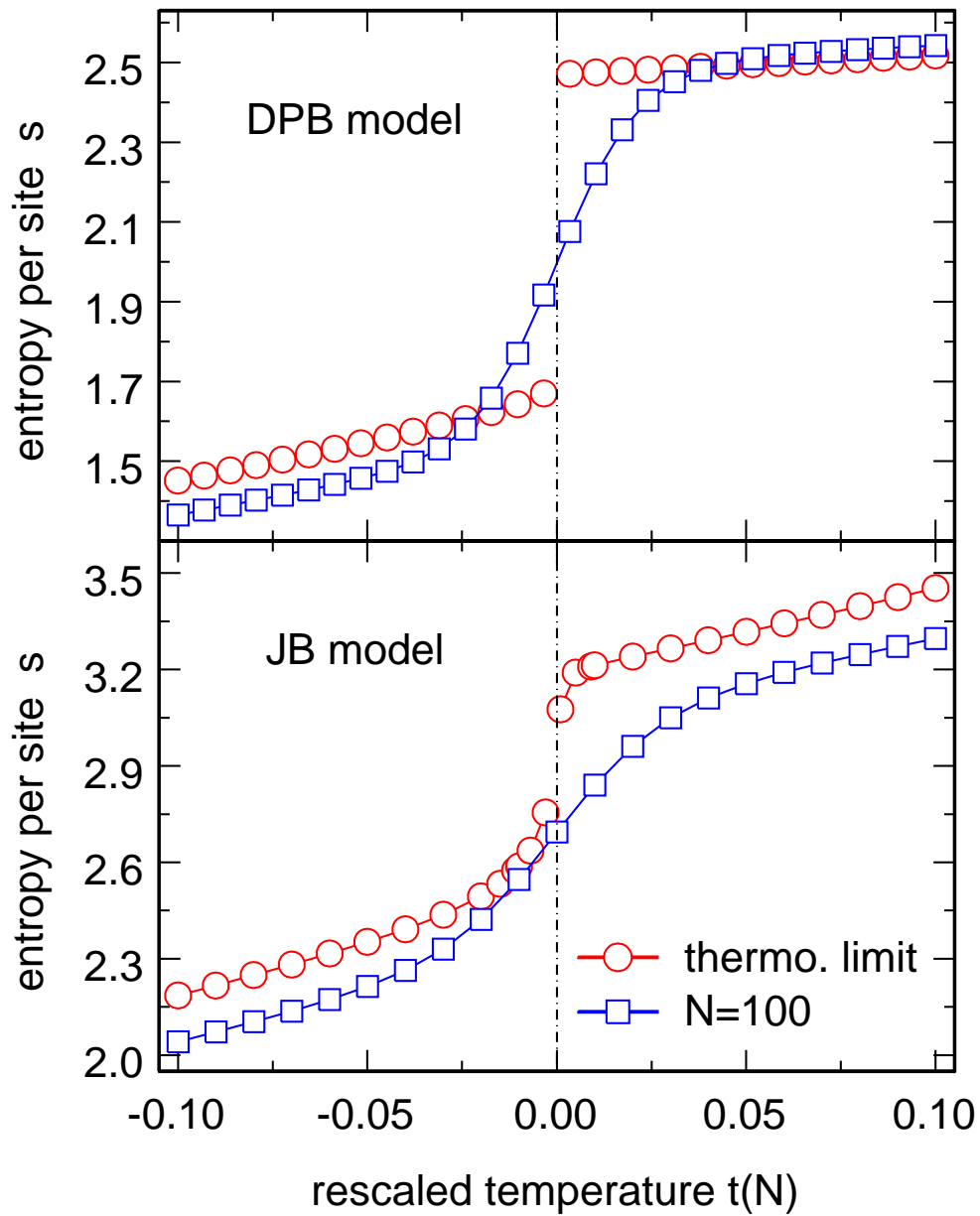


Fig. 5.2: Entropie par particule s , en fonction de la température réduite $t(N)$, pour une séquence de longueur infinie (cercles) et une séquence finie avec $N = 100$ paires de bases (carrés), pour le modèle DPB (en haut) et le modèle JB (en bas). s est tracée en unités de k_B .

successifs) à $T_c(N)$ en fonction de N : on voit que ξ est bien du même ordre de grandeur que N et qu'elle évolue comme $N^{0.97}$. On s'écarte légèrement de cette loi de puissance pour les trois derniers points correspondant à la limite $N \geq 3000$: j'y reviendrai à la fin de cette section. En bas de la figure 5.4, j'ai tracé ξ en fonction de la température pour sept valeurs de N variant de $N = 100$ à $N = \infty$. On observe le même lissage de la transition de phase que pour la chaleur spécifique (figure 5.3).

J'ai finalement calculé l'évolution de l'écart moyen entre les paires de bases $\langle y \rangle$ en fonction de la température pour le modèle JB et différentes longueurs de séquence N . Le calcul a été effectué à partir de la formule (5.9) et le résultat est présenté en haut de la figure 5.4. Pour cette observable, en plus des effets de finitude de taille déjà observés pour la chaleur spécifique (figure 5.3) et la longueur de corrélation (bas de la figure 5.4), on obtient un résultat remarquable : à la température critique, $\langle y \rangle$ semble tendre vers la même valeur quelle que soit la taille de la séquence. Cette valeur semble n'être rien d'autre que l'approximation de l'infini imposée par les limitations numériques de la méthode TI.

Avant de conclure cette section, je voudrais faire quelques remarques concernant la précision numérique de la méthode TI. Comme nous l'avons déjà vu dans le chapitre 4, la précision de cette méthode dépend de la taille des matrices TI que l'on diagonalise. Pour pouvoir nous approcher le plus près possible du point critique T_c , nous devons choisir des matrices très grandes. Les résultats présentés jusque ici ont été obtenus avec des matrices définies sur des grilles de taille 4201 s'étendant de $y_{min} = -200/a$ à $y_{max} = 4000/a$. Pour vérifier la convergence numérique de ces résultats, j'ai calculé quelques points sur une grille plus grande de taille 6201 qui s'étend de $y_{min} = -200/a$ à $y_{max} = 6000/a$. Les points calculés avec la nouvelle grille sont présentés en symboles pleins sur la figure 5.1 pour la chaleur spécifique et la longueur de corrélation à $T_c(N)$. On remarque que pour le modèle DPB les résultats obtenus avec la grille de taille 4201 sont déjà convergés jusqu'à $N = 10000$ tandis que les résultats du modèle JB sont faiblement convergés pour les valeurs de N au-dessus de quelques milliers. Notons que cela est plutôt rassurant

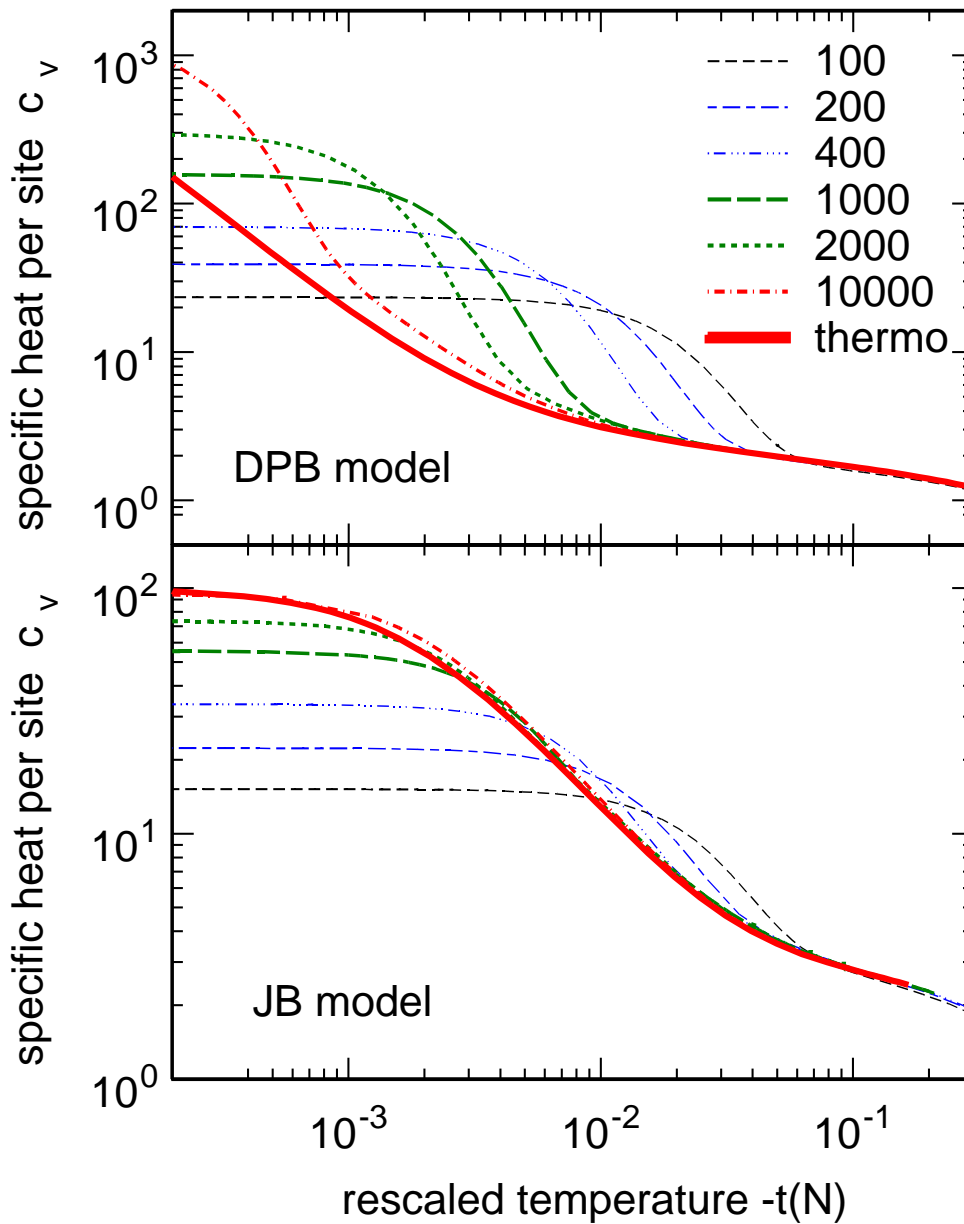


Fig. 5.3: Courbes log-log de la chaleur spécifique par particule c_V en fonction de $-t(N)$, pour le modèle DPB (en haut) et le modèle JB (en bas), tracées pour sept longueurs de séquence N variant de $N = 100$ à $N = \infty$. c_V est tracée en unités de k_B . En réalité, à la limite thermodynamique $N \rightarrow \infty$, c_V diverge à la température critique mais les limitations numériques de la méthode TI ne permettent pas d'observer cette divergence.

