

**Conformational sampling by Molecular Mechanics and Dynamics
simulations applied to the flexibility of Nucleic Acids**

by

Jérémy Curuksu

A thesis submitted in partial fulfilment
of the requirements for the degree of

Doctor of Philosophy in Biology

Approved, Thesis Committee:

Prof. Dr. Martin Zacharias, Chair

Prof. Dr. Krystyna Zakrzewska

Prof. Dr. Philippe Derreumaux

Prof. Dr. Mathias Winterhalter

Prof. Dr. Ulrich Kleinekathöfer

Date of Defense: January 27, 2009

School of Engineering and Science

Jacobs University, Bremen, Germany. Université Denis Diderot, Paris, France.
Funding sources from the Volkswagen foundation.

*To my parents, my brother
and former teachers*

Acknowledgements

I would like to express the best of my thankfulness to my PhD advisor Martin Zacharias, for his continuous support and his presence throughout the period of my PhD studies at the Jacobs University in Bremen, Germany. Following some directives in front of obstacles may only appear rewarding a long time afterward in research, but when they become appreciated they become so all the more that they were the fruit of scientific acquaintance.

For sharing such acquaintance I also want to thank Krystyna Zakrzewska (co-advisor of my PhD) and Richard Lavery. I occupied three months within their laboratory in Lyon and these were a highly profitable personal development in research. With marked thanks to Prof. Lavery for having offered to me a project at the first place, within his laboratory in Paris.

I acknowledge each member of the Jury herein Mathias Winterhalter, Philippe Derreumaux and Ulrich Kleinekathöfer, for having agreed on appraising the content of my PhD thesis.

There are different persons that I would thank for discussions or comments which contributed to extend my motivation in the realm of nucleic acid structures, as well as in the arenas of biomolecular computer simulations. In particular John Maddocks, Jiri Sponer, Georgi Muskhelishvili, Victor Zhurkin, Carlos Simmerling and Martin Karplus who afforded me open discussions on fascinating scientific subjects. For the others, I hope it has been possible to find a reciprocal contentment from both scientific and non scientific interactions that occurred with me during the graduate period of my academic studies. If not, may the readings of this PhD thesis fill the gap.

Institutions where the work was carried-out, founding sources as well as more specific acknowledgments are mentioned where appropriate in the following chapters.

CONTENTS

Acknowledgements	3
List of Figures	7
List of Tables	9
Abstract, English version	10
Abstract, French version	11
General Introduction	15
Goal of my PhD	18
I Structure and flexibility of nucleic acids	20
1.1 Chemical composition of nucleic acids	20
1.2 Polymorphism	26
1.3 Analysis of nucleic acid conformations	27
1.3.1 <i>Transformation operations based on three Euler angles</i>	27
1.3.2 <i>Alternative models for the transformation operations</i>	29
1.4 On the conformational flexibility of DNA	32
1.4.1 <i>Protein/DNA interactions</i>	32
1.4.2 <i>DNA bending and curvature</i>	36
1.4.3 <i>The phosphodiester backbone</i>	40
1.4.4 <i>DNA mismatches and abasic sites</i>	41
1.5 On the conformational flexibility of RNA	43
1.5.1 <i>Non Watson-Crick base pairing</i>	43
1.5.2 <i>Overview of RNA motifs</i>	44

II	Computer simulations in molecular mechanics	49
2.1	Nature and Implication of complexity when sampling the conformational landscape of biomolecules	49
2.2	Energy function	50
2.3	Energy minimisation	53
2.4	Molecular Dynamics simulation: a finite difference method	55
2.5	Molecular Dynamics sampling of Ensembles	56
2.6	Practical aspects and standard protocol	58
2.7	The challenge	62
2.8	Umbrella sampling simulations	63
2.9	Replica exchange sampling methods	64
2.9.1	<i>Temperature REM</i>	65
2.9.2	<i>Hamiltonian REM</i>	67
III	RNA bending and the flexibility of ribosomal kink-turn motifs	68
3.1	Abstract	68
3.2	Introduction	69
3.3	Method	70
3.4	Results	72
3.5	Discussion	80
IV	DNA bending, curvature and the spontaneous occurrence of a Kink between two base pairs	83
4.1	Abstract	83
4.2	Introduction	84
4.3	Method	87
4.4	Results	92
4.5	Discussion	120

V	Flexibility of the phosphodiester backbone: metastable ϵ/ζ conformations in severely bent and/or damaged DNA	125
5.1	Abstract	125
5.2	Introduction	126
5.3	Method	127
5.4	Results	133
5.5	Discussion	142
VI	Variation on the method: replica-sampling with a server-client relation.	144
6.1	Abstract	144
6.2	Introduction	145
6.3	Background	145
6.4	Method	147
6.5	Results	153
6.6	Discussion	158
	General Conclusion	160
	References	164
	List of Publications	185
	Annexes	186

List of Figures

- 1.1 Building blocks of nucleic acids
- 1.2 Sugar puckers of nucleic acids
- 1.3 Backbone dihedral angles of nucleic acids
- 1.4 Oligomerisation of nucleotide residues and common shortage notations
- 1.5 Standard B-form dsDNA and A-form dsRNA
- 1.6 Helical parameters used to describe nucleic acids conformations
- 1.7 Global and local description of helical parameters
- 1.8 Space-fixed axis of a total rotation between two orthogonal frames
- 1.9 Molecular model of the looped upstream region of the tyrT Promoter in *E. coli*
- 1.10 Abasic site in DNA
- 1.11 Mismatched base pairing in DNA
- 1.12 Non standard base pairing in RNA and the A-minor interaction
- 1.13 Cloverleaf structure of tRNA and kink-turn motif
- 2.1 Standard stages of the molecular dynamics recipe
- 2.2 Schematic representation of the molecular dynamics phase space
- 2.3 Time evolution in molecular dynamics and replica exchange simulations
- 3.1 Sequence, base pairing and A-minor interactions of the kt38 structure
- 3.2 Potential of mean force (PMF) for the opening of the kt38 kink-turn motif
- 3.3 RMSD with respect to starting structure during umbrella sampling of the kt38 opening
- 3.4 Kink angle and twist angle between C and NC stems during the kt38 opening
- 3.5 Correlation of the inter-stem kink and twist angles during the kt38 opening
- 3.6 Structural motifs characterizing the closed, semi-open and open conformations of kt38
- 3.7 The two A-minor interactions which stabilize kt38 in the semi-open conformation
- 3.8 Final structure corresponding to a fully open geometry of kt38
- 4.1 Definition of a DNA bending angle restraining coordinate
- 4.2 Correlation of the current DNA bend angles with the Curves parameters
- 4.3 Bending angle probability distributions in adiabatic mapping energy minimization
- 4.4 Comparison of two implicit solvent models in adiabatic mapping energy minimization
- 4.5 Influence of initial DNA bending direction in adiabatic mapping energy minimization
- 4.6 Helical parameters for the most stable bent conformations
- 4.7 Illustration of steric clash in the most stable bent conformations

- 4.8 Iso-probability surfaces for directional DNA bending: usual DNA sequences
- 4.9 Iso-probability surfaces for directional DNA bending: unusual DNA sequences
- 4.10 Free energy for bending of an A-tract DNA in umbrella sampling (parm94)
- 4.11 Helical parameters of an A-tract DNA in umbrella sampling (parm94)
- 4.12 Correlated changes of global conformational variables upon induced DNA bending
- 4.13 Free energy for bending of DNA oligomers in umbrella sampling (parmbse0)
- 4.14 Mean base pair hinge angle of DNA oligomers in umbrella sampling (parmbse0)
- 4.15 Representative conformational snapshots upon induced DNA bending
- 4.16 Illustration of type I and type II kinks in DNA
- 4.17 Stereo view of kinks observed in umbrella sampling (parmbse0)
- 4.18 Base pair bending force constants in umbrella sampling (parmbse0)
- 4.19 Analysis of replica exchange umbrella sampling for poly-GC oligomer (parm94)
- 4.20 Analysis of replica exchange umbrella sampling for poly-GC oligomer (parmbse0)
- 5.1 Dihedral angle biasing potential in one and two dimensions
- 5.2 Schematic in one dimension of the free energy surface smoothing technique
- 5.3 Convergence and location of sampled dihedral angle centers during D-BP-RexMD
- 5.4 Comparison of backbone ϵ/ζ dihedral angle sampling for adenine dinucleotide system
- 5.5 Convergence rate of current sampling methods
- 5.6 Comparison of backbone ϵ/ζ dihedral angle sampling for the DNA abasic site system
- 5.7 Coupled clusters of (ϵ,ζ) BI/BII substates and bent conformations
- 6.1 Efficiency of D-BP-RexMD vs. E-BP-RexMD
- 6.2 Energy surface smoothing in the SCREM method
- 6.3 Flow chart of the SCREM method
- 6.4 A practical example of the server vs. client relation in the SCREM method
- 6.5 Comparison of backbone ϵ/ζ dihedral angle sampling for adenine dinucleotide system
- 6.6 Comparison of ϵ/ζ sampling in a 2-replicas SCREM simulation.

List of Tables

- 1.I Geometry of A-, B- and Z polymorphs of nucleic acids
- 1.II DNA base pair stacking interaction energy from quantum mechanics
- 1.III DNA bend angle in 71 x-ray crystal protein / DNA complexes
- 1.IV Direct read-out protein/DNA recognition: functional groups available on each groove
- 5.I Cluster analysis of ϵ/ζ dihedral angle sampling of the adenine dinucleotide system
- 5.II Cluster analysis of ϵ/ζ dihedral angle sampling of the DNA abasic site system
- 6.I Cluster analysis of ϵ/ζ dihedral angle sampling of the adenine dinucleotide system
- 6.II Efficiency of current replica exchange methods

Abstract

My PhD dissertation deals with the application of atomic-scale computer simulation (Molecular Dynamics) on different aspects of the conformational flexibility of Nucleic Acids (i.e. DNA, RNA). The probabilistic nature of states which are searched for in molecular dynamics simulation of Nucleic Acids with explicit water involves a very high number of dimensions and is thus associated to the curse of dimensionality. Some transitions between important families of conformational states are rare or absent. I have used diverse statistical and computational methods such as Umbrella Sampling and Replica Exchange to extract entropic properties, e.g. free energy, characterising (i) the bending of DNA on short length-scale, (ii) the folding of recently discovered RNA ribosomal motifs (the *kink-turn* motif) and (iii) backbone dihedral conformations accessible to damaged DNA. One achievement is the reproduction of the experimental curve for the probability of very high bend angles observed for short fragment of DNA which demonstrates a non linear (softer) bending flexibility of DNA. Indeed the results of my thesis predict that DNA kinks (local defects unstacking neighbour basepairs) occur *in vivo* and some of them induce a 90°-turn in the helix. They are associated with a systematic decrease of the local DNA stiffness constant (half an order of magnitude) which was quite unexpected. DNA bending up to 150° on the 5 nm length scale requires on average 12 kcal/mol. It is slightly less expensive, ~10 kcal/mol when a run of consecutive adenines is present (Atract sequences which occur frequently in every genome). The same kind of information is provided for a larger kink-turn motif in RNA (~10 base pairs) for which an almost iso- energetic twist / bend coupled mode has been characterised. Methodological development of Hamiltonian replica exchange sampling techniques enables to characterize several competing DNA backbone conformations accessible to damaged DNA (an abasic site). The method can be used for any combination of backbone dihedral variables which constitute an example of DNA metastable states with very short relaxation time. Despite fast transition such transition is rare or absent during conventional molecular dynamics test-simulations. More generally the PhD thesis presents the development of new methods to tackle the accurate sampling of particular nucleic acid helical propensities, and this is closed with a brief section on an original effort to create a self-learning approach in the context of replica exchange sampling with molecular dynamics.

Abstract (French version)

Les thématiques adressées dans ma Thèse de Doctorat sont d'une part la flexibilité conformationnelle des Acides Nucléiques étudiée par simulations en Dynamique Moléculaire. D'autre part, mes travaux de Thèse adressent deux problèmes principaux rencontrés lors de la pratique des simulations en dynamique moléculaire, à savoir l'échantillonnage limité et le champ de force spécifique utilisé. Trois issues scientifiques qui ont récemment émergées en particulier dans le domaine expérimental à propos de la flexibilité des molécules d'ADN et d'ARN font l'objet de trois chapitres distincts. Premièrement la possibilité d'une dynamique spontanée formant un coude flexible dans les motifs d'ARN ribosomales appelés kink-turns. Deuxièmement la formation spontanée d'un motif coudé très local dans les molécules d'ADN fortement courbées (« kink » entre deux paires de bases). Troisièmement un équilibre entre plusieurs sous états meta-stables dans les angles dièdres du squelette phosphodiester des molécules d'ADN endommagées, dans le cas présent un site abasique (base qui n'a pas de partenaire dans la double hélice d'ADN). Sur le niveau méthodologique nous avons implémenté une coordonnée de réaction spécifique pour chaque projet utilisant la méthode d'Echantillonnage Parapluie (*Umbrella Sampling*), et développé une nouvelle méthode d'échantillonnage dans le domaine des techniques dites en échange de répliques (*Replica Exchange*). Finalement nous montrons qu'en combinant ces deux approches (donc « échantillonnage parapluie en échange de répliques ») il en résulte un échantillonnage plus efficace et plus précis.

Résumé détaillé:

Nos résultats permettent d'avoir une meilleure idée sur les coûts d'énergie libre impliqués dans le repliement et la courbure de divers motifs d'acide nucléique. Les simulations en échantillonnage parapluie sur le motif d'ARN kink-turn *kt-38* utilisant l'ouverture (ou la fermeture) de la seconde interaction dite « A-mineur » en tant que coordonnée de réaction ont indiqué un mouvement en coude impliquant un changement d'énergie libre d'environ 1,5 kcal/mol en faveur de la forme fermée. Les simulations indiquent un changement globale de courbure et de twist entre le brin appelé C (appariement canonique des bases) et le brin appelé NC (pour Non Canonique) qui ont lieu en concert avec l'ouverture de la structure du motif *kt-38* et sont contrôlés par la conformation de la seconde interaction A-mineur. Cette dernière

passer du type I au type 0 dans la conformation semi-ouverte de kt-38. La conformation complètement ouverte, aussi observée expérimentalement, requiert un changement d'énergie libre plus important (4 kcal/mol) et est caractérisée par une disruption complète des motifs A-mineurs. Les faibles coûts d'énergie libre qui ont été calculés supportent l'idée que les motifs kink-turn peuvent constituer des éléments flexibles à l'origine de changements plus globaux entre brin de double hélice dans les molécules d'ARN larges telles que le ribosome.

Nous avons développé une nouvelle contrainte permettant d'induire une courbure dans l'ADN, basée sur la définition d'un axe de visse moyen pour deux segments terminaux dans un oligonucléotide d'ADN. Cette définition permet d'induire systématiquement une déformation de courbure pour différentes molécules d'ADN sans restriction sur la relaxation conformationnelle de l'ADN lors de l'introduction de la contrainte de courbure. Les résultats sur des séquences contenant un *A-tract* (segment de 3 à 6 adénines consécutives) ont prédit une courbure préférentielle vers le petit sillon et des changements en terme de paramètres hélicoïdaux en bon accord avec des résultats expérimentaux. Les résultats sur les mésappariements G:A ou C:C ont prédit une variété significativement élargie de structures possibles montrant différentes directions et amplitudes de la courbure globale par rapport à l'ADN-B canonique. La flexibilité de courbure peut contribuer aux propriétés de reconnaissance des ADN endommagés par des enzymes de réparation. La méthode a été étendue afin d'étudier la courbure de l'ADN pendant les simulations en dynamique moléculaire incluant une concentration minimum d'ions et des molécules d'eau de façon explicite. Nous avons trouvé un angle de courbure optimal proche de la valeur expérimentale pour les différentes séquences étudiées et une corrélation de la courbure globale avec les autres variables globales caractérisant l'ADN. Les fluctuations de ces variables globales n'ont pas changé significativement sous le stress de courbure jusqu'à 100° (en terme de la nouvelle définition correspondant à ~ 80° dans la définition standard du logiciel Curves).

En utilisant le champ de force le plus récent d'Amber, parmbsc0, le changement d'énergie requis pour former un angle de courbure de 90° sur une échelle de 5 nm (~15 paires de bases) est proche de 5 kcal/mol pour les séquences contenant un *Atract* et 6 kcal/mol pour les séquences alternées de bases G et C (appelée [GC]) ou A et T (appelée [AT]). Pour une courbure proche de 150° le coût d'énergie libre est en moyenne 12 kcal/mol pour toutes les séquences à part la séquence appelée [*Atract-1*] (*Atract* situé au milieu du segment d'ADN) pour laquelle ce coût est réduit à 10 kcal/mol. Nous avons trouvé une relation linéaire (non harmonique) pour l'énergie et extrait les coefficients de linéarité pour différentes séquences

de bases, ce qui peut être utilisé dans les modèles appelés Sub-Elastic Chain (SEC). Pour la forte courbure ($>100^\circ$) le comportement des oligomères dépend de sa séquence. [GC] et [Atract-1] produisent le kink type II, [AT] se dénature et [Atract-2] (Atract situé en périphérie du segment d'ADN) répond au stress imposé par une série de petits kinks dans la partie G/C de la séquence. Le kink type II a été observé deux fois, dans [GC] et dans [Atract-1]. Le kink type II dans [GC] est associé à une courbure locale vers le grand sillon et une déformation des paires de bases plus légère que dans le cas de [Atract-1] où ce motif induit une courbure locale vers le petit sillon. Les deux motifs sont associés à un angle de kink d'environ 90° entre des axes linéaires superposés à quatre paires de bases de chaque côté du motif. En utilisant la technique d'échantillonnage parapluie en échange de répliques (REUS) il a été démontré que le kink type II est un motif d'ADN élastique et donc réversible. Au contraire, le kink type I a été observé seulement avec le champ de force parm94 et est apparue comme une déformation plastique liée à la localisation de l'énergie de déformation mécanique (« *strain* ») le long de la séquence, en d'autres termes le type I n'a pas été observé avec la technique REUS.

Nos résultats montrent que, dans les oligomères d'ADN court chaque jonction entre paires de bases se comporte différemment, en fonction de leur contexte dans la séquence, sans avoir recours à des événements rares tels que des motifs coudés (kinks) entre paires de bases. Toutefois ces derniers sont induits systématiquement à certain di- ou trinucleotides dans l'ADN dans le régime de forte courbure. Les « constantes harmoniques de rigidité de courbure » sont couramment utilisés dans la théorie des tubes élastiques afin de modéliser la dynamique à grande échelle des molécule d'ADN et une caractéristique qui a émergée pour chacun des motifs coudés observés (i.e. $\text{roll} > 20^\circ$) dans ces simulations est une constante de rigidité de courbure plus faible (de 4/5 jusqu'à 1/10), ce qui est un résultat surprenant.

Il n'y a pas eu de transition significative dans les angles de torsion du squelette phosphodiester associée avec la courbure de l'ADN dans l'étude qui vient d'être résumée (Chapitre 4 de la thèse). Ensuite (Chapitre 5), une nouvelle méthode en échange de répliques (Hamiltonian Replica Exchange) a été conçue, laquelle utilise un potentiel de biais en tant que coordonnée de réplique qui promeut spécifiquement des transitions dans les angles dièdres du squelette des acide nucléiques. La conclusion dans le chapitre 5 est alors que cette méthode est nécessaire dans les simulations en dynamique moléculaire afin de décrire correctement certaines conformations métastables dans le squelette de l'ADN pour les motifs non standard, un site abasique en l'occurrence, sur les échelles de temps de simulations courts (5 à 20 ns). En conclusion, dans l'étude qui a été décrite précédemment (Chapitre 4), plusieurs sous-états

compatibles avec une conformation courbée donnée, en raison de barrières entre les sous-états conformationnels du squelette, n'ont peut-être pas été tous explorés. L'étude du chapitre 5 permet de caractériser la structure fine de l'ADN double brins systématiquement sous des conditions réaliste et pourrait aussi permettre une évaluation systématique de la performance et la précision des champs de forces actuels. Une autre possibilité serait d'utiliser un des autres champs de force disponibles pour les simulations en dynamique moléculaire des acides nucléiques et de voir si le kink type II est toujours présent. Dans l'affirmative les kinks dans l'ADN, d'après l'outil que constitue les simulations en dynamique moléculaire, pourraient avoir lieu de façon spontanée *in vivo*, confirmant alors une théorie proposée il y a plus de vingt ans par Francis Crick et Aaron Klug.

Il a été démontré que l'association des simulations en échantillonnage parapluie avec l'approche dite en échange de répliques améliore significativement la convergence des calculs d'énergie libre et permet un meilleur fit des simulations *forward* et *backward* en comparaison avec l'échantillonnage parapluie conventionnel. Finalement, le formalisme nommé *Hamiltonian Replica Exchange* a été revu sur la notion que son avantage principal est de produire un Ensemble de conformations canonique dans le réplica non-biaisé. Les répliques biaisées furent considérées étant des copies de recherche dont le but est d'éviter une frustration cinétique dans le réplica non-biaisé. Ce dernier fut considéré étant un « client » dans un schéma où l'échelle des répliques complémentaires pourrait répondre effectivement en fonction de l'état actuel et de l'histoire enregistrée dans le réplica-client. Cela m'a amené à développer un potentiel de biais multiple généré automatiquement, fondé sur la technique dite de recherche Tabu et définie par l'intégration de fonctions de mémoire à court terme et à long terme, lesquelles s'ajoutent à la fonction Hamiltonian des répliques dits « servers ». Malgré des efforts dans cette direction la paramétrisation idéale qui permettrait des applications générales aussi sur des systèmes unusuellement complexes tels qu'un site abasique ou bien un kink type II dans l'ADN est pour l'heure encore une perspective. Considérant les simulation-tests d'un système avec deux répliques (1 Server vs. 1 Client) sur un modèle moléculaire simple cependant, cette technique d'Hamiltonian Replica Exchange adaptative peut en principe être paramétrée en vue d'un outil efficace pour explorer les paysages énergétiques associés à l'espace des phases des angles dièdres des acides nucléiques et biomolécules en générale. Un échantillonnage amélioré a été démontré à travers chacun des chapitres dans le cas des méthodes d'Hamiltonian Replica Exchange standards qui utilisent un potentiel de biais ou un potentiel parapluie.