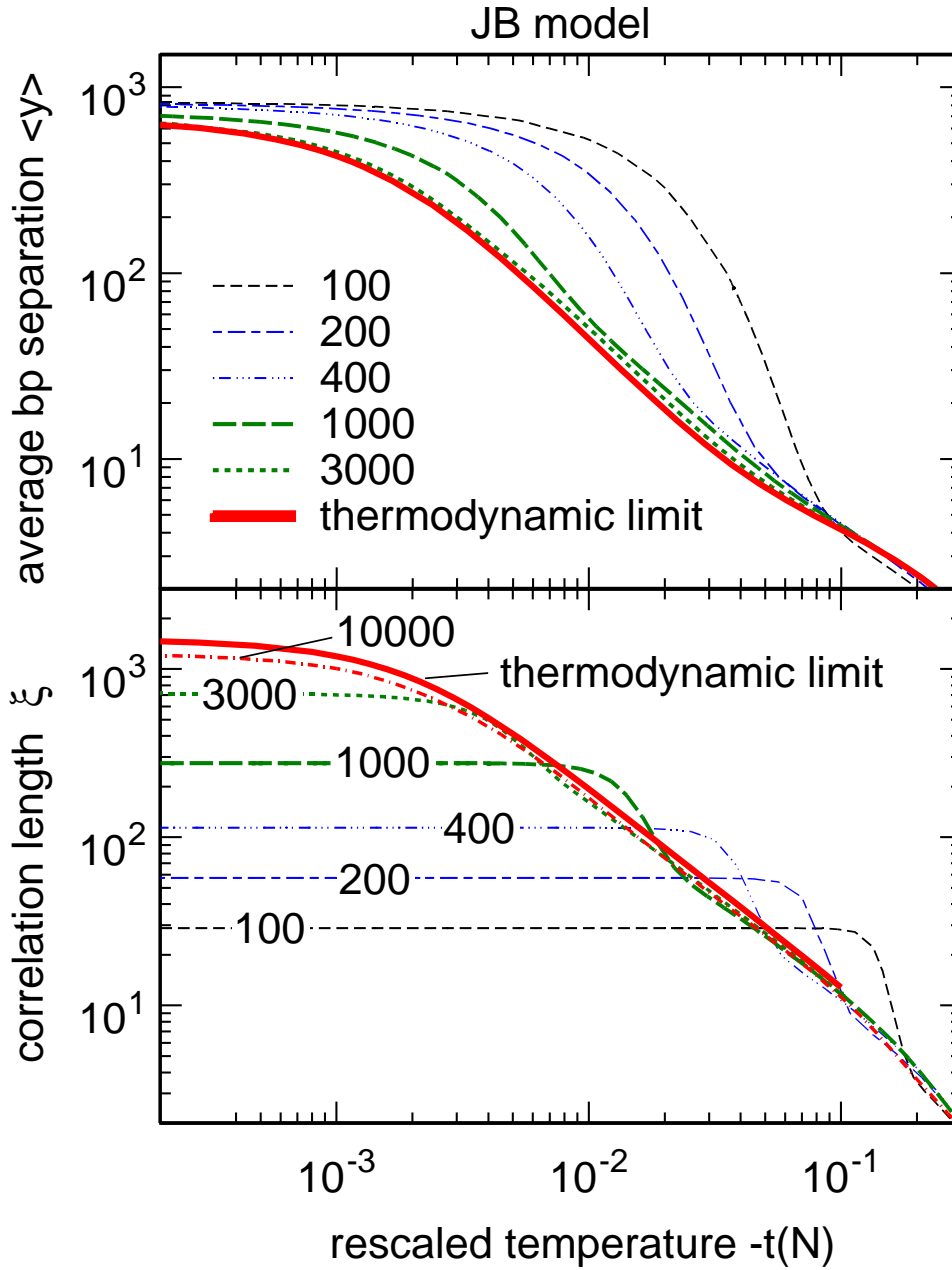


Fig. 5.4: Courbes log-log de la longueur de corrélation ξ (en bas) et de l'écart moyen des paires de base $\langle y \rangle$ (en haut) en fonction de $-t(N)$, pour le modèle JB et sept longueurs de séquence N variant de $N = 100$ à $N = \infty$. ξ est calculée en unités de la distance entre sites successifs, et $\langle y \rangle$ en unités de l'inverse $1/a$ du paramètre du potentiel de Morse. En réalité, à la limite thermodynamique $N \rightarrow \infty$, $\langle y \rangle$ et ξ divergent à la température critique, mais les limitations numériques de la méthode TI ne permettent pas d'observer cette divergence.

puisque c'est également dans cette région de valeurs de N que les lois de puissance de la figure 5.1 sont violées pour le modèle JB. Ce résultat est également en bon accord avec les figures 5.3 et 5.4 qui montrent que les séquences de longueur $N = 10000$ sont encore éloignées de la limite thermodynamique pour le modèle DPB, ce qui explique le bon fonctionnement de la méthode TI jusqu'à $N = 10000$. En revanche, pour le modèle JB avec $N = 1000 - 2000$, on est déjà très proche de la limite thermodynamique, ce qui explique la faible convergence de la méthode TI pour des séquences avec $N \gg 1000$.

5.3 Renormalisation de l'énergie libre

L'analyse des effets de taille dans la théorie du finite size scaling est basée sur l'hypothèse selon laquelle la longueur de corrélation ξ est la seule longueur caractéristique importante proche de T_c . Cela signifie que le comportement critique du système dépend de l'évolution de ξ comparée à la taille L du système : en fait, la singularité au point critique commence à être lissée dès que $L/\xi \sim 1$. En tenant compte de l'équation (4.19), valable pour les grandes valeurs de L (ou de N), le rapport L/ξ peut s'écrire comme une puissance de $|t|N^{1/\nu}$. En l'absence de champ externe, il est donc naturel d'écrire la partie singulière de l'énergie libre sous la forme

$$f_{sing} = N^{-d} Y(|t|N^{1/\nu}). \quad (5.18)$$

où Y est une fonction homogène dont la forme explicite ne nous intéresse pas. En dérivant deux fois l'équation (5.18) par rapport à t , on obtient pour la chaleur spécifique

$$c_V = N^p G(|t|N^\sigma), \quad (5.19)$$

où

$$\begin{aligned}\rho &= \frac{2}{\nu} - d \\ \sigma &= \frac{1}{\nu}\end{aligned}\tag{5.20}$$

et G est une fonction homogène qui est proportionnelle à la dérivée seconde de Y . En utilisant la loi de Josephson (4.32) et en comparant l'équation (5.19) avec l'équation (4.15) à la limite $N \rightarrow \infty$, on peut également écrire ρ et σ sous la forme

$$\begin{aligned}\rho &= \frac{\alpha}{\nu} \\ \sigma &= \frac{1}{\nu}.\end{aligned}\tag{5.21}$$

Dans les articles concernant les effets de finitude de taille, on rencontre plus souvent la forme (5.21) que la forme (5.20) des exposants ρ et σ (voir par exemple l'éq. (1.7) de la réf. [77]).

Afin de pouvoir traiter des cas plus complexes pour lesquels il peut exister plusieurs longueurs caractéristiques significatives au point critique, comme c'est le cas pour la transition de dénaturation de l'ADN, Binder et ses collaborateurs ont proposé une approche [78] basée sur l'introduction d'une nouvelle variable significative u dans la relation (5.18), ce qui donne

$$f_{sing} = N^{-d} F(tN^{y_t}, uN^{y_u})\tag{5.22}$$

où y_t et y_u sont les exposants du groupe de renormalisation, avec $y_t = 1/\nu$. Après avoir fait quelques approximations et un peu d'algèbre, les auteurs obtiennent une généralisation des équations (5.21),

	modèle DPB		modèle JB	
	ρ	σ	ρ	σ
Eq. (5.20)	0.79 (0.70-0.88)	0.89 (0.85-0.94)	0.63 (0.55-0.71)	0.81 (0.77-0.85)
Eq. (5.21)	1.29 (1.17-1.43)	0.89 (0.85-0.94)	0.92 (0.83-1.02)	0.81 (0.77-0.85)
Eq. (5.23)	1.78 (1.04-3.25)	1.39 (1.03-2.11)	1.82 (0.98-3.84)	1.41 (0.99-2.42)
ajusté	0.85 (0.80-0.90)	1.00 (0.98-1.02)	0.45 (0.35-0.55)	0.90 (0.85-0.95)

Tab. 5.1: Valeurs des coefficients ρ et σ de l'éq. (5.19) pour les modèles DPB et JB. Les trois premières lignes montrent les valeurs calculées à partir des éqs. (5.20), (5.21), (5.23) et les exposants critiques du tableau 4.1. La dernière ligne présente les exposants ajustés à la main pour que les courbes de c_V/N^ρ en fonction de tN^σ soient superposées pour différentes valeurs de N . Les incertitudes présentées entre parenthèses ont été obtenues en supposant des incertitudes additives de 5% sur chacun des exposants critiques du tableau 4.1. L'incertitude des valeurs ajustées est plus subjective. Elle a été estimée en observant l'évolution de c_V/N^ρ en fonction de tN^σ lorsque l'on fait varier ρ et σ .

$$\rho = \frac{2d}{2\beta + \gamma} - d \quad (5.23)$$

$$\sigma = \frac{d}{2\beta + \gamma},$$

où β et γ sont respectivement les exposants critiques du paramètre d'ordre (4.16) et de la susceptibilité (4.18).

Le tableau 5.1 montre les valeurs de ρ et σ obtenues à partir des équations (5.20), (5.21) et (5.23), et les valeurs ajustées. Ces dernières ont été obtenues en variant les exposants ρ et σ à la main jusqu'à ce que les courbes représentant c_V/N^ρ en fonction de tN^σ soient superposées pour les différentes valeurs de N . En posant $t = 0$ dans l'éq. (5.19), on remarque que la valeur de c_V à la température $T_c(N)$ varie comme N^ρ . ρ a donc été ajusté au voisinage de la valeur de la pente caractérisant la dépendance de c_V en fonction de N (voir la figure 5.1), alors que l'exposant σ a été ajusté au voisinage de $1/\nu$. Les incertitudes sur les exposants ρ et σ calculés, qui sont présentées dans les trois premières lignes du tableau 5.1, ont été estimées en supposant une erreur additive de 5% sur les valeurs des exposants critiques présentées dans le tableau 4.1. L'estimation de l'incertitude concernant les valeurs ajustées (dernière ligne du tableau 5.1) est cependant plus subjective dans la mesure où elle n'est basée que sur l'évolution de c_V/N^ρ en fonction

de tN^σ lorsque l'on varie ρ et σ . Le tableau 5.1 montre que les valeurs de ρ et σ obtenues à partir des équations (5.20) sont en bon accord avec les valeurs ajustées, ce qui n'est pas le cas des exposants obtenus avec les équations (5.21) et (5.23). Les figures 5.5 et 5.6 montrent l'évolution de la chaleur spécifique réduite c_V/N^ρ en fonction de tN^σ , pour les modèles DPB et JB et des exposants ρ et σ obtenus à partir des expressions (5.21) (figures du haut) et ajustés (figures du bas). En haut de ces deux figures, on voit que les courbes tracées pour différentes valeurs de N sont loin d'être superposées. Nous avons vérifié que la superposition est encore moins satisfaisante avec les équations (5.23). En revanche, comme on peut le constater en bas des figures 5.5 et 5.6, les courbes obtenues avec les valeurs ajustées de ρ et σ sont bien superposées, comme c'est également le cas lorsque ρ et σ sont obtenus à partir de l'éq. (5.20).

Certes, la courbe correspondant à $N = 10000$ ne respecte pas la superposition pour le modèle JB (bas de la figure 5.6). Cela n'est cependant pas surprenant dans la mesure où nous avons vu dans la section précédente que pour le modèle JB la limite thermodynamique est atteinte à $N \sim 1000 - 2000$. Cela signifie que pour $N = 10000$, on est clairement à la limite thermodynamique, alors que j'ai précisé lors de la présentation de la méthode TI que celle-ci n'est pas capable de reproduire les divergences de la limite thermodynamique à T_c .