General Introduction

Structure and flexibility of Nucleic Acid molecules is important for their many biological functions, from the replication of the support of genetic information, to its translation into macromolecular assemblies constituting biological organisms. Many biological functions induced by the bending flexibility of nucleic acids reflect dynamical hinge properties of specific base pair components. This property may extend over a few base pairs which respond collectively into a concerted way and thereby define some flexible hinge *motifs*. Therein two important notions ensue on the conformational flexibility of nucleic acids, which are the notion of *sequence dependence* and the notion of *local versus global* deformability. The fine sequence of nucleic acids can now be easily manipulated (e.g. PCR) within the context of current bio- nano- technologies (1,2) but the local or atomic details are out of reach of a real time observation by any available experimental technique in biology (3). State-of-the-art the unveiling area of single molecule experiments opens the doors to short length scales up to the few DNA/RNA base pair level (4,5,6), producing unexpected results, generating new ideas and bringing the possibility to check mathematical models on this short length scale. Computer simulation of nucleic acids and other biomolecules will thus make an increasingly important contribution to twenty-first century science (3,7), because they stand in this development as an opportunity to develop and test scientific hypothesis at any scale.

Molecular Dynamics (MD) simulation is a sampling technique for a molecular system based on discrete time integration of Newton's second law. The force acting on each particle is derived from a high-dimension potential energy function (*the force field*) describing bonded and non-bonded interactions in the molecular system (typically a box of water molecules with some solute molecules). The elementary particle in classical molecular dynamics is the net-charge atom whereby the electronic shell property and the molecular environment are subsumed based on the Born-Oppenheimer approximation (8). By definition the molecular dynamics simulation is a tool (also called computer experiment) which enables to investigate the dynamics of chemical events as well as entropic properties in a given thermodynamics ensemble (i.e. fixed extensive properties e.g. NVT or NPT). From the probability density of conformation in a sample resulting from molecular dynamics (*a trajectory*) the free energy along a given reaction coordinate can be extracted (9a).

However looking at the statistical mechanical expression for the probability density of conformation in a canonical ensemble (NVT) at equilibrium, it appears that a conformation with relatively high energy is exponentially unlikely to be sampled. Consequently, molecular dynamics simulations of biomolecules are often limited by poor sampling due to kinetic trapping events in low-energy regions of the complex Free Energy Surface (FES). The rugged nature of the FES is in particular due to frequent atomic surface proximity inducing hard core van-der-Walls repulsion and the allowable set of rotamers in the subspace of protein and nucleic acid backbone dihedral angles. This limited sampling at physiological temperature can be improved by several techniques some of which will be explained, applied and further developed in the current thesis. The many recent developments on the methodological front, noteworthy in free energy calculations, is said to have made molecular dynamics simulations come of age (10). I acknowledge this statement and here also append “at the just time” given the recent experimental developments.

Through this thesis I have made use of two standard methods for enhanced sampling and developed them further. The first enhanced sampling method is the Umbrella Sampling (11) which consists to decompose the trajectory into several sampling windows generated independently with an energetic penalty (added to the force-field) that restricts consecutive values of an order parameter. The consecutive sampling windows ensure a good probability of sampling along this order parameter. The sampled probabilities are thermodynamically biased due to the presence of an additional energetic penalty in the force-field. At the cost of losing information about the time evolution (the dynamics), an optimal unbiased probability density along the order parameter is recoverable by statistical methods based on the analysis of weighted histograms (9b). The corrected probability density in turn equates the free energy along the generalized coordinate. Obviously a challenge in umbrella sampling lies in finding generalized coordinates which accurately describe events such as folding or bending of biomolecules (or some of their motifs) because a high number of orthogonal coordinates is involved in these complex conformational events. We begin (Chapter III) by a simple example on the ribosomal kink-turn RNA motif which illustrates how key interactions (the A-minor tertiary interaction) may mediate larger scale conformational changes and as such define a reaction coordinate of free energy calculation for the elbow like dynamics of these motifs. In chapter IV a more elaborate reaction coordinate based on a screw-representation of the double helix is developed. This generalized coordinate defines two handles that can be used on both side of a fragment of interest to monitor its bending and enables systematic

study of the sequence dependent bending properties of nucleic acids. Our application enables to provide stereochemical explanation for a recent controversial issue about the short length scale bending flexibility of DNA. Indeed new experimental observations (by now several) concluded to an enhanced probability of high bend angles compared to the linearly elastic theory of DNA on short length scale (4,5,6,12). We have reproduced the experimental curve for the probability of high bend angles on same length scale and our results predict that DNA kinks (local defects unstacking neighbour base pairs) can occur in vivo, some of them inducing a 90° turn between the two adjacent double helical stems.

The second enhanced sampling method used and further developed in this thesis is Replica Exchange (13). As mentioned, the challenge of finding an accurate generalized coordinate in umbrella sampling reflects the fact that a given umbrella window may not overcome all relevant barriers at the given value of the umbrella coordinate. A good example is given if the umbrella coordinate is an order parameter describing the folding of a non-trivial molecular motif. In such case a sampling window at value of order parameter corresponding to a folded state does not induce transitions to some secondary minima when the latter require the crossing of high energy barrier along eigenvectors *orthogonal* to the umbrella. One solution to enhance sampling is then to couple the independent umbrella windows by the replica exchange method which consists of periodically swapping conformations between adjacent replicas *i* and *j*, every pair *i, j* being tested in turn, according to an exchange probability which satisfies detailed balance. The exchange criterion takes the form of an equation of micro-reversibility which takes into account the conformational probability density of the two replicas *i* and *j*, i.e. Boltzmann factors based on the biased potential energy function of replica *i* and replica *j*. As a result of these periodical exchanges the conformations trapped in a low energy region of the conformational landscape easily diffuse to neighbour replicas which have a different scaling (i.e. different biased Hamiltonian). The neighbour replicas provide an enhanced variety of conformations (in the above example some partially unfolded motifs), accessible to every replica in the limit of a high number of exchanges between the replicas, i.e. long-time limit.

Recently, the replica exchange in Hamiltonian parameter space described above (referred to as Hamiltonian replica exchange method, HREM, 14) was combined with the umbrella sampling methodology for free energy calculation (15-17). Despite comments on the parameterisation of one of these studies (18), the use of replica-exchange moves during

umbrella sampling systematically leads to dramatic improvement in conformational sampling. Such improvement will be demonstrated on two examples in chapter III and IV, respectively the umbrella sampling study on the metastability of RNA kink-turn motifs and of DNA base pair kink motifs. In chapter V a Hamiltonian replica exchange method based on an additional term in the potential energy function which flattens the potential energy surface associated to DNA backbone dihedral transitions is developed. We found that despite the recent development of a new force-field for more accurate representation of nucleic acid backbone dihedral transitions (19), used here, several ϵ/ζ backbone conformations in damaged DNA (an abasic site) were sampled only with such enhanced sampling on short simulation time scale (5 ns). The advantage of the HREM approach is that an unbiased probability distribution is directly obtained if we append a replica without any additional term. In other words thermodynamics properties such as Free Energy are available from the data output without further statistical treatment. For this reason I devoted the last chapter of my thesis to an extension of the method. The latter aims to make use of the information in the unbiased replica *on-the-fly* and adjust the conformational searching in phase space of the biased copies directly during the simulation. First results on a simple test model DNA system as well as outlooks are discussed in this chapter.

Goal of my PhD

In addition to new methodological development in biomolecular simulations the goal of my PhD is to better understand the flexibility of nucleic acids based on quantitative arguments. It addresses three scientific issues which have arisen notably in the experimental field of investigations: in RNA molecules the possibility of a spontaneous elbow-like dynamics of kink-turn ribosomal motifs, in DNA molecules the possibility of spontaneous kinking events between two pairs of bases and in damaged DNA molecules, an equilibrium between a variety of backbone dihedral conformations. I organized the results of my PhD such as to evidence a complementarity between modelling and experiments in biology: information not derivable from any available experimental technique. This first goal will require in every case some assertion that our methodology is able to reproduce basic experimental observations. Some connections between the different chapters will be made in terms of specific issues concerning the flexibility of nucleic acids. However, methodological developments provide the general guideline to my PhD dissertation.

Several advanced sampling methods are available in molecular dynamics and the best choice depends on the problem at hand. To this aim I introduce an order parameter in each issue for the problem at hand. By combining replica exchange with standard umbrella sampling it was intended to provide a self-consistency check of the sampling obtained. In the course of time allocated for my PhD this goal became more prominent. Combining these two powerful sampling strategies resulted in an ansatz significantly more efficient and by avoiding important artefacts, sometimes leads to modify our conclusions.

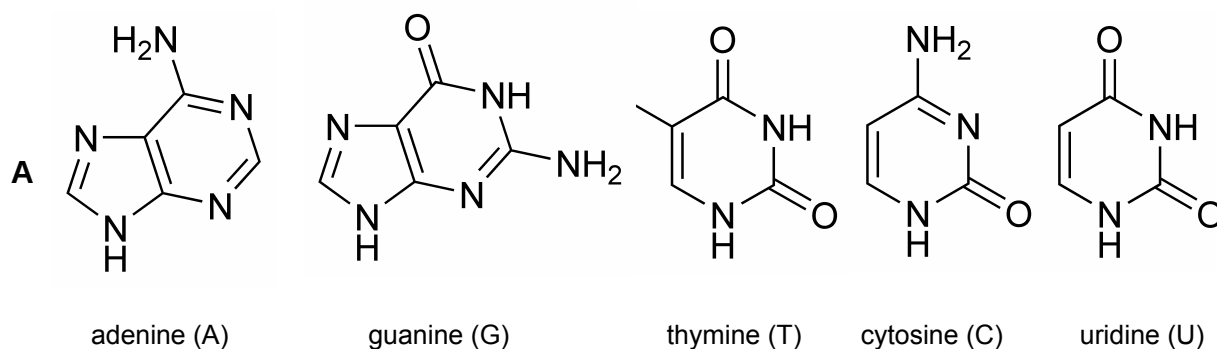
The last part of my PhD thesis goes a step further with the replica exchange sampling approach (in molecular dynamics). I present a new development about a self-regulating searching strategy which aims to take full advantage of the Hamiltonian replica exchange formalism. This final part is just a proof of principle, not an application in biology. However the notion brought about, that of retailers copies (replicas with adaptive sampling), introduces the reader to most recent developments in the replica exchange field as well as to enhanced sampling methods beyond the replica exchange method. It also set marks of a common conclusion which encompasses all methods of current dissertation and the challenge of devising hybrid, multi-scale molecular modelling in contemporary biological sciences.

Chapter I

Structure and Flexibility of Nucleic Acids

1.1 Chemical composition of Nucleic Acids

Nucleic Acids such as DNA and RNA are polymers made up of a linear array of building blocks. Each residue is constructed from three components: a nitrogen heterocyclic base, a pentose D-ribose (ribonucleotide in RNA or 2'-desoxyribonucleotide in DNA) and a phosphate residue PO_4^- (pH 7) that makes phosphodiester bonds between successive sugars. There are most significantly five types of monomeric units to build up molecules of nucleic acid and they differ between their aromatic base constituents (Fig.1.1). The aromatic base can be either monocyclic (*pyrimidines* C, T, U) or bicyclic (*purines* G, A). The sugar is locked into a five member furanose ring twisted out of plane in order to minimize non-bonded interactions between its constituents. This sugar puckering (Fig.1.2) is identified through the relative displacement of carbons 2' and 3' from the median plane of C1'-O4'-C4' in the pseudo-rotation cycle of the furanose ring. It is notably influenced by the orientation of the base and intra-strand hydrogen bond from O2' in one residue to O4' in the next, and as a result the C2'-endo pucker dominates in DNA but the C3'-endo is the most frequent in RNA (and some unusual DNA conformations such as Z-DNA).



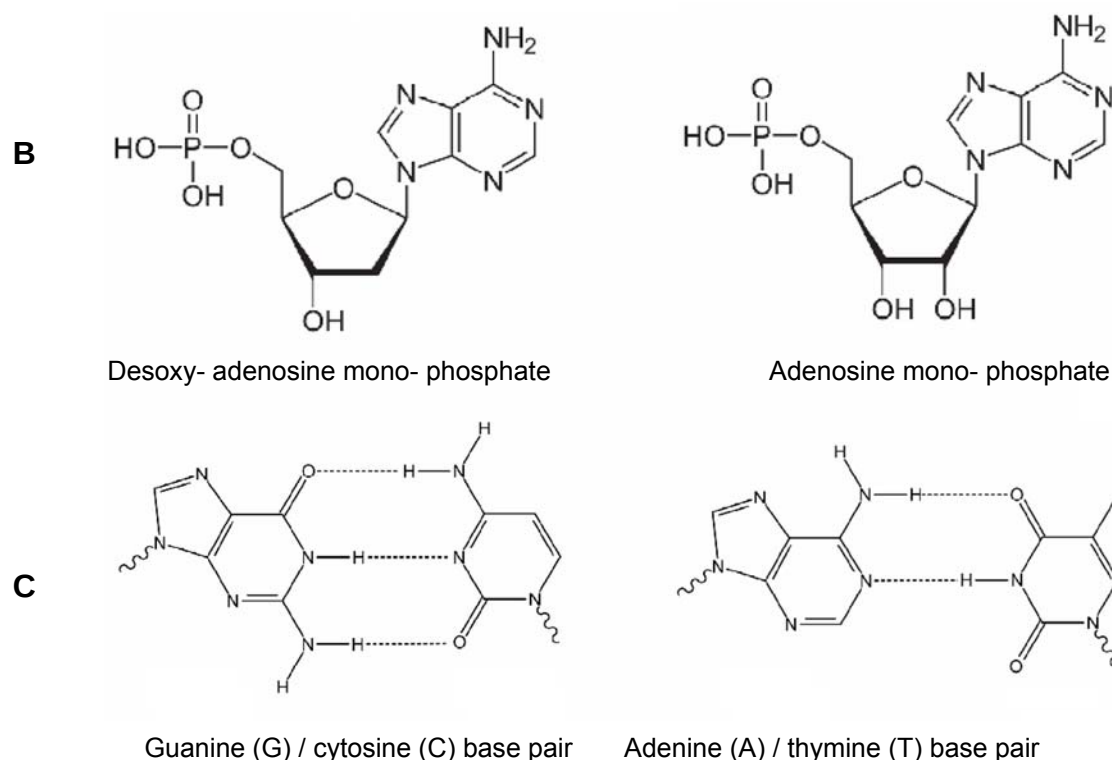


Figure 1.1. Building blocks of nucleic acids: (a) Aromatic bases, (b) Nucleotides (example of an adenine) and (c) Base pairs. The source of these schematics is *Wikipedia*.

The furanose ring is linked by the bond from C1 to N1 of C, T or U and to N9 of A or G. The bi-component sugar-base (Fig.1.1) is called “nucleoside”. Nucleosides are linked together to form a polynucleotidic chain through the phosphodiester backbone which restricts the conformational space of the attached nucleosides.

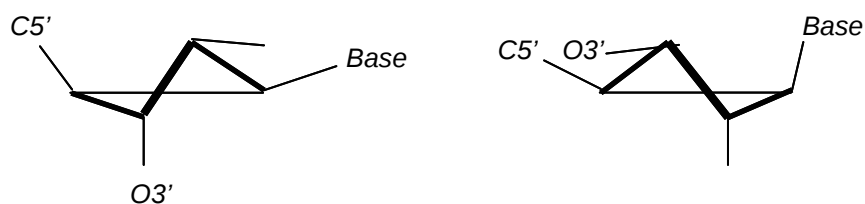


Figure 1.2. The two usual sugar-puckers presents in the B-form (C2'-endo, left) and the A-form (C3'-endo, right) conformation of nucleic acids.

This conformational space is defined by the six torsion angles α , β , γ , δ , ϵ and ζ as well as the glycosidic torsion angle χ joining the furanose ring to the aromatic base (Fig.1.3). A

discussion on nucleic acid backbone flexibility is given in section 1.4.3. Each tri-component nucleoside + phosphate group is then called a nucleotide.

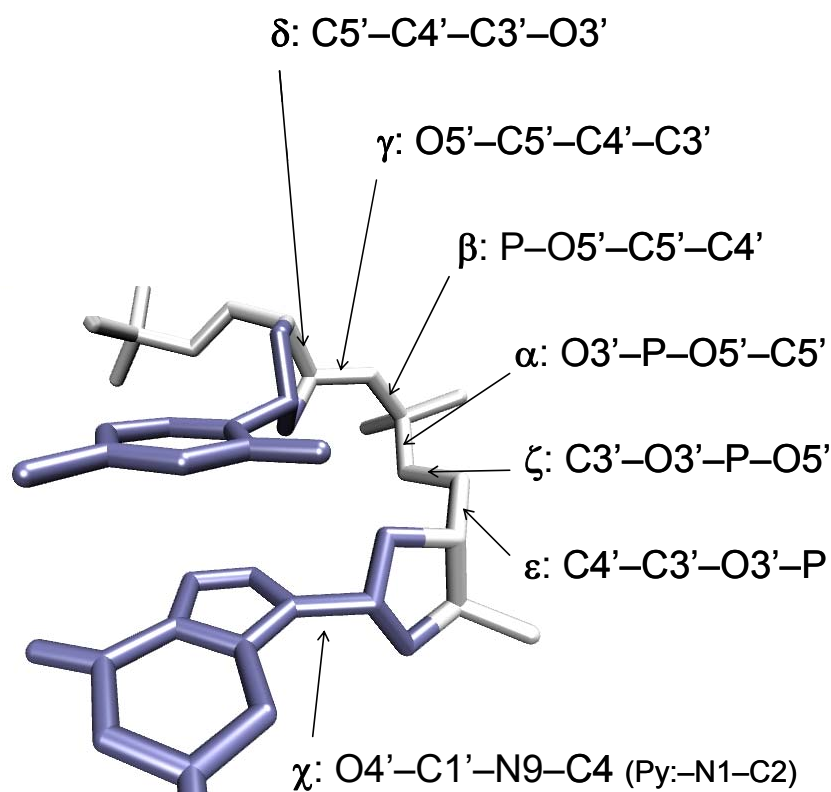


Figure 1.3. Backbone dihedral angles of the phosphodiester backbone and the glycosidic torsion χ .

The primary structure of DNA was established in 1935 by Klein and Thannhauser (20) who found that oligomerisation of nucleotides into a single-stranded (ss) polymer was always through the 5'-hydroxyl group to 3'-hydroxyl group between the nucleoside residues, never through 5'-5' or 3'-3' linkage (Fig.1.4). This implied that the uniqueness of a given DNA primary structure resides solely in the sequence of its bases. Fig.1.4 indicates three of the common shorthand notations used to describe a sequence of ss-DNA. In the following chapters the condensed notation will be used and often without the prefix *d* (for desoxy-) or *p* (for phosphate), will refer to RNA or DNA implicitly because Chapter III deals *only* with RNA and the others *only* with DNA, and will refer explicitly to double stranded sequences. Note functional RNA frequently adopts a single-stranded secondary structure in contrast to functional DNA (but some viral DNAs are single-stranded).

In DNA two single strands wrap and run in opposite direction into a right handed double helical conformation which presents an important structural polymorphism referred to by a prefix letter (Z-DNA is left handed). Far most frequent polymorphs are the A- (RNA) and B-

(DNA) families (Fig.1.5). Table 1.1 compares standard helical parameters for A-, B- and Z-conformational families.

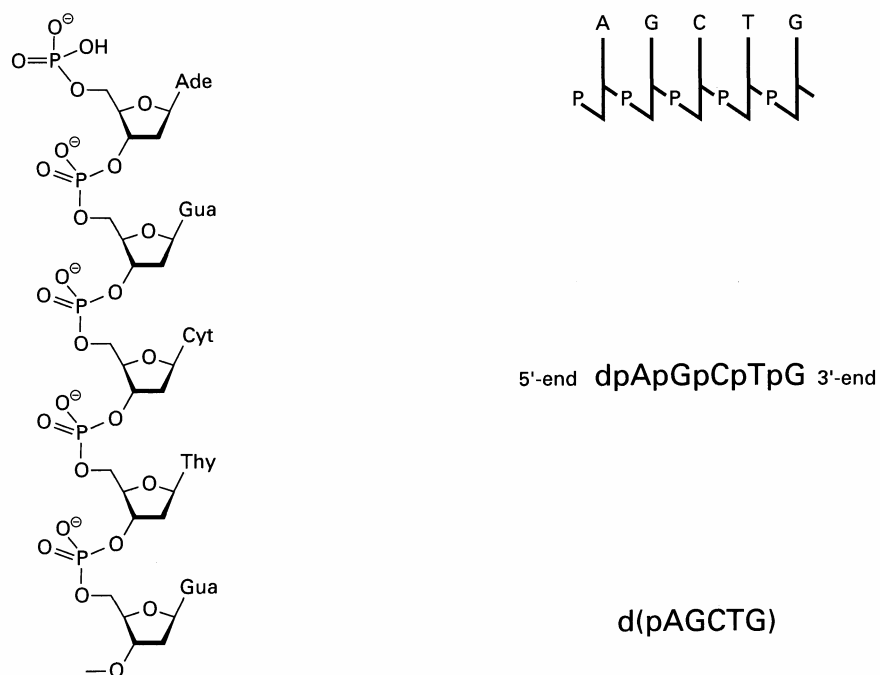


Figure 1.4. Oligomerisation of nucleotide residue and common shortage notations: Fischer (upper right), linear alphabetic (centre right) and condensed alphabetic (lower right). [source: (21)].