Le fait que les courbes tracées à partir des éq. (5.20) soient correctement superposées montre que l'hypothèse conduisant à l'éq. (5.18) est satisfaite. D'autre part, le fait que les exposants de l'éq. (5.21) ne conduisent pas à une bonne superposition des courbes reflète simplement le fait que la loi de Josephson (4.32) n'est vérifiée pour aucun des deux modèles en question, une conclusion à laquelle nous étions déjà arrivés dans le chapitre 4 (voir en particulier le tableau 4.2). Enfin, on peut remarquer que les expressions de ρ et σ de l'équation (5.23) peuvent très facilement s'obtenir à partir de celles de l'équation (5.20) en utilisant les lois de Rushbrooke et Josephson. Si la loi de Rushbrooke est valide, ce n'est, encore une fois, pas le cas pour celle de Josephson, ce qui explique que les exposants ρ et σ calculés à partir de l'équation (5.23) ne conduisent pas à une

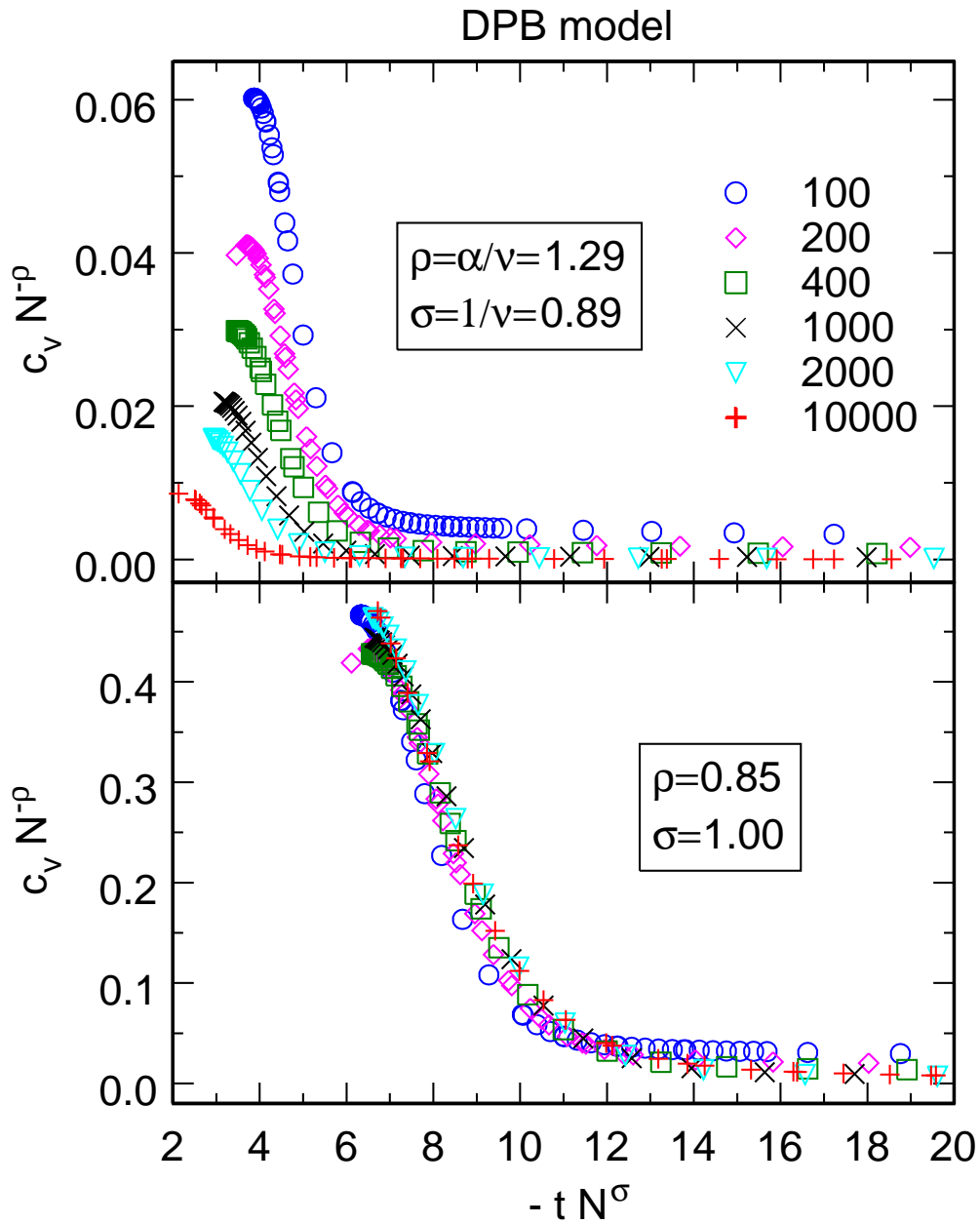


Fig. 5.5: Tracé, en fonction de $-tN^\sigma$, de la chaleur spécifique réduite c_V/N^ρ pour le modèle DPB, pour six longueurs de séquence N variant de $N = 100$ à $N = 10000$. Les coefficients ρ et σ sont soit obtenus à partir de l'équation (5.20) (en haut), soit ajustés à la main (en bas).

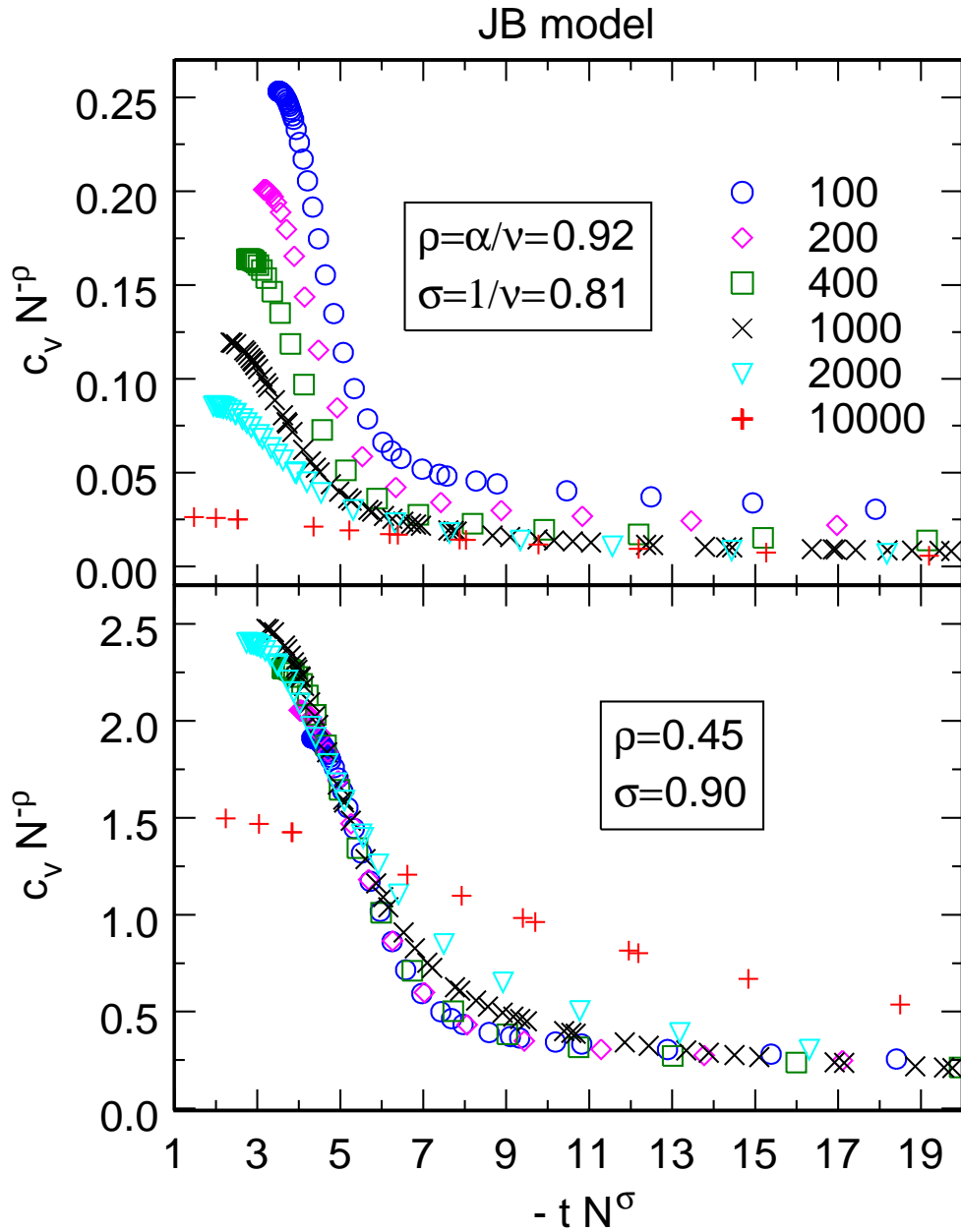


Fig. 5.6: Tracé, en fonction de $-tN^\sigma$, de la chaleur spécifique réduite c_V/N^ρ pour le modèle JB, pour six longueurs de séquence N variant de $N = 100$ à $N = 10000$. Les coefficients ρ et σ sont soit obtenus à partir de l'équation (5.20) (en haut), soit ajustés à la main (en bas).

superposition correcte des courbes.

5.4 Conclusion

Dans ce chapitre, j'ai présenté une version généralisée de la méthode TI adaptée au calcul des fonctions thermodynamiques des séquences d'ADN finies aux bords libres. Bien que le calcul des quantités intensives comme la distance moyenne des paires de bases $\langle y \rangle$ et la longueur de corrélation ξ soient longs et complexes, j'ai appliqué la méthode TI modifiée au calcul de ces observables et utilisé ce nouvel outil pour caractériser les effets de la finitude de taille sur la transition de dénaturation de l'ADN. J'ai montré que la disparition progressive de la singularité du point critique lorsque l'on réduit la taille de la séquence se manifeste clairement à travers l'évolution des observables thermodynamiques en fonction de la température. J'ai constaté que l'énergie libre des systèmes étudiés peut être écrite sous forme d'une fonction homogène, mais que, la loi de Josephson (4.32) n'étant pas valide, le choix des exposants ρ et σ caractérisant la renormalisation de la chaleur spécifique doit être fait avec précaution.

La méthode de l'intégrale de transfert est visiblement le seul outil fiable pour étudier la physique statistique de la dénaturation de l'ADN. Cependant, l'approche présentée dans ce chapitre ne peut être appliquée qu'aux séquences homogènes. Or, nous savons que le désordre peut avoir des effets très significatifs sur les propriétés critiques d'un système. Pour cette raison, je vais présenter dans le chapitre 6 une version encore plus sophistiquée de la méthode TI, qui est adaptée aux séquences hétérogènes, afin d'étudier la dénaturation de l'ADN en présence de désordre.

6. TRANSITION DE DÉNATURATION DES SÉQUENCES D'ADN HÉTÉROGÈNES

Les chapitres 4 et 5 de cette thèse ont été consacrés à l'étude de la transition de dénaturation des séquences d'ADN homogènes pour deux modèles d'ADN. Comme je l'ai précisé dans le chapitre 5, l'homogénéité de la séquence permet de se concentrer sur les propriétés d'un seul processus de dénaturation. Cependant, les molécules d'ADN réalistes, c'est à dire celles ayant un rôle dans les processus biologiques, possèdent forcément une structure génétique hétérogène. Plus précisément, les gènes sont codés par une distribution des paires de bases (A, T, G, ou C) d'aspect quasi-aléatoire. Il est donc crucial de comprendre les effets de cette hétérogénéité sur la transition de dénaturation de l'ADN.

Dans le langage de la physique statistique, une distribution hétérogène des constituants individuels d'un système complexe est appelé du *désordre*. Les études concernant les effets du désordre sur les transitions de phase considèrent en général deux types de désordre. Le premier correspond à une distribution hétérogène aléatoire du champ externe couplé à chaque particule, comme celui de l'éq. (4.2) (*field disorder*). Le deuxième type de désordre est caractérisé par une distribution hétérogène des énergies d'interaction entre les différents composants du système (*bond disorder*). C'est précisément le cas d'une chaîne d'ADN décrite par le modèle JB hétérogène (2.15).

Pour un système qui subit une transition de phase du second ordre, Harris a montré que l'introduction de désordre ne lisse pas la transition si la condition

$$\nu \geq \frac{2}{d} \tag{6.1}$$

est vérifiée [79]. Dans l'inégalité (6.1), qui est appelée *critère de Harris*, ν est l'exposant critique de la longueur de corrélation du système homogène (pur) et d désigne la dimension du système. La signification physique de l'expression (6.1) est claire : si la longueur de corrélation, qui mesure l'étendue des corrélations, est suffisamment divergente dans le cas homogène, alors, même en présence du désordre, le système se comporte à grande échelle comme un système homogène et la classe d'universalité du point critique n'est pas modifiée.

Les effets du désordre sur une transition discontinue ont été analysés plus tard par Imry et Wortis [23]. Cette analyse est basée sur l'existence de deux énergies caractéristiques intervenant dans la formation des domaines de cohérence, autrement dit des régions qui changent d'état indépendamment du reste du système. Pour un domaine de taille L , une telle transition locale diminue l'énergie libre du système de $\epsilon_+ \sim L^{d/2}$ où d est la dimension du système. En revanche, la différence de phase entre ce domaine et le reste du système coûte une énergie d'interface de l'ordre de $\epsilon_- \sim L^{d-1}$. Pour qu'il y ait la formation d'un tel domaine, la réduction d'énergie libre ϵ_+ doit largement compenser le coût énergétique ϵ_- , ce qui selon Imry et Wortis est toujours le cas pour $d < 2$. Cela signifie que chaque domaine de cohérence du système vérifiant l'inégalité $\epsilon_+ > \epsilon_-$ change de phase à sa propre température critique T_c , transformant le système en un mélange de phases et lissant ainsi la transition du système entier sur une plage de températures comprises entre la température d'apparition du premier domaine et celle du dernier. Selon Berker [50], une transition discontinue d'un système de dimension $d < 2$, impliquant une brisure de symétrie pour le système homogène, devient ainsi du second ordre en présence du désordre. En revanche, si la transition n'implique aucun changement de symétrie dans le cas homogène, alors le point critique disparaît tout simplement pour le système hétérogène.