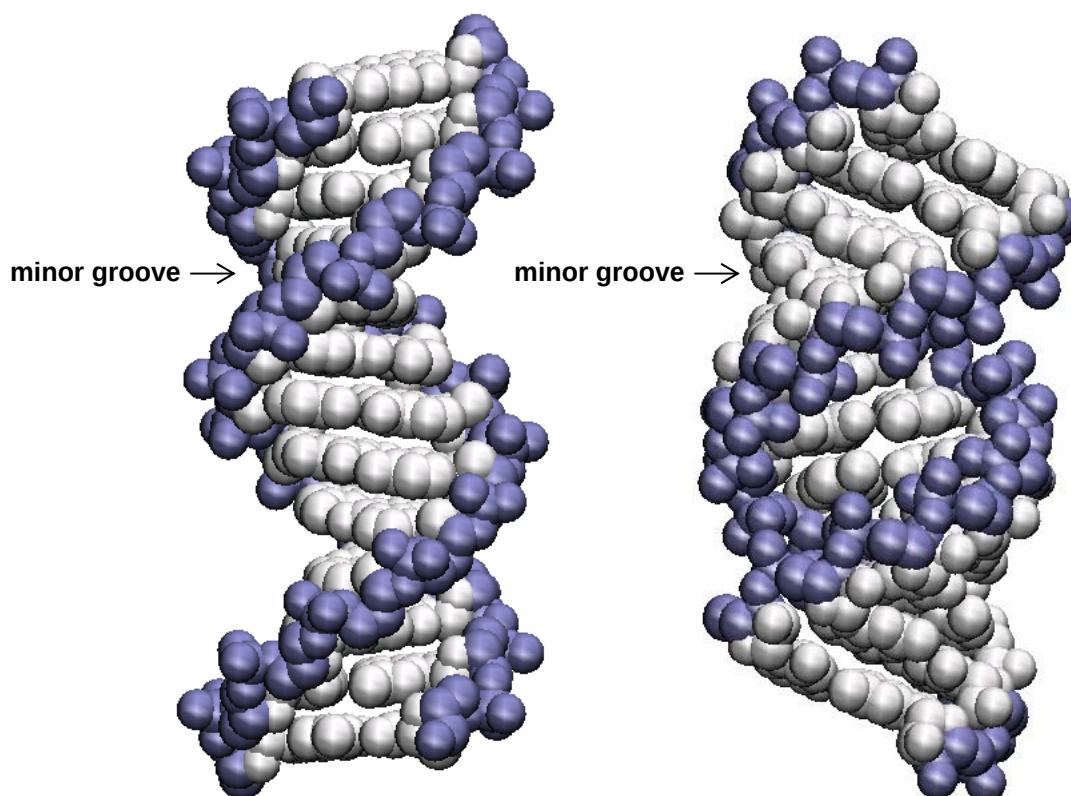


Figure 1.5. Molecular frame picture of a typical B-form dsDNA (left) and A-form dsRNA (right) (generated with Jumna (22), same length, same sequence of bases).

	A	B	Z
Condition	RNA and low humidity DNA	Hydrated DNA	High salt DNA
Helix sense	Right handed	Right-handed	Left handed
Repeating unit	1 bp	1bp	2 bp
Rotation/bp	33.6°	35.9°	60°/2
Mean bp/turn	10.7	10.0	12
Inclination of bp to axis	+19°	-1.2°	-9°
Rise/bp along axis	2.3 Å	3.32 Å	3.8 Å
Pitch/turn of helix	24.6Å	33.2Å	45.6Å
Mean propeller twist	+18°	+16°	0°
Glycosyl angle	anti	anti	C: anti, G: syn
Sugar pucker	C3'-endo	C2'-endo	C: C2'-endo, G: C2'-exo
Diameter	26	20	18
Groove	Base pair displaced from axis, Deep Major	Base pair on axis, Both Grooves available	Base pair stick out into Major Groove, Wide minor

Table 1.1. Geometry of A-, B- and Z polymorphs of nucleic acids

The natural twisting of the double helix can be concisely explained by the strong requirement for stacking interaction between base pairs (details below) and the total length of the backbone-segment joining two successive base pairs since this length is twice as long as the thickness of a base pair (23). The double helix can be characterized *globally* by its handedness (right / left handed) and also the depth of its grooves (see Table 1.1): the B-form DNA helix has one small (narrow) and one wide groove, of similar depth.

Usual interactions stabilizing the secondary structure of nucleic acids are the hydrogen bonds between bases in pairs, stacking interactions between neighbour bases along the helical axis of the double helix, and long-range intra- and inter backbone forces (24). Standard hydrogen

bonding between G and C or between A and T (or U in RNA) is of the type proposed by Crick and Watson (Fig.1.1). Non standard base pairing is treated in sections 1.4.4 and 1.5.1. GC pairs are more stable ($E_{GC} = 16.79$ kcal/mol (in vacuum, 25)) than AT pairs ($E_{AT} = 7.00$ kcal/mol) because they involve three hydrogen bonds and two hydrogen bonds respectively. Stacking interactions hold one base over the next and form a stack of bases which minimizes the exposure of hydrophobic aromatic base rings to the solvent. In addition to hydrophobic forces, stacking interactions are also contributed by dipole-dipole interactions, π -electron systems and London dispersion forces (25). They can easily be detected by marked reduction in the intensity of UV absorption (hypo-chromicity) because the shared π - π electron interactions affect greatly the transition dipoles of the bases. The minimum energy distance that results between the overlapping base pairs, is close to 3.4 Å. Stacking interactions depend on the sequence of base pairs in double stranded nucleic acids and recent estimates from quantum chemical calculations (26) are shown in Table 1.2. If we compare the energies of both hydrogen bonding and stacking interaction with those of covalent bonds (on the order of 50-100 kcal/mol, 25) we conclude that the first two are of the same order of value but both at least an order of magnitude weaker than the covalent structure. Thus a significant degree of conformational change can be expected under environmental stress before the covalent architecture changes. Moreover we can expect

	Assuming propeller twist 0°	With optimized propeller twist	Optimized Propeller twist (°)
AA	-15.01	-16.54	20
AT	-14.78	-15.85	14
TA	-14.67	-14.80	6
GG	-13.57	-13.98	6
GC	-16.12	-16.12	0
CG	-16.41	-16.58	0
GA	-14.29	-14.29	0
AG	-14.65	-15.25	8
TG	-15.81	-16.02	2
GT	-14.98	-15.09	6

Table 1.2. DNA base pair stacking interaction energy (kcal/mol) from recent quantum mechanics calculations (in vacuum, 26).

base pair disruption to be induced by perturbation of the base pair stacking pattern, such as DNA bending. Backbone forces arise from the charged phosphate group (PO_4^-) at

physiological pH. The distance between phosphates on a same strand is relatively very small ($\sim 5\text{\AA}$) and the mutual repulsion is avoided by counter-ions from the medium shielding the phosphates. In a solution of $5 \cdot 10^{-1}$ to $5 \cdot 10^{-4}$ M of NaCl, on average 0.88 Na^+ shield each phosphate group (24). Any factor that alters the counter-ion shielding effect (nature of salt present, pH, large amount of strain affecting local charge distribution) can in turn be responsible for conformational transition. For example it was shown that DNA bending is induced by increased Na^+ or Mg^{2+} salt concentration (27) and that phosphate-phosphate electrostatic repulsion account for at least 30% of DNA bending free energies (28a,b). Also sharp local bends (kinking) were observed in DNA minicircles on AFM media in presence of zinc ions for $> 10^{-6}$ M Zn^{2+} (29).

1.2 Polymorphism

The sequence of bases influences groove width, helical twist, curvature, mechanical rigidity and resistance to bending. The three relatively frequent conformational families of nucleic acids have been mentioned (A-, B- and Z-, Table 1.1). In addition more subtle conformational changes are essential for the biological function of nucleic acids and a continuum of such changes between A- and B polymorphs has been proposed (30). In particular the intrinsic curvature and the bending flexibility of particular sequences play a major role in recognition of a nucleic acid with Proteins and in its cellular compaction (eukaryotic nuclei, prokaryotic cells, viruses) and this is described in section 1.4.1 and 1.4.2. Another important sequence-dependent feature of nucleic acids is the modulation of the contra-rotation of bases inside each base pair, referred to as propeller twist (Fig. 1.6). Propeller twist can improve face-to-face contact between adjacent bases in the same strand and increase stacking stability between base pairs (23). The importance of propeller twist in sequence dependent DNA curvature is notably investigated in Chapter IV. The presence of propeller twist is the main signature of curved DNA fragments known as the Atract sequences (run of consecutive Adenines) which adopt heteronomous conformations in between the A- and B- forms. More generally the polymorphism of DNA was interpreted by Calladine in term of base clashes dependent on the sequence (31). These famous *Calladine's rules* involve four major components of motion to relieve steric clashes in DNA: {1} propeller twist the base pairs, {2} rotate the base pairs to increase steric space across the groove (roll), {3} translate the base pairs to push the bicyclic

purines away from the helical axis (slide) or {4} unwind the helix locally to decrease inter strand purine-purine overlap (un-twist).

1.3 Analysis of Nucleic Acid conformations

To analyze the three dimensional disposition of base pairs and base pair steps, these repeating units are seen as quasi-rigid body systems. Because of chemical variability (A, C, G, T or U) and conformational variability the rigid body superposition requires a proper choice of atoms in each residue (a rigid fraction) to define a coordinate axis system and then fitting the coordinate frame to a standard reference frame which defines the “0 deformation” of the residue unit. The choice of the standard reference frame, for example latest fibre coordinates of B-DNA conformation (32) in (33, 34), some coordinates based on a survey of high resolution crystal structure database in (35) or based on principal axes of inertia for base pairs in (36), as well as the choice of the fitting method, give rise to a level of degeneracy when interpreting a given nucleic acid structure. Moreover methodological “degeneracy” arises also from which definition is adopted for analysing the transformation between base or base pair coordinate systems (37, 38). Thanks to a consortium held at Churchill College, Cambridge in 1988 (39) a narrow range of standard methods are most widely used and one such standard is briefly sketched in the section “transformation operations based on three Euler angles”. The two “alternative models for the transformation operation” develop the limits of the local parametric description (global parameters, model 1) and then *pro vs. cons* of a simpler description (screw theory, model 2) which will be used in Chapter IV.

1.3.1 Transformation operations based on three Euler angles

Because the aromatic bases in nucleic acid are nearly planar and they have a central role in fixing the fine structure of nucleic acids through stacking and pairing interaction, they are most often used to define the quasi rigid body unit in contrast to the backbone. For a chosen base i this is achieved in the standard software Curves (33,34) by defining the axis system $(XYZ)_i$ so that Z_i is the base normal defined by the vector product between the glycosidic bond and the bond N9-C4 for purine or N1-C2 for pyrimidine, X_i is the local dyad direction in the B reference conformation (pointing toward the major groove side) and Y_i is the vector product between X_i and Z_i (pointing toward the sugar-phosphate backbone). In the case of

non-planar base coordinate a standard (planar) base is fitted (40) to the input coordinate before to construct the (XYZ) coordinate frame. Having chosen a fixed axis system which describes the position in space of quasi rigid body nucleotidic bases, the set of schematics reported in Fig.1.6 define a complete set of helical parameters for describing any nucleic acid conformation. Details of mathematical treatments to compute the helical parameters of Fig.1.6, either matrix based method or axis-projection based methods or a combination of both for seven different softwares in current use can be found in (37,38). Note that it is straightforward to derive the parametric description of base pair orientation (Fig.1.6, right) from the treatment of single stranded oligomers and I refer to (33) for an example.