En ce qui concerne les effets du désordre sur la dénaturation thermique de l'ADN, les avis sont partagés. A partir de l'étude d'un modèle de self-avoiding walks (SAWs) appliqué au modèle de Poland-Scheraga (PS) en présence du désordre [80], Coluzzi et

Yeramian ont conclu que la transition de dénaturation de ce modèle, qui est discontinue dans le cas homogène, devient du second ordre en présence de désordre. En revanche, l'étude du modèle de PS par Garel et Monthus [20] suggère que l'ordre de cette transition n'est pas affecté par la présence de désordre. Il faut préciser que, pour le calcul des exposants caractéristiques, les auteurs de ces références n'ont pas étudié les observables pour une séquence particulière mais au contraire ont tiré ces conclusions du comportement des fonctions thermodynamiques moyennées sur le désordre, c'est à dire sur un grand nombre de réalisations d'hétérogénéité de la séquence. Cela ne correspond pas vraiment à ce qui se passe dans les systèmes biologiques, puisque dans la réalité (et également dans les expériences) les processus génétiques impliquent une séquence hétérogène particulière. D'autre part, comme la dimension effective est $d = 1$ pour l'ADN, on s'attend à ce que la formation des domaines, définis plus haut, soit toujours favorisée lors de la dénaturation thermique. C'est également ce qui est observé lors des expériences, puisque les courbes de dénaturation expérimentales (voir la figure 1.5) montrent un processus de dénaturation en plusieurs étapes. En d'autres termes, un processus de dénaturation peut faire intervenir des transitions locales. La localisation d'une température critique unique pour le calcul des exposants critiques est donc souvent impossible lorsqu'il s'agit d'une séquence hétérogène particulière. La seule tâche ayant un sens du point de vue physique serait donc de comprendre la nature de la transition pour chacune de ces dissociations locales, en supposant bien sûr que les portions correspondantes soient suffisamment grandes (comme je l'ai précisé dans le chapitre 5, il ne s'agirait pas de vraies transitions de phase dans le cas contraire). Cela est plus ou moins équivalent à découper virtuellement la séquence totale en sous-ensembles complémentaires, dont chacun serait caractérisé par un seul point critique pour lequel il serait possible de calculer les exposants caractéristiques.

Dans ce chapitre, je vais exposer le travail qui est en cours dans ce domaine et présenter les résultats déjà obtenus. Dans la section 6.1, je vais décrire une méthode TI modifiée, développée par Zhang et al. [15], que j'ai adaptée à une version "hétérogénéisée" du modèle JB (2.11). Ce dernier peut s'écrire sous la forme d'un hamiltonien à N degrés

de liberté

$$H = \sum_{n=1}^N \{V_M(y_n) + W^{(n)}(y_n, y_{n-1})\}, \quad (6.2)$$

où

$$W^{(n)}(y_n, y_{n-1}) = \frac{\Delta H^{(n)}}{2} \left(1 - e^{-b(y_n - y_{n-1})^2}\right) + K_b (y_n - y_{n-1})^2, \quad (6.3)$$

avec les énergies d'empilement $\Delta H^{(n)}$ qui prennent les valeurs définies dans le tableau 1.1 et $D = 0.04$ eV, $a = 4.45 \text{ \AA}^{-1}$, $K_b = 10^{-5}$ eV \AA^{-2} et $b = 0.10 \text{ \AA}^{-2}$ comme dans les chapitres 4 et 5 (on rappelle que $1 \text{ eV} \approx 96.5 \text{ kJ/mol}$). Pour le modèle DPB hétérogène de la réf. [15] on a en revanche

$$H = \sum_{n=1}^N \left\{V_M^{(n)}(y_n) + W(y_n, y_{n-1})\right\} \quad (6.4)$$

où l'énergie d'empilement

$$W(y_n, y_{n-1}) = \frac{K}{2} (y_n - y_{n-1})^2 [1 + \rho e^{-\alpha(y_n + y_{n-1})}] \quad (6.5)$$

est indépendante de la nature des paires de bases, alors que, dans le potentiel de Morse décrivant l'attraction de deux bases appariées

$$V_M(y_n) = D_n (1 - e^{-ay_n})^2 \quad (6.6)$$

on suppose que $D_n = 0.038$ eV pour une paire de bases A-T et $D_n = 0.042$ eV pour une paire de bases G-C. Les autres paramètres de ce modèle sont $a = 4.2 \text{ \AA}^{-1}$, $K = 0.042 \text{ eV \AA}^{-2}$, $\rho = 0.5$ et $\alpha = 0.35 \text{ \AA}^{-1}$.

Dans la section 6.2, je vais ensuite montrer la validité de la méthode TI modifiée en comparant les profils de dénaturation d'une séquence hétérogène obtenus par cette méthode et par dynamique moléculaire pour les modèles des éqs. (6.2) et (6.4). Finalement, je décrirai, grâce au comportement de la chaleur spécifique, les effets du désordre sur

le comportement critique des molécules d'ADN hétérogènes. Je conclurai en expliquant brièvement le travail qui reste à accomplir pour pouvoir estimer de façon quantitative l'ordre de la transition de dénaturation en présence de désordre.

6.1 Méthode TI pour les séquences d'ADN hétérogènes

Dans cette section, je vais présenter une forme généralisée de la méthode TI pour l'étude des séquences hétérogènes. Cette méthode a également été développée par Zhang et al. [15] et je l'ai adaptée au modèle JB hétérogène de l'éq. (6.2).

Pour une chaîne d'ADN hétérogène, la fonction de partition canonique (comparer à l'éq. (5.1)), peut s'écrire

$$Z = \int dy_1 dy_2 \cdots dy_N e^{-\beta V_M^{(1)}(y_1)/2} K_2(y_2, y_1) K_3(y_3, y_2) \cdots K_N(y_N, y_{N-1}) e^{-\beta V_M^{(N)}(y_N)/2} \quad (6.7)$$

où le noyau TI d'une paire de bases n qui interagit avec la paire voisine est de la forme

$$K_n(y_n, y_{n-1}) = \exp \left[-\beta \left\{ \frac{1}{2} V_M^{(n)}(y_n) + \frac{1}{2} V_M^{(n-1)}(y_{n-1}) + W^{(n)}(y_n, y_{n-1}) \right\} \right]. \quad (6.8)$$

La non-symétrie de ce noyau pour le modèle DPB ($K(x, y) \neq K(y, x)$) rend sa diagonalisation numériquement difficile. Il devient donc nécessaire d'utiliser une technique de symétrisation pour contourner ce problème numérique, ce qui conduit à quatre types de noyaux symétriques à diagonaliser. Puisque, dans ce chapitre, nous allons très peu nous intéresser à la physique statistique du modèle DPB hétérogène, la technique de symétrisation et la suite du calcul conduisant à l'évaluation de la fonction de partition (6.7) pour le modèle DPB, présentées en détail dans la réf. [15], ne seront pas discutées ici. Je vais tout de suite décrire l'adaptation de cette technique au modèle JB.

Dans le chapitre 2, nous avons vu que pour le modèle JB le potentiel de Morse ne dépend pas de la nature de la paire de bases. Cela signifie que $V_M^{(n)}(y) = V_M^{(n-1)}(y) =$

$V_M(y)$, ce qui conduit naturellement à un noyau (6.8) déjà symétrique et évite l'étape de symétrisation. J'ai également précisé que, pour ce modèle, le contenu génétique de la séquence est caractérisé par dix valeurs différentes de l'énergie d'empilement $\Delta H^{(n)}$. Par conséquent, il existe au total dix types de noyaux symétriques de la forme (6.8) qui interviennent dans la fonction de partition (6.7) de ce modèle. L'astuce de la méthode TI généralisée consiste à développer chacun de ces noyaux sur une base orthonormale, comme dans l'éq. (4.39), sous la forme

$$K_n(y_n, y_{n-1}) = \sum_i \lambda_i^{(n)} \Phi_i^{(n)}(y_n) \Phi_i^{(n)}(y_{n-1}) \quad (6.9)$$

où $\{\Phi_i^{(n)}\}$ et $\{\lambda_i^{(n)}\}$ sont les vecteurs et valeurs propres de l'opérateur de transfert, qui satisfont

$$\int dx K_n(x, y) \Phi_i^{(n)}(x) = \lambda_i^{(n)} \Phi_i^{(n)}(y). \quad (6.10)$$

La diagonalisation des dix noyaux a été effectuée sur un réseau discret de taille 2901, avec la valeur du degré de liberté d'élongation qui variait de $y = -100/a$ à $y = 2800/a$ par pas de $1/a$. Les intégrations numériques ont été effectuées sur le même réseau. En définissant

$$\begin{aligned} a_i^{(1)} &= \int dy_1 e^{-\beta V_M(y_1)/2} \Phi_i^{(2)}(y_1) \\ a_i^{(N)} &= \int dy_N e^{-\beta V_M(y_N)/2} \Phi_i^{(N)}(y_N) \\ B_{ij} &= \sqrt{\lambda_i^{(N)} \lambda_j^{(2)}} a_i^{(N)} a_j^{(1)} \\ D_{ij}^{(n)} &= \sqrt{\lambda_i^{(n-1)} \lambda_j^{(n)}} \int dy_{n-1} \Phi_i^{(n-1)}(y_{n-1}) \Phi_j^{(n)}(y_{n-1}), \end{aligned} \quad (6.11)$$

et en substituant le développement (6.9) dans l'expression (6.7), la fonction de partition devient

$$Z = \sum_{i_2, \dots, i_N} B_{i_N i_2} D_{i_2 i_3}^{(3)} D_{i_3 i_4}^{(4)} \cdots D_{i_{N-2} i_{N-1}}^{(N-1)} D_{i_{N-1} i_N}^{(N)}. \quad (6.12)$$

Cette sommation multiple peut être exprimée comme la trace du produit de $N - 2$ matrices

$$Z = Tr \{ \mathbf{B} \mathbf{D}^{(3)} \mathbf{D}^{(4)} \cdots \mathbf{D}^{(N-1)} \mathbf{D}^{(N)} \}, \quad (6.13)$$

qui est plus appropriée que l'éq. (6.12) pour une évaluation numérique. Le calcul de la fonction de partition (6.13) permet, à partir des éqs. (4.4), (4.7) et (4.8), d'évaluer les fonctions thermodynamiques extensives comme l'énergie libre, l'entropie et la chaleur spécifique.

Le calcul des quantités intensives s'effectue de la même façon. Par exemple, dans l'espace configurationnel, l'élongation moyenne de la n -ème paire de bases s'écrit

$$\langle y_n \rangle = \frac{1}{Z} \int dy_1 dy_2 \cdots dy_N y_n e^{-\beta V_M(y_1)/2} K_2(y_2, y_1) K_3(y_3, y_2) \cdots K_N(y_N, y_{N-1}) e^{-\beta V_M(y_N)/2}. \quad (6.14)$$

En définissant

$$\begin{aligned} b_i^{(1)} &= \int dy_1 e^{-\beta V_M(y_1)/2} \Phi_i^{(2)}(y_1) y_1 \\ b_i^{(N)} &= \int dy_N e^{-\beta V_M(y_N)/2} \Phi_i^{(N)}(y_N) y_N \\ C_{ij}^{(1)} &= \sqrt{\lambda_i^{(N)} \lambda_j^{(2)}} a_i^{(N)} b_j^{(1)} \\ C_{ij}^{(N)} &= \sqrt{\lambda_i^{(N)} \lambda_j^{(2)}} b_i^{(N)} a_j^{(1)} \\ Y_{1,ij}^{(n)} &= \sqrt{\lambda_i^{(n-1)} \lambda_j^{(n)}} \int dy_{n-1} \Phi_i^{(n-1)}(y_{n-1}) \Phi_j^{(n)}(y_{n-1}) y_{n-1}, \end{aligned} \quad (6.15)$$

et en injectant le développement (6.9) dans l'équation (6.14), on obtient finalement

$$\langle y_n \rangle = \frac{1}{Z} Tr \left\{ \mathbf{B} \mathbf{D}^{(3)} \mathbf{D}^{(4)} \dots \mathbf{D}^{(n)} \mathbf{Y}_1^{(n+1)} \mathbf{D}^{(n+2)} \dots \mathbf{D}^{(N)} \right\} \quad (6.16)$$

pour $n \neq 1, N$ et

$$\langle y_{1,N} \rangle = \frac{1}{Z} Tr \left\{ \mathbf{C}^{(1,N)} \mathbf{D}^{(3)} \mathbf{D}^{(4)} \dots \mathbf{D}^{(N-1)} \mathbf{D}^{(N)} \right\} \quad (6.17)$$

aux extrémités de la chaîne. Les fonctions de corrélation, dont le comportement ne sera pas étudié dans cette thèse, peuvent être calculées de façon très similaire.