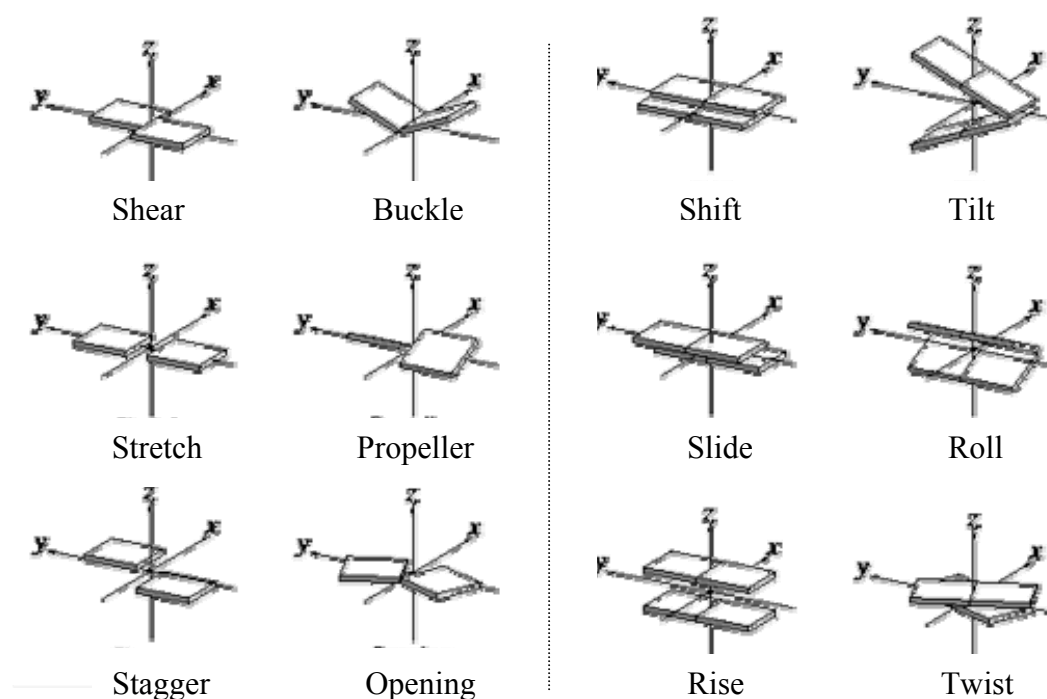


Figure 1.6. Helical parameters describing the arrangement of bases inside a base pair (left) and of base pairs inside a double-helix (right) in nucleic acids (from 39).

The six parameters necessary to describe the relative orientation of two base pair are three translations along the X, Y, Z axis respectively and three rotations (Fig.1.6, right). The rotation about the axis normal to the base pair plane (Z axis) defines the twist rotation, the rotation about the long base pair axis (Y axis) is called the roll and the rotation about the short base pair axis (X axis) is called the tilt. Another rotational parameter of particular importance for the discussions in this thesis is the propeller twist rotation, defined as the rotation of two complementary bases, in one base pair, about the long base pair Y axis (Fig.1.6, left). It is known that the Euler angles description of quasi rigid body transformation is not commutative

and that any sequential combination of the three axes for rotation can be used at the condition that the chosen rotated axis are not redundant (41). The *ZYX* convention for Euler rotation, $(X,Y,Z)\mathbf{i} + \mathbf{1} = \{Z(\text{twist}) \times Y(\text{roll}) \times X(\text{tilt})\} \times (X,Y,Z)\mathbf{i}$, is widely used for nucleic acids. It translates to twist - roll - tilt successive rotations. Finally, a middle reference frame between two adjacent quasi rigid units (bases or base pairs) is used at every 5'-3' or 3'-5' junction to ensure that the magnitude of computed parameters are independent of the direction from which the strands are read, as required by one of the major conventions fixed at the 1988 Cambridge workshop (39).

1.3.2 Alternative models for the transformation operations

Model 1: global description

The Curves software (33,34) is unique in its generation of an optimal helical curvilinear axis threading along the nucleic acid analysed. This global axis gives the viewer an instantaneous picture of the curved helical path (global bending or curvature) for the input molecule.

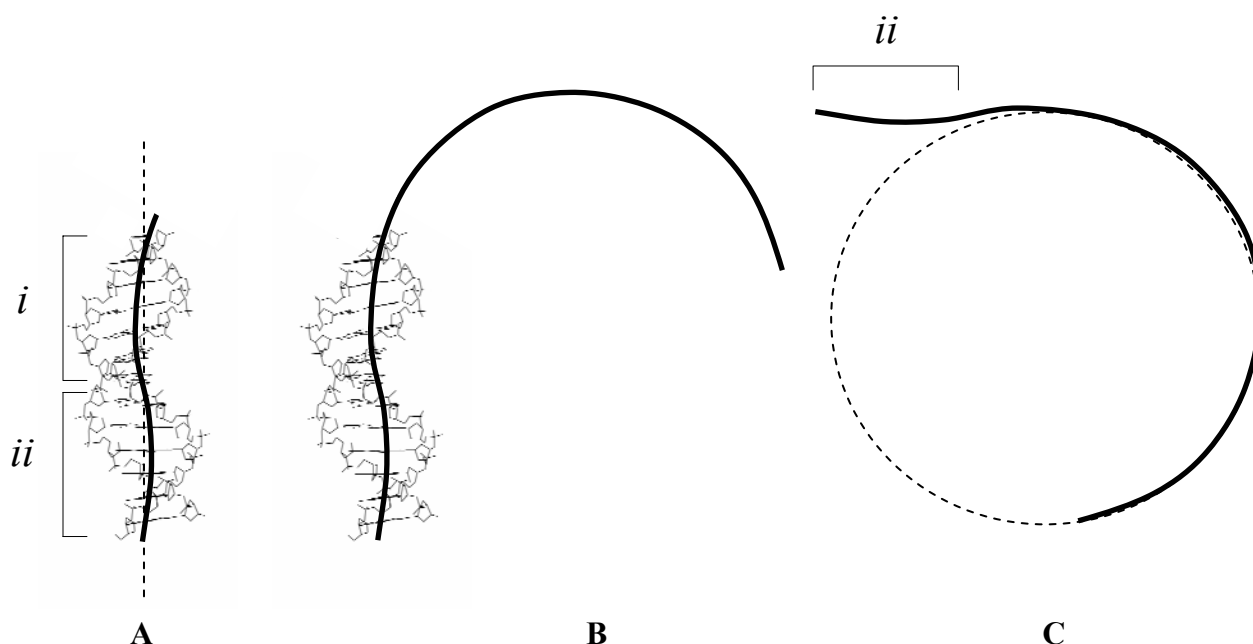


Figure 1.7. Importance of the context in describing local helical parameters. The full line schematizes the helical axis of a nucleic acid obtained by joining independently local screw axis between bases or base pairs. In **A** the helix is nearly straight. An optimal global axis for this molecule would be a straight line and rotation parameters between residues would describe slight bends with symmetrically opposite direction between *i* and *ii*. However (**B**, **C**) for the same segment embedded in a large and

homogeneously curved helical axis (e.g. a minicircle), the rotation parameters can be increased in amplitude in *ii* and decreased in *i* to make clear that the segment “*ii*” is a local heterogeneity in such a minicircle. This can be achieved by optimizing a global helical axis into a near-circular axis (dashes) with respect to which are calculated the base pair parameters (**C** is same as **B** rotated 90° clockwise).

Moreover the helical parameters of Fig.1.6-right, in Curves, are calculated with respect to the local helical axis for each base pair junction. The set of successive helical axis segments produce a global curvilinear axis. Curves also minimises a function which sums-up energy terms for (i) the kink angle (+ lateral dislocation) between successive helical axis and (ii) the orientation (+ displacement) of successive pairs of bases with respect to their local helical axis systems. The balance between the terms (i) and (ii) for every basepair and junction in the fragment is a global curvilinear axis as much linear as possible while keeping the orientation of each base with respect to its local helical axis as much homogeneous as possible.

And thus in addition to the local description Curves also provides a set of global parameters by describing the transformation between successive base pairs (parameters of Fig.1.6, right) with respect to the set of local helical axis obtained as output of the *minimised* global curvilinear axis. These global helical parameters are less often used than local helical parameters, for Curves is almost unique (see also 42) as a complete tool of analysis for nucleic acid structures (global + local parameters). The global parametric description of nucleic acids is nevertheless an optimal helical description for frequent cases such as the overall bent double helix schematized in Fig.1.7.

Model 2: single parametric description

The use of sequential rotations about defined axis (Euler angles) is convenient to assess the specific components of base pair deformability. It is also most typically used in statistical treatment of mechanical properties of nucleic acids in solution (43) as well as coarse-grained models of DNA including the continuous rod theory (44-45). However the theory of screw motion brought about in the nineteen century by Chasles (46) provides a simpler specification of the total transformation between two identical rigid bodies. The original theorem of Chasles is:

« Il est toujours possible de transporter un corps rigide libre ayant une position et une orientation prédéfinie vers une autre position et une autre orientation par le biais d'un mouvement continue et unique le long d'un axe de rotation unique. » (46)

A weaker form of this theorem, often used in the statistical mechanical problems dealing with infinitesimal transformation, is that “any displacement of a rigid body can be accomplished by a single rotation and a single translation” (41). As a consequence the three consecutive rotations by Euler angles twist-roll-tilt can be replaced by a single rotation $\{A\}$ about a space fixed axis (Fig.1.8), i.e. an axis for which the orientation is a required and sufficient condition for resuming the actual rotation between the two base pairs (47). The concept of helical parameters based on a single inter-nucleotide rotation about a space fixed rotation axis in accordance with the guidelines of the 1988 Cambridge workshop (39) has been developed for the analysis of nucleic acid structure (using the three direction cosines of the rotation vector as Twist-Roll-Tilt equivalents, 48). The degeneracy with the standard model (Euler angles) was found to be small enough as to lead to energetically equivalent helical conformations (47). The great advantage of this reduced representation was illustrated in the work of Srinivasan et al. (49) where an *axoid* defined by the set of every successive space fixed rotation axis was taken as the global curvilinear axis threading along the nucleic acid analysed. Together with the set of unique rotation angle plotted as a function of residue number, the axoid description acts as a finger print of any given nucleic acid conformation (49). This single parametric approach will be used in chapter IV not for analysis purposes but for the generation of DNA conformations with a given global bending geometry and the reason is as follows. The main disadvantage of this rotation model is that detailed information on the rotational component such as roll-tilt-twist is lost and is recoverable via reasonable matrix operations only with a certain degree of degeneracy (47). However this does not mean that any of the standard rotational parameters (twist, roll and tilt) is differently taken into account (take the analogy between the amplitude of a vector in 3 dimensions and its three exact coordinates in a new coordinate axis system if a reference system was actually never given). A considerable advantage of this model is its high simplicity and the fact that the orientation of the rotation vector, i.e. deviation from the base normal, is due to non-zero x- and y- direction cosines and this equals the amplitude of bending between the two base pair coordinate frames (47, 48). In

consequence the parametric angle describing the average relative orientation of two axoids (*axoid* as defined in (49) in the context of structure analysis) will have the following features:

- It does not make any distinction between *roll*, *tilt* and *twist* but includes all of them.
- It corresponds to the average amplitude of bending *between* and *inside* the axoids.
- It is a relatively simple operation (i.e. computationally efficient).