6.2 Etude de la transition de dénaturation

Pour étudier la physique statistique de l'ADN en équilibre thermique, il existe, en plus de la méthode TI et de l'intégration numérique des équations de Langevin, un certain nombre d'outils numériques comme la méthode de produits matriciels [16] et la méthode de Monte-Carlo [14]. Bien que les résultats de ces références aient montré que ces méthodes sont capables de décrire correctement certains aspects de la dénaturation de l'ADN, les études en question étaient basées sur le calcul d'observables moyennées sur la séquence entière, comme la fraction de paires de bases ouvertes ou la fraction de bulles. Autrement dit, la performance de ces méthodes n'a jamais été testée à l'échelle des paires de bases. Cela est pourtant crucial puisque les modèles étudiés sont précisément définis à cette échelle. Dans cette section je vais donc d'abord comparer les profils de dénaturation obtenus par le calcul TI et les simulations de type Langevin (MD).

L'intégration numérique des éqs. de Langevin (2.12) a été décrite dans le chapitre 2. Les calculs MD de ce chapitre sont basés sur les mêmes principes. J'ai pris un coefficient de dissipation $\gamma = 5 \text{ ns}^{-1}$. Le système, initialement à $T = 0 \text{ K}$, a été chauffé jusqu'à la température désirée avec une pente de 10^{-2} ns/K . Une fois cette température atteinte, une intégration supplémentaire a été effectuée à température constante pendant le même intervalle de temps, afin d'assurer une meilleure mise en équilibre thermique du système. Ensuite, j'ai commencé le calcul de la moyenne temporelle de la distance y_n pour chaque

paire de bases. La durée de ce calcul sera précisée plus loin. Pendant ce processus, j'ai également enregistré la température physique moyenne (2.14) dont la proximité finale à la température imposée par l'éq. de Langevin (2.12) est un critère important pour avoir une estimation de la qualité de $\langle y_n \rangle$. Pour les résultats présentés dans ce chapitre, la différence entre la température imposée et la température physique (2.14) était strictement inférieure à 0.1 K.

La figure 6.1 montre les profils de dénaturation pour une séquence hétérogène de 1793 paires de bases (code d'entrée NCB NM 001101), à une température relativement basse $T = 322$ K (en bas) et à $T = 346$ K, c'est à dire 7 K au-dessous de la dissociation de la séquence entière (en haut). La courbe continue rouge correspond au résultat du calcul TI et la courbe en tirets bleue à la simulation MD. Chaque figure principale montre le profil de la séquence entière, alors que les vignettes sont des zooms sur quelques centaines de paires de bases. On constate qu'il y a un très bon accord entre les deux algorithmes sur toute la plage de températures considérée. Ainsi qu'on peut le voir sur le tracé à 322 K, les résultats MD et TI coïncident même pour les fluctuations à l'échelle d'un petit nombre de paires de bases. De plus, il est important de noter que la position et la taille des bulles qui conduisent la transition de dénaturation de l'ADN sont identiques avec les deux méthodes. Il en va de même pour l'ouverture d'une grande portion de la séquence à 345 K.

J'ai également calculé le profil de la même séquence pour le modèle DPB hétérogène (6.4) à $T = 350$ K. Le résultat est présenté sur la figure 6.2. Je voudrais rappeler que, pour ce modèle, le seuil de la dissociation est de l'ordre de 1-2 Å, contrairement au modèle JB pour lequel ce seuil est de 10-12 Å. Cela signifie que la courbe de la figure 6.2 est proche de la dissociation. On remarque qu'il y a toujours un bon accord entre les résultats obtenus par la méthode TI et les simulations de dynamique moléculaire. Cependant, la courbe TI a une forme inhabituelle. Au début de ce chapitre, nous avons vu que le modèle DPB se contente de modéliser l'hétérogénéité par l'introduction de deux types de liaisons hydrogène; en d'autres termes il ne tient pas compte de la différence

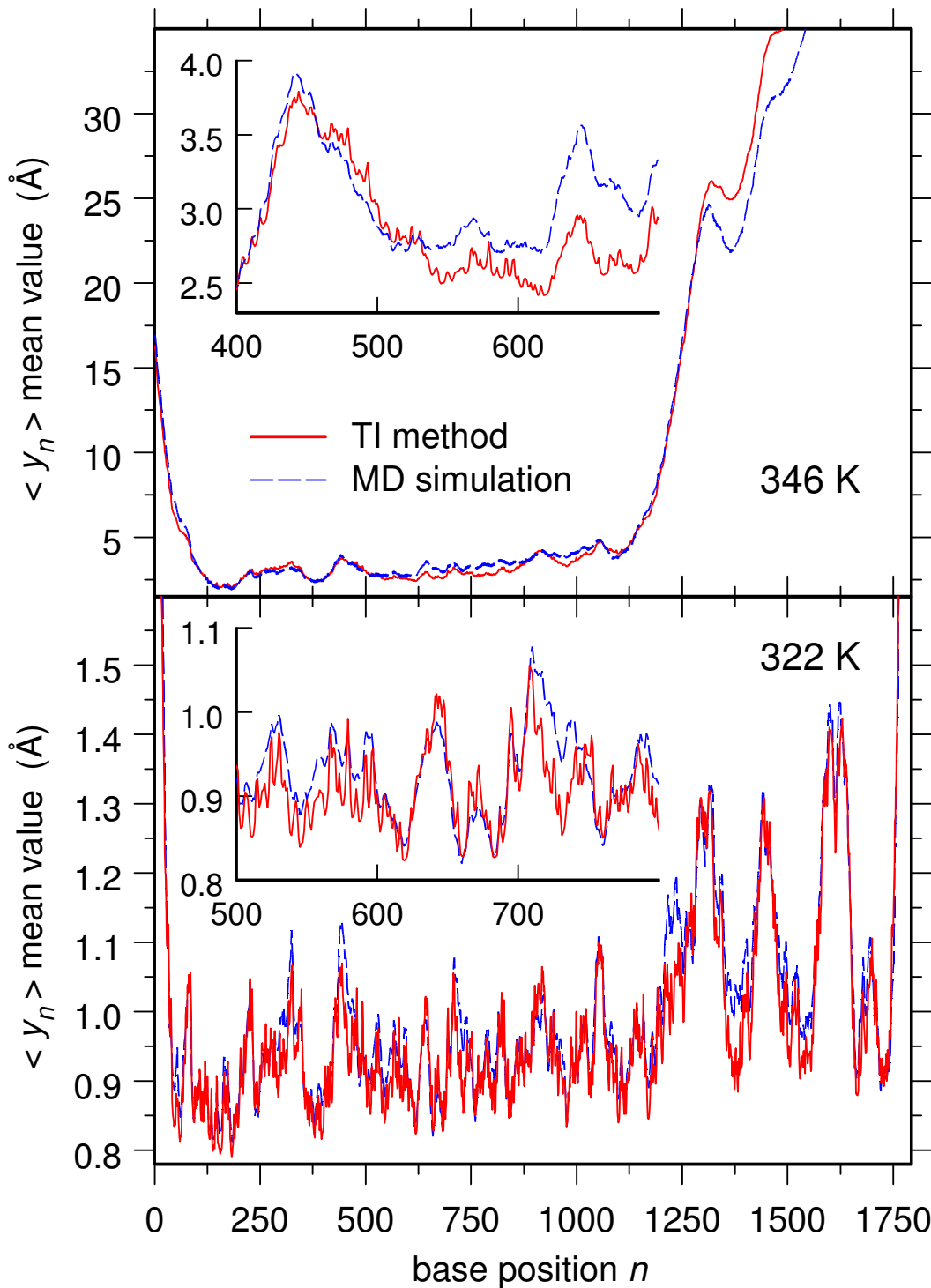


Fig. 6.1: Profils de dénaturation pour une séquence hétérogène de 1793 paires de bases (code d'entrée NCB NM 001101) calculés pour le modèle JB hétérogène (6.2), à $T = 322$ K (en bas) et à $T = 346$ K (en haut). La courbe rouge continue correspond au calcul TI et la courbe en tirets bleus à la simulation MD. Chaque figure principale montre le profil de la séquence entière, alors que les vignettes montrent un zoom sur quelques centaines de paires de bases.

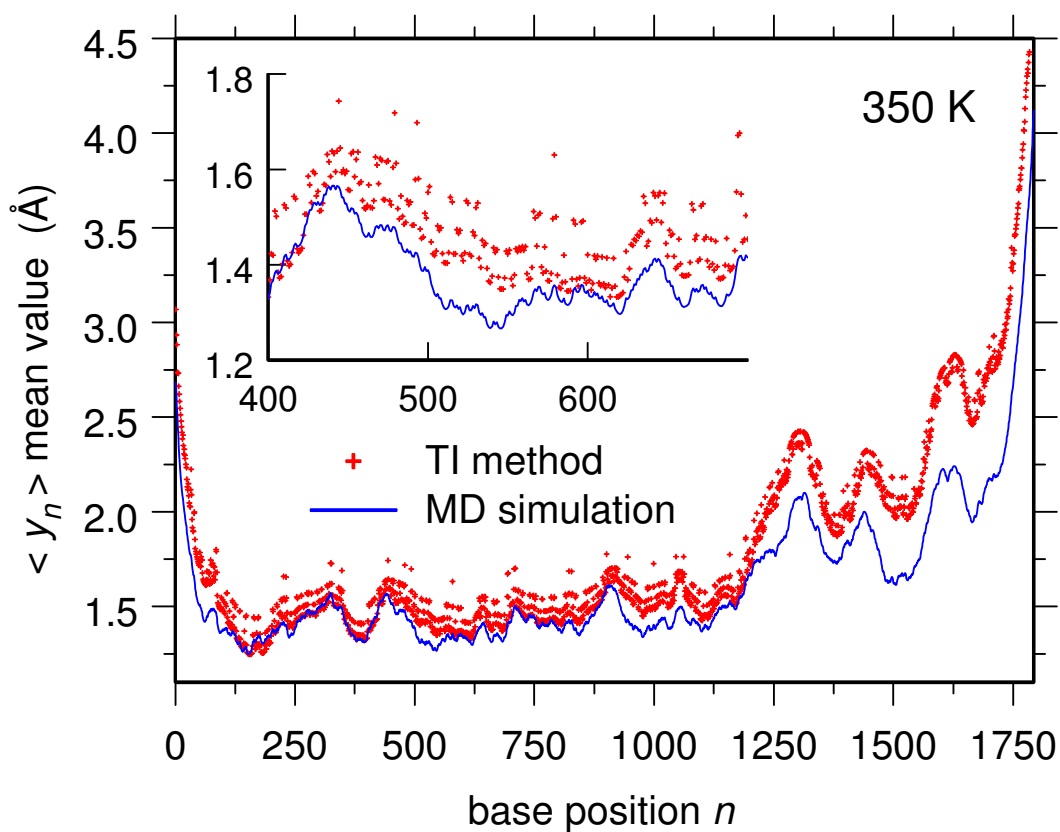


Fig. 6.2: Profil de dénaturation de la séquence hétérogène de 1793 paires de bases (code d'entrée NCB NM 001101) calculé pour le modèle DPB hétérogène (6.4). La courbe en croix rouges correspond au calcul TI et la courbe en tirets bleus à la simulation MD. La figure principale montre le profil de la séquence entière et la vignette un zoom sur quelques centaines de paires de bases.

des énergies d'empilement des paires de bases successives. Cela signifie que le modèle ne fait aucune différence entre, par exemple, les paires successives AA, AT, TT et TA ou GC, CG, CC et GG. Effectivement, on voit que la courbe qui correspond au résultat TI donne l'impression d'être constituée de quatre couches superposées (rappelons qu'il y a quatre types de noyaux pour le calcul TI de ce modèle). Cela est en fait un effet visuel qui résulte du fait que les paires de bases adjacentes, considérées comme équivalentes par le modèle DPB hétérogène, forment des blocs de répétitions et, à l'échelle de quelques centaines de paires de bases, on distingue ces répétitions comme des superpositions. Le modèle DPB hétérogène semble donc en difficulté lorsqu'il s'agit de décrire les détails d'une transition de dénaturation à l'échelle des paires de bases.