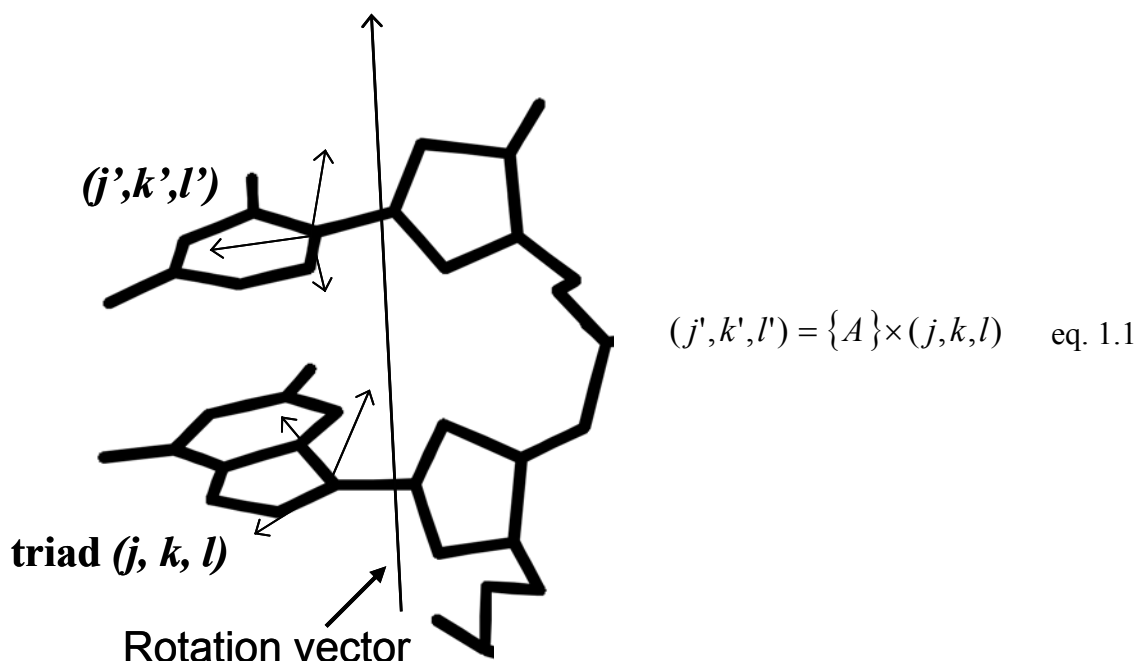


Figure 1.8. The space-fixed axis of a total rotation $\{A\}$ between two bases' orthogonal frames.

1.4 On the conformational flexibility of DNA

1.4.1 Protein / DNA interactions

Cellular compaction of DNA

The global deformability of DNA has obvious role in protein / DNA interaction when considering the length scale of DNA *in vivo*. A typical human chromosome contains 50 to 400 million base pairs and 46 chromosomes (total length > 1 meter) are confined in a nucleus of diameter $\sim 10^{-6}$ meter (21). The first stage in the hierarchical organisation leading to the required compaction of the genetic material in the nucleus of eucaryotes, and which is the aspect of concern here, is the successive wrapping of consecutive portions of ~ 145 DNA base pairs in $1\frac{3}{4}$ turn around independent spools of protein (nucleosomes), a sub-structure referred

to as “bead on a string” (50,51). Each nucleosome is made-up of an octamer of positively charged histones which help in stabilizing the 360° bending turn-around of ~80 negatively charged DNA base pairs as required for such organisation (52).

Experimental (53-57) and theoretical (58-61) studies have demonstrated that nucleosomes have preferences for certain nucleotide steps which form a periodic pattern of hinge-bending fractures, alternating with the helix phase. This nucleosome positioning frame directly encoded in the genome was derived from several species of eukaryotes (including chicken, yeast, mouse, human, 54,56,57) and can be used to predict *in vivo* nucleosome organisation as in the recent thermodynamics model of Widom (57). Long range correlations (150 bp and beyond) in the DNA bending profile were also found by wavelet transform analysis as a signature of nucleosomal organisation in eukaryotes (62, 63). Aim of these models is to integrate the chromatin structure into models of gene regulation (57). An aspect of great importance is the extent to which on-site spontaneous accessibility of nucleosomal DNA is possible *in vivo* (models / studies in 64) as notably suggested by FRET measurement of the rate of spontaneous DNA unwrapping from the nucleosome core (65,66). Based on a canonical persistence length of DNA (150 bp, 67) the role played by DNA flexibility would be a passive role and the answer to above question would be deceptive. However it is known that sharp local bending at DNA base pair junctions occurs in the nucleosome (“extreme kinking” in 55) and facilitate access to DNA at least by molecular agents probing DNA base pair bends such as singlet oxygen and permanganate (68). Most importantly this PhD thesis comes at the time of controversies about the true value of the DNA persistence length which started with the 2005 cyclisation kinetics experiment of Johnathan Widom (12). The Jacobson-Stockmayer (J) factor (ratio of cyclised versus linear polymers in solution) that they observed on nucleosomal sequences was order of magnitude ($\times 10^4$) higher than predicted by the linearly elastic theory based on a persistence length of 150 bp. Spontaneous sharp bending of DNA on short length scale was already observed in AFM experiment at unusual salt concentration (29) and predicted some years ago by computer simulation (45° (69), 22.5° (70)), but the J-factor observed in (12) occurred at physiological condition and also on random DNA sequences. Very recently the investigation was extended by AFM observations (4) of random DNA sequences over 750000 base pairs by the same authors and by various FRET measurements of others (5,6). They all confirm a softer DNA response in term of strong bending on short length scale. In other words the origin and stereochemistry of some given DNA kink motif(s) in free DNA in solution have not been explained, but a high prevalence of such DNA virtue is now an experimental proof. Along this thesis will come

some new insights onto the intrinsic bending deformability of DNA on short length scale, and indeed the outcome of above experiments is the development of a new series of both mesoscopic model for long pieces of DNA as it already started to appear (71-73) and of elastic potential (74) which estimates the indirect readout component in protein/DNA recognition (see below). In prokaryotic cells as well as viruses which do not have a nucleosomal architecture, DNA is continuously associated to abundant proteins along its thread, e.g. Fis, HU, IHF, that stabilize compaction (into *nucleoids*) and regulate DNA accessibility for replication, expression and repair (75). The structure of DNA in the X-ray crystal of these protein/DNA complexes is again most often kinked at some specific base pair. More generally sharp DNA bending/kinking is an essential property of many protein/DNA complexes. This can be appreciated in the Table 1.3 compiled by K. Zakrzewska in (76).

PDB-	Bend (°)	2CGP	59.4	1ZME	55.9	1AZP	35.5
1LMB	36.0	1DP7	29.7	2DGC	18.4	1BF4	28.6
1PER	27.2	1AWC	4.6	1AM91	20.3	1BNZ	30.9
1RPE	28.2	1BC7	7.8	1AM92	22.3	1CA5	35.7
2OR1	34.3	1BC8	8.1	2BOP	34.4	1NFK	7.9
3CRO	31.0	1BLO	39.0	1B3T	29.1	1B94	55.8
1AKH	52.5	1DUX	10.9	1SKN	36.1	1CKQ	13.8
1MNM	66.8	1PUE	17.5	4CRX	75.7	1ERI	3.6
1YRN	48.7	1AAY	4.3	1CKT	61.7	1BHM	77.1
6PAX	26.4	1MEY	13.2	1EGW1	15.5	1D2I	15.7
9ANT	16.3	1UBD	4.9	1EGW2	12.8	1DFM	15.9
1HCR	31.8	1A6Y	4.7	1AIS	50.0	1EBM	44.5
1PDN	7.8	1BY4	9.7	1CDW	94.7	1DNK	16.3
1TC3	83.0	1DSZ	7.2	1D3U	77.2	2BDP	17.5
2IRF1	20.7	1HCQ	8.7	1YTF	43.7	3KTQ	13.7
2IRF2	17.8	1TSR	20.9	1ECR	22.1	1CEZ	48.8
2IRF3	21.0	1HAP	43.8	1IHF	96.7	1QAJ	4.9
1BER	100.	1HWT	11.4	1XBR	14.7	1A73	53.8

Table 1.III. Overall DNA bend angle in 71 x-ray crystal protein / DNA complexes contained in the NDB and PDB databases (76).

Direct / Indirect readout

The digital (i.e. *direct*) recognition between protein and DNA which involves direct contact between amino acid side chain and specific DNA bases (Table 1.4) and/or non-specific DNA backbone, can be decomposed into a chemical (hydrogen bonding) and a shape (steric fit) recognition (77). The analog (i.e. *indirect*) recognition involves deformation in the DNA sequence not in contact with the protein but nonetheless mandatory for the direct readout to occur properly. The DNA deformation energy used to describe generic mechanisms involved in protein-DNA recognition has many different origins (78).

	Major Groove	Minor Groove
A	C6 N6 C5 N7 C5	C2 N3 C4 N9
G	C6 O6 C5 N7 C8	C2 N2 N3 C4 N9
C	C6 C5 C4 N4	O2 N1 C2
T	C6 C5 C4 O4 C5M	O2 C2 N1 C6

Table 1.IV. The direct read-out in protein/DNA recognition: Coding through diverse set of functional groups that can be used to form specific contacts with DNA binding proteins. Shown are the specific atoms which are accessible on specific side of the widened grooves.

Opening the major or minor groove is often necessary to “protrude” hydrogen bond donor or acceptor and accommodate (sterically) a ligand or part of protein side-chain. In that case the energetic signature appears as both a consequence of major vs. minor groove hydration property (79) and local bending property of the DNA segment (80). Indeed when the helix bends toward the major groove this widens the minor groove (80) and reciprocally (note there is no example of known protein/DNA complex where the helix bends away from the protein toward its minor groove, 81).

More generally the sequence dependent bending deformability of DNA affects the accessibility and position of hydrogen bonding groups in the grooves. A familiar and illustrative example is the helix-loop-helix (HLH) DNA binding motif (82) in dimers where two alpha-helices (“reading heads”) must each reach into one major groove but are separated by a distance shorter than the DNA helical repeat in the straight conformation, rendering specificity dependent upon the intermediary DNA segment (83). This mode of analog

recognition is by no mean universal. The reading head can be a non canonical alpha helix (84), a beta sheet (85), a zinc finger (86) or some more exotic protruding residues at the surface of the protein (87).