L'avantage de la méthode TI comparée à l'intégration numérique des équations de Langevin est une réduction considérable du temps de calcul. Avec un processeur à 3.2 GHz, le profil de la séquence est calculée par la méthode TI en 2 jours, quelle que soit la température, tandis que la durée d'une simulation par la dynamique moléculaire varie de 4 à 10 jours, pour un temps de simulation qui varie de 10^2 à 10^3 ns, en fonction de la proximité à la température de dénaturation. En effet, lorsque l'on s'approche de la température de dissociation, l'amplification des fluctuations (voir le chapitre 3) nécessite de moyennner ces fluctuations sur des intervalles de temps de plus en plus longs. Par conséquent, la méthode TI est beaucoup plus stable que les simulations MD lorsque l'on s'approche de T_c .

On peut donc conclure que la méthode TI modifiée est capable, lorsqu'elle est appliquée à notre modèle d'ADN hétérogène, i) de reproduire les détails du processus de dénaturation jusqu'à l'échelle des paires de bases, et ii) de réduire le temps de calcul par un facteur 2-5. Je voudrais aussi préciser que la méthode TI n'est pas une méthode de simulation. Il s'agit en fait d'un outil de calcul théorique numérisé, qui vise à évaluer de façon directe la fonction de partition. Cet aspect rend cette méthode plus appropriée que la dynamique moléculaire pour les études théoriques basées sur une évaluation de la fonction de partition.

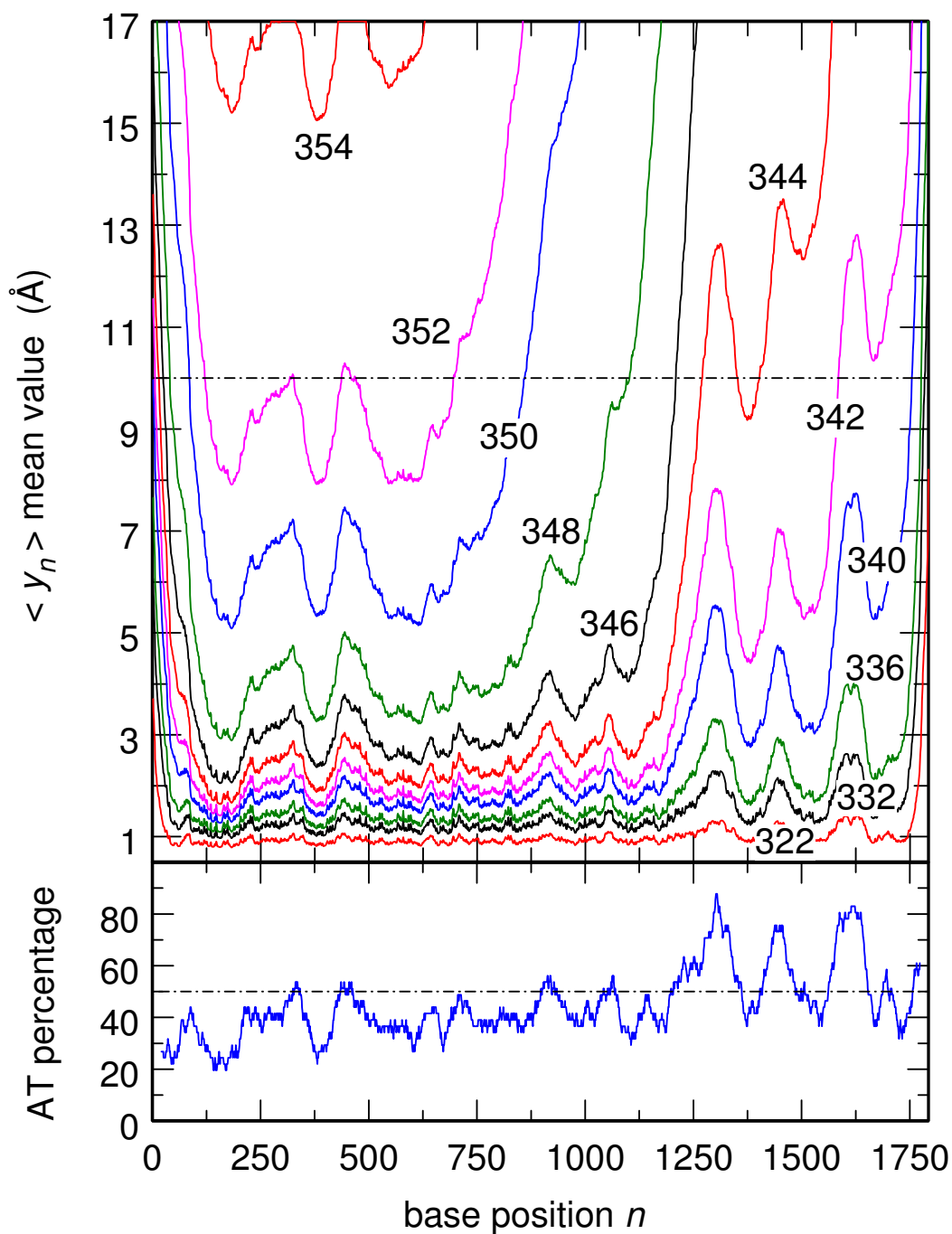


Fig. 6.3: En haut, profils de dénaturation de la séquence de 1793 paires de bases (code d'entrée NCB NM 001101) calculés par la méthode TI, pour le modèle JB hétérogène (6.2) et des températures qui varient entre 322 K et 354 K. En bas, pourcentage de AT moyenné sur 40 paires de bases successives.

Le processus de dénaturation de cette séquence a été décrit par des simulations de dynamique moléculaire dans le chapitre 2 et plus particulièrement les figures 2.5 et 2.10. En haut de la figure 6.3, j'ai présenté les profils de dénaturation calculés par la méthode TI pour des températures qui varient entre 322 K et 354 K et en bas de cette figure le pourcentage de AT moyenné sur 40 paires de bases successives. On remarque qu'il s'agit ici d'une séquence dont le pourcentage de AT augmente régulièrement avec n . La dissociation démarre légèrement au-dessus de $T = 320$ K par l'apparition de trois bulles situées autour de $n = 1300$, $n = 1450$ et $n = 1600$. Ces bulles correspondent exactement aux régions de la séquence les plus riches en AT (voir la courbe du bas). La dénaturation d'une première grande section de 600 paires de bases a lieu au voisinage de $T = 346$ K. Le reste de la séquence se dissocie brusquement à $T = 353$ K. Ce processus est également illustré par la courbe de dénaturation de la figure 6.4 obtenue par la méthode TI (comparer à la courbe de la figure 2.11). Il s'agit donc d'une dissociation à deux étapes. Afin d'illustrer l'effet de cette caractéristique sur les propriétés critiques de cette séquence d'ADN, j'ai calculé par la méthode TI sa chaleur spécifique c_V , que j'ai présentée en haut la figure 6.5. On constate que c_V possède deux pics, correspondant à deux points critiques distincts, séparés par un intervalle de 7 K. La position des deux pics correspond bien aux températures de dissociation des deux portions de la séquence. Cela montre sans ambiguïté que l'hétérogénéité de la séquence peut conduire à plusieurs transitions de dénaturation distinctes.

J'ai également calculé la chaleur spécifique pour une autre séquence de 2399 paires de bases, à savoir l'inhibiteur de l'activateur du facteur d'amplification de l'hépatocyte [35] dont j'avais représenté le profil de dénaturation, calculé par des simulations MD pour le modèle JB à $2N$ degrés de liberté (2.15), sur la figure 3.7. En observant ces courbes, on remarque qu'il s'agit également d'une séquence qui se dissocie en deux étapes, avec des températures de dissociation qui sont nettement plus rapprochées que celles de la séquence actine. On constate en bas de la figure 6.5 que cette proximité des températures critiques conduit à une asymétrie de la chaleur spécifique pour cette séquence plutôt

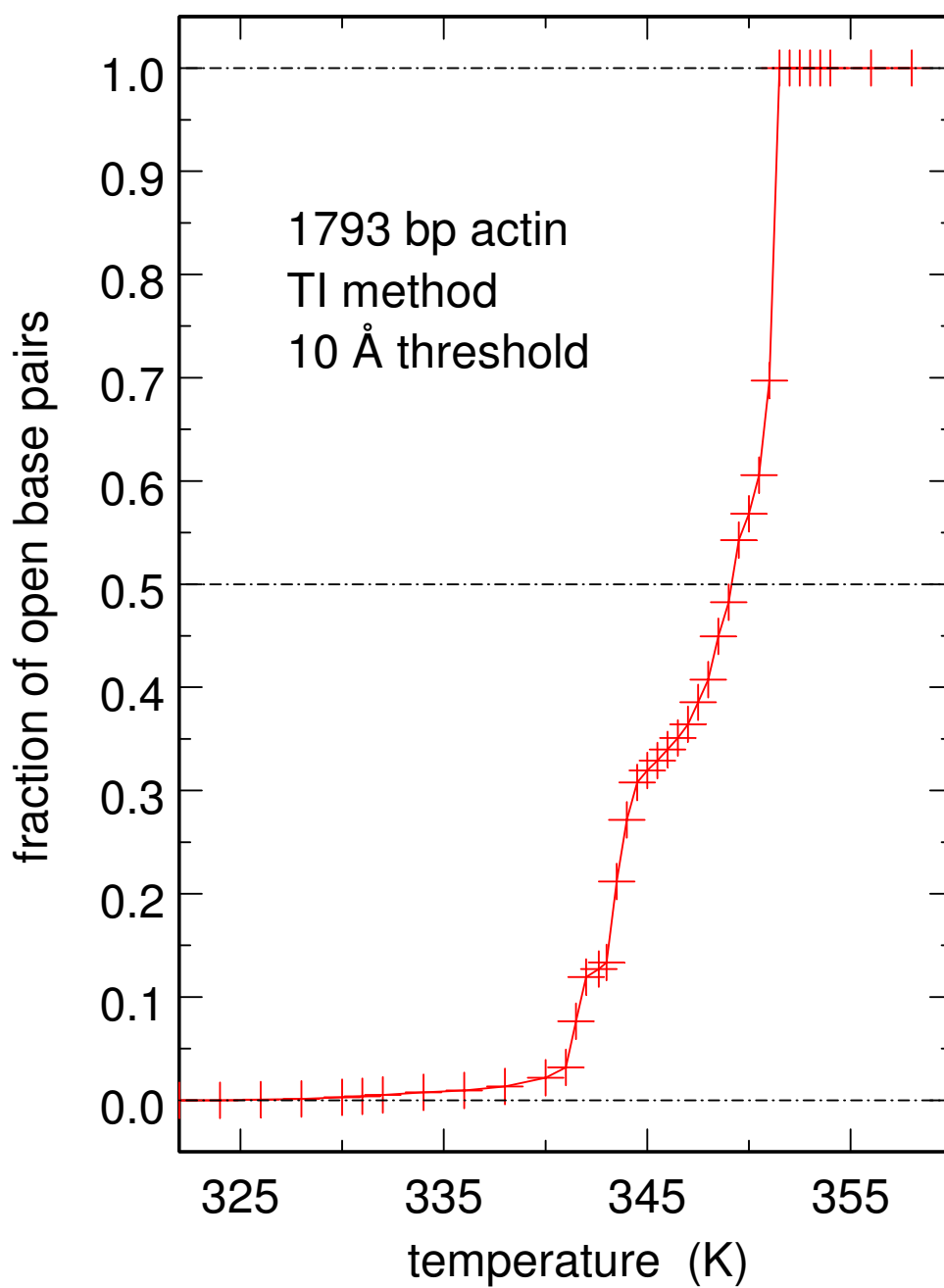


Fig. 6.4: Fraction de paires de bases ouvertes de la séquence de 1793 paires de bases (code d'entrée NCB NM 001101) en fonction de la température, calculée par la méthode TI pour le modèle (6.2). L'axe des ordonnées correspond à la probabilité $P(\langle y_n \rangle > 10 \text{ \AA})$.

qu'à la présence de deux pics. On peut en déduire que dans le cas où les températures de dissociation ne sont pas suffisamment distantes, l'hétérogénéité peut lisser la transition de phase au lieu de conduire à plusieurs transitions de phase distinctes.

Dans ce chapitre, j'ai présenté un travail qui est en cours et discuté les résultats déjà obtenus. J'ai appliqué la méthode TI modifiée au modèle JB hétérogène et montré que cette méthode est capable de reproduire pour ce modèle les détails microscopiques du processus de dénaturation pour une séquence hétérogène. En calculant par cette méthode modifiée les chaleurs spécifiques de deux séquences de structures génétiques distinctes pour le modèle JB (6.4), j'ai montré que l'hétérogénéité d'une séquence peut lisser la transition à travers une asymétrie de la singularité, ou bien conduire à plusieurs transitions de phase si les températures de dissociation sont suffisamment séparées. La suite de ce travail portera sur le calcul de certains exposants critiques caractérisant la transition de dénaturation des séquences hétérogènes, afin d'obtenir une certitude quantitative quant à l'ordre de la transition de phase de l'ADN en présence de désordre.

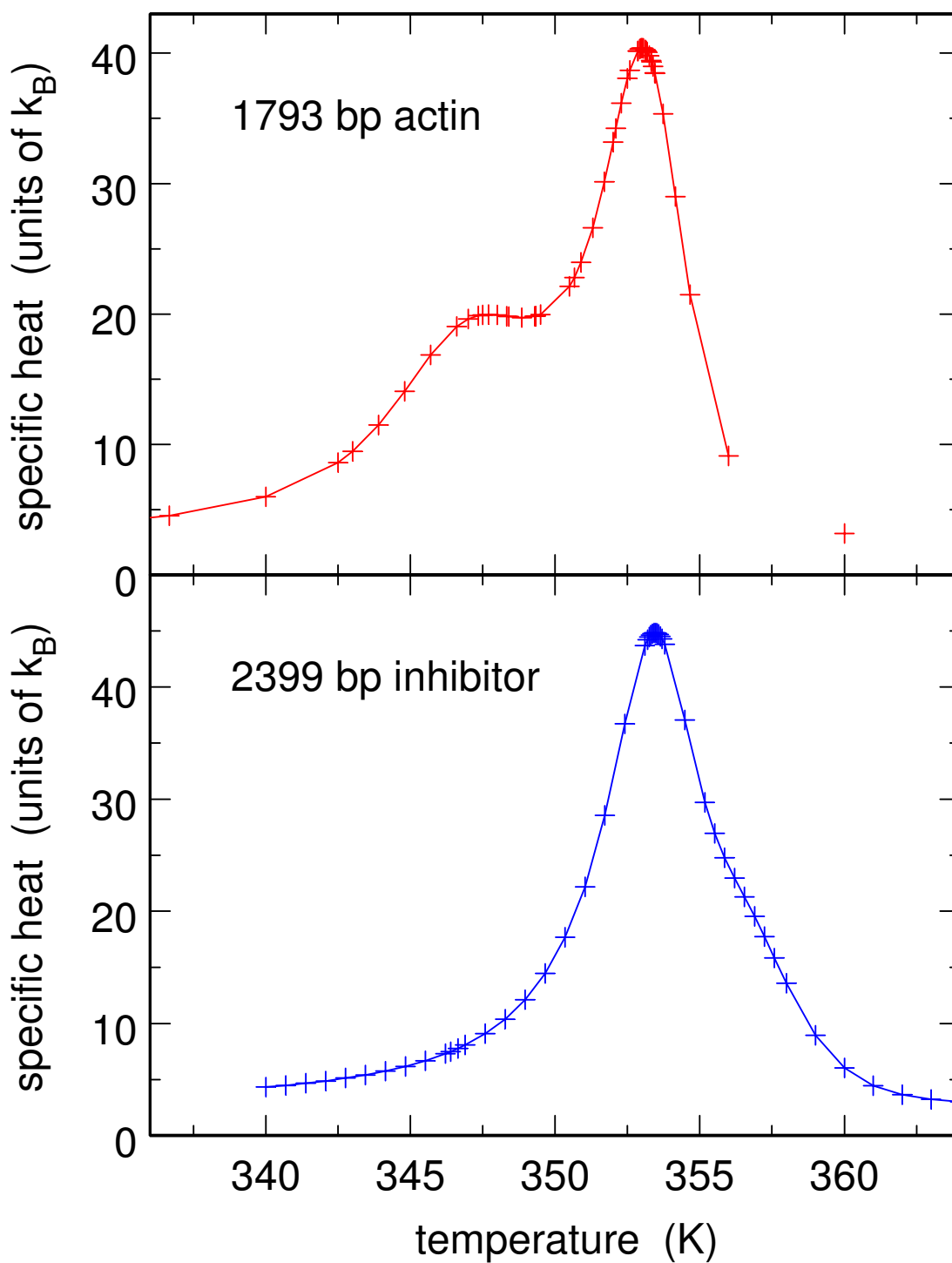


Fig. 6.5: Chaleur spécifique par particule en fonction de la température, calculée par la méthode TI, pour le modèle JB hétérogène (6.2) et la séquence de 1793 paires de bases (code d'entrée NCB NM 001101) (en haut) et une autre séquence de 2399 paires de bases [35] (en bas).

CONCLUSION

L'axe principal de ce travail de thèse était la mise au point et l'étude théorique de modèles d'ADN non-linéaires capables de décrire la dynamique de ce système sur une plage de températures étendue, afin de mieux comprendre son processus de dénaturation thermique. Les trois premiers chapitres ont été en grande partie consacrés à la dynamique et les chapitres 4, 5 et 6 à la mécanique statistique de l'ADN. Nous avons commencé par proposer un modèle d'ADN unidimensionnel basé sur des énergies d'empilement finies et dépendantes de la séquence et montré que ce modèle reproduit très bien les aspects caractéristiques des résultats expérimentaux. Nous avons ensuite amélioré ce modèle en tenant compte des degrés de liberté de torsion. Une observation expérimentale récente de Nagapriya et al. [21], à savoir l'amplification des fluctuations de température de l'ADN à la dissociation, a été étudiée et reproduite dans un contexte théorique. Dans le cadre de ce même travail, nous avons mis en évidence la dépendance en $1/f$ du spectre des fluctuations de température et proposé un début d'explication quant à l'origine de cette dépendance. Nous avons ensuite effectué une étude complète du comportement critique des modèles d'ADN homogènes à la limite thermodynamique. Pour cela, nous avons calculé les exposants critiques pour deux modèles, démontré quantitativement la discontinuité de la transition de dénaturation de l'ADN et montré que deux lois d'échelle, parmi les quatre habituellement utilisées, ne sont pas vérifiées par l'ADN. Enfin, nous avons analysé, pour les modèles homogènes, les effets de finitude de taille sur la transition de phase de l'ADN et testé les hypothèses de la théorie de *finite size scaling*. Nous avons étudié le comportement des fonctions thermodynamiques et de la température de dissociation en fonction de la taille du système et montré en particulier que l'évolution de la température critique est en parfait accord avec le comportement en $1/N$ observé

expérimentalement.

Ces résultats ont été obtenus en utilisant des variantes de la méthode de l'intégrale de transfert, ainsi qu'en intégrant numériquement les équations de Hamilton et de Langevin. Les langages de programmation que nous avons utilisés sont Mathematica et Pascal. Nous avons écrit la totalité des programmes, à l'exception de l'intégrateur de Runge-Kutta d'ordre 4 et du générateur de nombres aléatoires. Les calculs numériques ont été effectués sur des mono et bi-processeurs Xeon cadencés entre 2.8 et 3.6 GHz.

Il y a au moins trois problématiques, qui découlent plus ou moins directement du travail déjà effectué, dont j'aimerais achever l'étude ou sur lesquelles j'aimerais me pencher dans un futur proche. La première, dont j'ai déjà tracé les grandes lignes dans le chapitre 6, concerne les séquences hétérogènes. Dans les chapitres 4 et 5, j'ai exposé notre travail sur la mécanique statistique d'équilibre de la transition de phase des séquences d'ADN homogènes. Or, les molécules d'ADN naturelles correspondent à des séquences hétérogènes. Le fait d'étudier une séquence homogène permet de comprendre les propriétés propres à un seul point critique, tandis que les courbes de dénaturation expérimentales montrent que l'hétérogénéité peut conduire à une transition de dénaturation qui a lieu en plusieurs étapes, chaque étape correspondant à une dissociation locale. Il est donc important de comprendre les effets du désordre sur le comportement critique du système et de vérifier si le caractère discontinu de la transition de phase des séquences homogènes, que nous avons démontré quantitativement dans le chapitre 4, est maintenu en présence du désordre. Au dernier chapitre de cette thèse, nous avons montré que l'hétérogénéité de la séquence peut effectivement conduire à plusieurs points critiques correspondant à des dissociations locales et que cela se manifeste sur le tracé de la chaleur spécifique par l'apparition de plusieurs pics ou d'un seul pic asymétrique. Il faut maintenant estimer l'ordre de la transition pour chacun de ces points critiques en calculant les exposants associés.

Un autre problème que nous avons commencé à étudier concerne la dénaturation forcée de l'ADN. Les expériences de traction qui étudient les propriétés élastiques des

séquences d'ADN homogènes et hétérogènes ont mis en évidence un phénomène intéressant, à savoir la *dénaturation mécanique* de l'ADN. Lorsque l'un des deux brins d'ADN est fixé à un support et l'autre brin tiré avec une force constante, la molécule se dissocie d'un coup à une force critique F_c . La dénaturation mécanique de l'ADN est un très joli problème de physique statistique. Puisque les résultats expérimentaux, comme les courbes d'extension ou le diagramme des phases (F_c en fonction de T) [83, 84], sont calculables par une méthode TI adaptée à ce problème, cette étude théorique permettrait en premier lieu de tester de façon poussée notre modèle pour des séquences homogènes et hétérogènes. D'autre part, on peut également envisager d'estimer du point de vue théorique l'ordre de la transition de dénaturation mécanique de l'ADN. Les propriétés critiques de l'ADN sous contrainte externe pourraient également s'avérer instructives sur un plan plus biologique, dans la mesure où il y a une certaine similarité entre la réplication et la dénaturation mécanique.

Le dernier point concerne justement l'application de l'approche décrite dans le paragraphe précédent à la compréhension de l'interaction ADN-protéines. L'un des buts les plus ambitieux de la modélisation des processus génétiques est la description des processus de réplication et de transcription en tenant correctement compte de l'interaction réelle entre la dynamique intrinsèque de l'ADN et celle des moteurs moléculaires. En fait, un moteur moléculaire doit écarter le double brin pour pouvoir avancer et, dans la plupart des études qui ont précédé, cette dynamique est décrite par des énergies effectives intégrées dans un modèle cinétique où la protéine diffuse sur un espace discret, c'est à dire d'un site à un autre (voir par exemple [81, 82]). Une première piste à explorer consisterait à utiliser dans ces modèles discrets des énergies réalistes dépendant de la température, c'est à dire des spectres obtenus en calculant l'énergie nécessaire pour écarter une ou plusieurs paires de bases pour une séquence particulière et une température donnée. Le pas suivant serait bien évidemment de proposer un Hamiltonien décrivant, en fonction de variables continues, la dynamique de protéines couplées, par exemple, à notre modèle d'ADN.

BIBLIOGRAPHIE

- [1] Molecular structure of Nucleic Acids, James D. Watson et Francis H. Crick. Nature 171, 737 (1953).
- [2] R. Thomas, Bull. Sté. Chim. Biol., 7, 609 (1953)
- [3] R. Tomas, Biochimica et Biophysica Acta, 14, 231 (1954)
- [4] J. Marmur, R. Rownd et C. L. Schildkraut, Prog. Nucleic Acid Res. 1, 231 (1963)
- [5] D. Poland et H. A. Scheraga, J. Chem. Phys. 45, 1456 (1966); 45, 1464 (1966).
- [6] R. Blossey et E. Carlon, Phys. Rev. E 68, 061911 (2003)
- [7] Y. Gao et E. W. Prohofsky, J. Chem. Phys, 80,2242 (1984)
- [8] M. Peyrard et A. R. Bishop, Phys. Rev. Lett. 62, 2755 (1989).
- [9] T. Dauxois, M. Peyrard et A. R. Bishop, Phys. Rev. E 47, R44 (1993).
- [10] S. Cocco, M. Barbi, et M. Peyrard, Phys. Lett. A, 253, 161 (1999).
- [11] T. Dauxois et M. Peyrard. Physics of Solitons, Nonlinear Science and Fluid Dynamics, Cambridge, 2006.
- [12] S. Komineas, G. Kalosakas et A. R. Bishop, Phys. Rev. E, 65, 061905 (2002)
- [13] G. Kalosakas, K. Rasmussen et A. R. Bishop, J. Chem. Phys, 118, 3731 (2003)
- [14] S. Ares, N. K. Voulgarakis, K. Rasmussen et A. R. Bishop, Phys. Rev. Lett. 94, 035504 (2005)
- [15] Y.-L. Zhang, W.-M. Zheng, J. X. Liu et Y. Z. Chen, Phys. Rev. E 56, 7100 (1997).

- [16] T. Michoel et Y. Van de Peer, *Phys. Rev. E* 73, 011908 (2006)
- [17] R. D. Blake, J. W. Bizzaro, J. D. Blake, G. R. Day, S. G. Delcourt, J. Knowles, K. A. Marx et J. SantaLucia, *Bioinformatics* 15, 370 (1999).
- [18] D. Poland, *Biopolymers* 13, 1859 (1974)
- [19] E. Carlon, E. Orlandini et A.L. Stella, *Phys. Rev. Letters* 88, 198101 (2002)
- [20] T. Garel et C. Monthus, *Journal of Statistical Mechanics-Theory and Experiment*, Art. No. P06004 (2005)
- [21] K.S. Nagapriya, A.K. Raychaudhuri et D. Chatterji, *Phys. Rev. Lett.* 96, 038102 (2006)
- [22] T. Geisel, A. Zacherl et G. Radons, *Phys. Rev. Lett.* 59, 2503 (1987)
- [23] Y. Imry et M. Wortis, *Phys. Rev. B*, 19, 3580 (1979)
- [24] O. Gotoh, *Adv. Biophys.* 16, 3 (1983).
- [25] S. W. Englander, N. R. Kallenbach, A. J. Heeger, J. A. Krumhansl, et S. Litwin, *Proc. Natl. Acad. Sci. USA*, 77, 7222 (1980).
- [26] S. Yomosa, *Phys. Rev. A*, 30, 474 (1984).
- [27] C. T. Zhang et K. C. Chou, *Chem. Phys.*, 191, 17 (1995).
- [28] G. Gaeta, *Phys. Lett. A* 168, 383 (1992)
- [29] G. Gaeta, *Phys. Lett. A* 190, 301 (1994)
- [30] G. Gaeta, *J. Biol. Phys.* 24, 81 (1999)
- [31] G. Gaeta, *Phys. Rev. E* 74, 021921 (2006)
- [32] M. Techera, L.L. Daemen et E.W. Prohofsky, *Phys. Rev. A* 40, 6636 (1989)
- [33] M. Joyeux et S. Buyukdagli, *Phys. Rev. E* 72, 051902 (2005).
- [34] A. Brünger, C. B. Brooks et M. Karplus, *Chem. Phys. Lett.* 105, 495 (1984).
- [35] T. Shimomura, K. Denda, A. Kitamura, T. Kawaguchi, M. Kito, J. Kondo, S. Kagaya, L. Qin, H. Takata, K. Miyazawa et N. Kitamura, *J. Biol. Chem.* 272, 6370 (1997).

- [36] E. Carlon, M. L. Malki et R. Blossey, *Phys. Rev. Lett.* 94, 178101 (2005).
- [37] A. Campa et A. Giansanti, *Phys. Rev. E* 58, 3585 (1998).
- [38] S. Buyukdagli, M. Sanrey et M. Joyeux, *Chem. Phys. Lett.* 419, 434 (2006)
- [39] M. Peyrard, *Nonlinearity* 17, R1 (2004).
- [40] S. Buyukdagli and M. Joyeux, *Phys. Rev. E* 73, 051910 (2006)
- [41] R. Balian, *From microphysics to macrophysics - methods and applications of statistical physics* (Springer-Verlag, Berlin, 1991)
- [42] J. Barré et T. Dauxois, *Europhys. Lett.* 55, 164 (2001)
- [43] N. Theodorakopoulos, T. Dauxois et M. Peyrard, *Phys. Rev. Lett.* 85, 6 (2000).
- [44] A.R. Bizzarri et S. Cannistraro, *Phys. Lett. A* 236, 596 (1997)
- [45] M. Takano, T. Takahashi et K. Nagayama, *Phys. Rev. Lett.* 80, 5691 (1998)
- [46] A.R. Bizzarri et S. Cannistraro, *Physica A* 267, 257 (1999)
- [47] P. Carlini, A.R. Bizzarri et S. Cannistraro, *Physica D : Nonlinear phenomena* 165, 242 (2002)
- [48] B. Derrida, *Physics Reports*, 103, 29 (1984)
- [49] M. Aizenman, J. Wehr, *Physical Review Letters*, 62, 2503 (1989)
- [50] A. Nihat Berker, *J. Appl. Phys.* 70, 5941 (1991)
- [51] M. Le Bellac, *Des phénomènes critiques aux champs de jauge*, EDP Sciences (1998)
- [52] T. Schneider et E. Stoll, *Phys. Rev. B* 22, 5317 (1980).
- [53] T. Dauxois, N. Theodorakopoulos et M. Peyrard, *J. Stat. Phys.* 107, 869 (2002).
- [54] B. D. Josephson, *Proc. Phys. Soc. London* 92, 269 (1967).
- [55] A. D. Sokal, *J. Stat. Phys.* 25, 25 (1981).
- [56] I.E. Scheffler, E.L. Elson et R.L. Baldwin, *J. Mol. Biol.* 48, 145 (1970)

- [57] J.-R. Sarasua, R.E. Prud'homme, M. Wisniewski, A. Le Borgne et N. Spassky, *Macromolecules*, 31, 3895 (1998)
- [58] M. Kuwahara, M. Arimitsu, M. Shigeyasu, N. Saeki et M. Sisido, *J. Am. Chem. Soc.* 123, 4653 (2001)
- [59] D. De Luchi, C. Gouyette et J.A. Subirana, *Analytical Biochemistry* 322, 279 (2003)
- [60] K.J. Breslauer, R. Frank, H. Blöcker et L.A. Marky, *Proc. Natl. Acad. Sci. USA* 83, 3746 (1986)
- [61] J.G. Duguid, V.A. Bloomfield, J.M. Benevides et G.J. Thomas, *Biophys. J.* 71, 3350 (1996)
- [62] R.M. Myers, S.G. Fischer, T. Maniatis et L.S. Lerman, *Nucleic Acids Res.* 13, 3111 (1985)
- [63] J. Zhu et R.M. Wartell, *Biochemistry* 36, 15326 (1997)
- [64] A.M. Paiva et R.D. Sheardy, *Biochemistry* 43, 14218 (2004)
- [65] S. Amrane, B. Sacca, M. Mills, M. Chauhan, H.H. Klump et J.-L. Mergny, *Nucleic Acids Res.* 33, 4065 (2005)
- [66] Y. Zeng, A. Montrichok et G. Zocchi, *Phys. Rev. Lett.* 91, 148101 (2003)
- [67] Y. Zeng, A. Montrichok et G. Zocchi, *J. Mol. Biol.* 339, 67 (2004)
- [68] S. Ares, N.K. Voulgarakis, K.O. Rasmussen et A.R. Bishop, *Phys. Rev. Lett.* 94, 035504 (2005)
- [69] T.S. van Erp, S. Cuesta-Lopez et M. Peyrard, *Eur. Phys. J. E* 20, 421 (2006)
- [70] M.E. Fisher et M.N. Barber, *Phys. Rev. Lett.* 28, 1516 (1972)
- [71] H.W.J. Blöte et E. Luijten, *Europhys. Lett.* 38, 565 (1997)
- [72] E. Luijten, K. Binder et H.W.J. Blöte, *Eur. Phys. J. B* 9, 289 (1999)
- [73] Z. Merdan et R. Erdem, *Phys. Letters A* 330, 403 (2004)
- [74] S. Clar, B. Drossel, K. Schenk et F. Schwabl, *Phys. Rev. E* 56, 2467 (1997)

- [75] M. Masihi, P.R. King et P. Nurafza, Phys. Rev. E 74, 042102 (2006)
- [76] K. Binder et D.P. Landau, Phys. Rev. B 30, 1477 (1984)
- [77] V. Privman, in "Finite size scaling and numerical simulation of statistical systems", edited by V. Privman (World Scientific, Singapore, 1990)
- [78] K. Binder, M. Nauenberg, V. Privman, A.P. Young, Phys. Rev. B 31, 1498 (1985)
- [79] A. B. Harris, J. Phys C 7, 1671 (1974)
- [80] B. Coluzzi et E. Yeramian, Philosophical Magazine, 87, 517 (2007)
- [81] F. Julicher et J. Prost, Phys. Rev. Lett, 75, 4510 (1997)
- [82] Y. Kafri, D. Lubensky et D. Nelson, Biophysical Journal, 86, 3373 (2004)
- [83] U. Bockelmann, Ph. Thomen, B. Essevaz-Roulet, V. Viasnoff et F. Heslot, Biophysical Journal, 82, 1537 (2002)
- [84] R. Krautbauer, M. Rief et H. E. Gaub, Nano Letters, 3, 493 (2003)

RÉSUMÉ

Résumé : L'axe principal de ce travail de thèse est la mise au point et l'étude théorique de modèles d'ADN non-linéaires afin de mieux comprendre le processus de dénaturation thermique. Nous avons commencé par proposer un modèle d'ADN unidimensionnel basé sur des énergies d'empilement finies et dépendantes de la séquence et montré que ce modèle reproduit très bien les aspects caractéristiques des résultats expérimentaux. Nous avons ensuite amélioré ce modèle en tenant compte de certains degrés de liberté de torsion. Une observation expérimentale, à savoir l'amplification des fluctuations de température de l'ADN à la dissociation, a été étudiée et reproduite dans un contexte théorique. Dans le cadre de ce même travail, nous avons mis en évidence une dépendance en $1/f$ du spectre des fluctuations de température et proposé une explication à cette observation. Nous avons ensuite effectué une étude complète du comportement critique des modèles d'ADN homogènes à la limite thermodynamique. Pour cela, nous avons calculé les exposants critiques pour deux modèles d'ADN, démontré quantitativement la discontinuité de la transition de dénaturation et montré que deux lois d'échelle, parmi les quatre habituellement utilisées, ne sont pas vérifiées. Enfin, nous avons analysé, pour les modèles homogènes, les effets de finitude de taille sur la transition de phase de l'ADN et testé les hypothèses de la théorie de finite size scaling. Nous avons étudié le comportement des fonctions thermodynamiques et de la température de dissociation en fonction de la taille du système et montré en particulier que l'évolution de la température critique est en parfait accord avec le comportement en $1/N$ observé expérimentalement. Nos études en cours portent sur l'effet du désordre sur la transition de dénaturation de l'ADN.

Abstract : The main goal of this thesis consists in proposing new non-linear DNA models and studying their properties in a theoretical context, in order to better understand the thermal denaturation process. We first proposed a one dimensional DNA model based on finite, site dependent stacking enthalpies. We showed that this model correctly reproduces the main characteristics of experimental melting curves. We then improved it by taking torsional degrees of freedom into account. Using these models, we confirmed a recent experimental observation, namely the amplification of temperature fluctuations of DNA at thermal denaturation. We furthermore pointed out the $1/f$ dependence of the power spectrum of temperature fluctuations and proposed a possible explanation for this result. In the second part of this PhD work, we performed a complete statistical analysis of the critical behaviour of homogeneous DNA models, quantitatively checked the discontinuous character of the melting transition, and showed that two scaling laws (out of four) are not satisfied. Finally, we investigated finite-size effects on the melting transition of homogeneous DNA models and tested the hypotheses on which the finite size scaling theory relies. We also studied the behaviour of thermodynamical functions and the melting temperature as a function of the sequence length and showed that the evolution of the melting temperature perfectly matches the experimental $1/N$ dependence. Ongoing work aims at understanding the effects of the heterogeneity of natural DNA sequences on their melting transition.