In 1988 Matthews summarized the importance of analog readout in protein / DNA recognition by the maxim “No code for recognition” (88). In 1998 Dickerson tried to established a set of principles that would summarize DNA bending as a factor in protein / DNA recognition based on every experimental data then available (89). Today, ten years further, the canon of Dickerson is still supported by the many different known recognition motifs and is thus summarized below before to close this section.

1. Bending in B-DNA is almost always the result of roll between adjacent base pairs, never tilt.
2. The dynamic bendability of B-DNA is anisotropic, with bends toward the major groove (positive roll) being easier than those into the minor groove (negative roll).
3. Bends in DNA are of three types: localized kinks (large roll at one or two discrete base steps), three-dimensional writhe (positive roll at a series of successive steps) or smooth curvature (alternation of positive and negative roll every half-turn).
4. The most significant aspect of roll bends is not that they are necessarily present in unstressed DNA, but that they can be *induced under** stress by environmental factor.
5. Significant roll deformation under stress are a property of relatively weak Y-R steps, whereas R-R and R-Y are straighter on average and more resistant to bending.
6. The overall bending effect of local writhe depends on its phasing with the natural rotational periodicity of DNA (~10 bp). It is maximal when repeated in phase.
7. Because of statement 2, a low level of continuous writhe is the normal condition of an otherwise unbent helix. For only a few specific cases will satisfy 2 in a planar helix ‡.

1.4.2 DNA bending and curvature

The bending deformability of DNA is involved in many biological functions and the indirect readout component of protein/DNA interaction encompasses most of them. Indeed the propensity of nucleic acids to adopt the right conformation for a given function like nuclear/cellular compaction is a special case of indirect readout (77). Other such special cases include the ejection of the genetic material from viral capsides (64) or the formation of regulatory loop in multi-component complexes that initiate transcription or replication (64).

* Italic font in original

‡ Reformulation from original (89) statement.

DNA looping can reduce statistical noise by bringing distal DNA sites close to each other and increase locally the concentration of regulatory factors at the transcription start site (90).

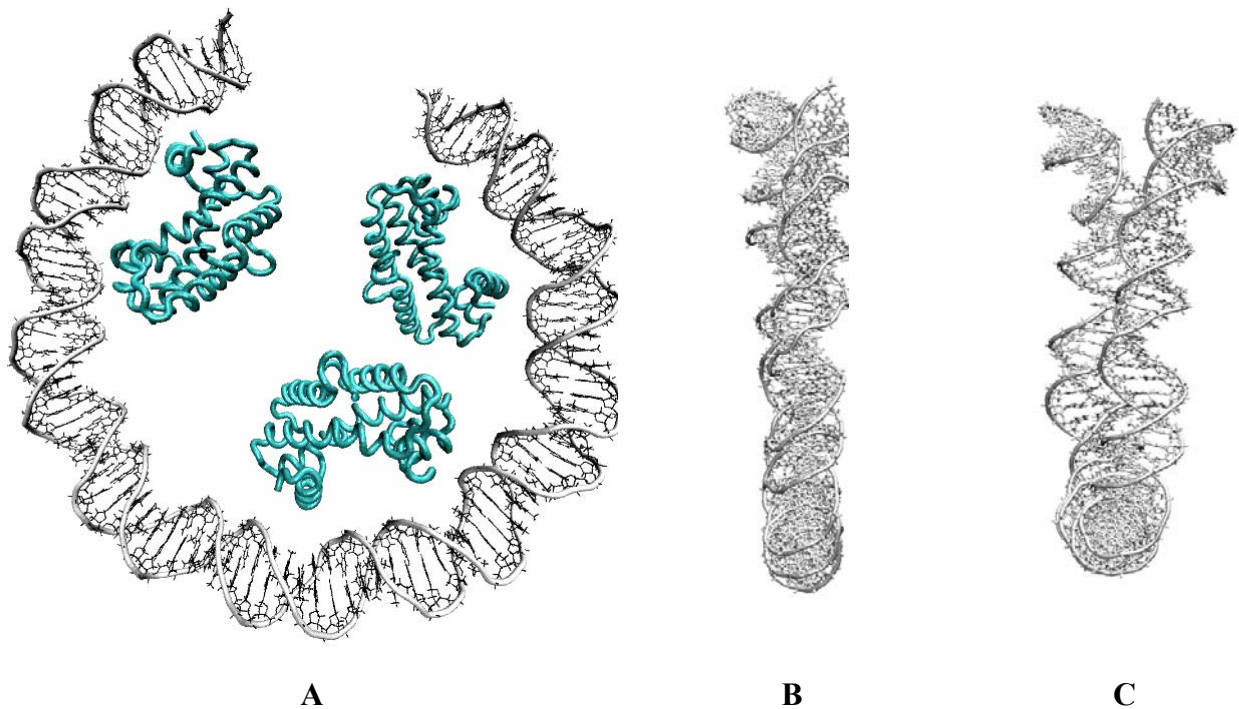


Figure 1.9. Computer generated model of the three Fis/DNA binding sites at the upstream region of the tyrT Promoter in *E. coli* (A) with three Fis dimers. (B): side view of the loop as initially generated with a planar superhelix; Fis dimers are not shown. (C): Same but after 5 ns of molecular dynamics simulation including Fis bind to their receptor sites, which induced a global left-handed writhe.

For example the binding of three Fis protein-dimers with HLH motif (91) to three upstream regions of the tyrT Promoter in *E. coli* (illustrated in Fig.1.9, personal work) forms a transcription-regulatory loop (92). AFM and UV photo- footprinting experiments indicate that the microloop (Fig.1.9) could modulate the kinetic of transcription activation by torsional transmission (93). Importantly all these examples imply mechanical properties on length scale smaller or comparable to the $\sim 150\text{bp}$ persistence length in DNA. When considering the short length scale bending of DNA molecules, a distinction is made (see 43) between a static bend angle characterising a specific DNA sequence, referred to as intrinsic curvature, and the quantitative propensity with which a bend angle can be dynamically accessible upon external stress such as protein binding, referred to as induced bending. Now I discuss briefly these two properties of DNA.

DNA curvature

Certain nucleotide sequences lead to intrinsic curvature and this property can be an essential component in the recognition process (indirect readout). The most widely studied examples of intrinsically curved DNA sequence are the so-called A-tract elements. Indeed a tract of 3 to 6 consecutive adenine (A) nucleotides repeated in phase with the helical repeat of DNA (A-tracts) lead to global curvature, which can cause an unusual slow migration during electrophoresis in an acrylamide gel (94-97). The reduction of electrophoretic mobility compared to a straight DNA of the same length has been used to quantify the sequence-dependence of curvature and for a single A-tract bend angles of 17°-21° have been measured (94). Both A-tract length and the nature of the spacer sequence between A-tracts can influence the bending magnitude (96). Some years ago, Hagerman (97) found that a sequence of the form (dA₄T₄CG)_n has a significant electrophoretic abnormality indicating strong curvature whereas a sequence with the same nucleotide content (dT₄A₄CG)_n shows normal electrophoretic behavior. Note a caution has been emphasized in (98) regarding detailed measures by electrophoretic mobility since the contribution of out of plane curvature to the extent of migration can bias the estimation.

Several models have been developed to explain the molecular origin of intrinsic A-tract induced DNA curvature. Early models like the junction model (99) view the sequence dependent curvature as a result of kinks created at junctions between two types of B-DNA structures [in this model A-tracts adopt a heteronomous B'-structure that differs from the normal B-DNA, see (100)]. Supports for this model come in part from pre-melting transitions of DNA that can be observed in A-tract containing DNAs but not in other sequences (101) and from early modeling studies (102). The wedge model explains the curvature as the effect of dinucleotide specific roll and tilt angles that if appropriately spaced along the helix can give rise to overall DNA curvature (103,104). Structural analysis of A-tract containing DNA using X-ray crystallography indicated a straight A-tract DNA and located the origin of curvature within the intervening sequences (105). Other high-resolution techniques such as NMR spectroscopy have been applied but the relatively low proton density of nucleotides limit the number of nuclear overhauser effects for nucleic acid fine structure determination. These limits of NMR have been partially overcome by the application of residual dipolar couplings and lead recently to the proposal of a more unified model for A-tract induced curvature where bending results from phased combinations of roll and tilt contributions with relatively low amplitude, delocalized over the whole sequence (Delocalized Bend Model)

(106). The A-tract induced DNA curvature problem is developed at length in chapter IV. Similar to A-tracts some other sequences can also adopt intrinsically curved structures or may also exhibit a higher *bending flexibility* of DNA (12, 107-109).

DNA bending flexibility

Beside the static curvature discussed above, any DNA base pair *junction* (the step between two consecutive base pairs) can undergo a certain degree of hinge deformation, i.e. induced bending. The linear elastic theory of polymer physics (110), can be used to describe the deformation up to a breakdown above which non linear contribution have to be taken into account into the strain/stress relation (111). If the strain induced above this breakdown is reversible, we refer to as the non linear elastic theory while if it is non reversible we refer to as the plastic theory (111). A suitable sequence-dependent energy function describing the linearly elastic deformability of the DNA double helix can be extracted from the fluctuation and correlation of structural parameters (in that case an inter-basepair bend parameter) in any thermodynamic ensemble:

$$k_{i,j} = 1/2 \times \frac{k_B T}{\langle (x_i - x_i^0) \rangle \times \langle (x_j - x_j^0) \rangle} \quad \text{eq. 1.2}$$

where $k_{i,j}$ is the elastic force constant corresponding to the coupling of two generalized coordinate x_i and x_j , x_i^0 and x_j^0 are the reference value of highest probability around which are observed the fluctuation, k_B is Boltzmann constant and T the temperature in Kelvin. If $i = j$ then $k_{i,j}$ is the effective elastic force constant of the degree of freedom x_i (without coupling).

This expression is based on the relation first derived by A. Einstein between the probability distribution of a fluctuating variable and the system-entropy (in 112), extended to the analysis of all atoms dynamic trajectories by M. Karplus (113) and to the analysis of helical parameters in DNA by W. Olson (114). In the latter study the formalism was applied to a set of DNA/protein complexes with fictive temperature obtained by scaling the data against solution measurement of the standard DNA persistence length. Assuming harmonic behaviour of base pair step parameter fluctuations (linear elasticity) the authors culled steps with obvious structural irregularities for restricting attention to input parameters that clustered around quasi-normal distributions (114). In so doing they could extract general trends of dimer deformability in DNA noteworthy the importance of pyrimidine-purine step in

exhibiting real time flexibility within protein/DNA complexes. The question of high DNA bending on short length scale could however not be abarded because of very limited data on the nonharmonic, i.e. bimodal, bent base pairs which are in some cases highly distorted (the culled data of 114).

Here it should be clear to the reader from above sections, that DNA molecules do undergo relatively severe functional bending on length scale much shorter than its persistence length (~ 50 nm). As discussed in 1.4.1₁ current efforts for understanding the short length scale complexity of DNA molecules have shed light on a high curvature softening in cyclisation assays (12), molecular force sensor (5), FRET (6) and AFM (4) experiments. The theoretical models which attempt to reproduce this property (71-73) have found a good experimental fit with the general sub-elastic chain (SEC) model (5,72). In SEC the non-linearly elastic bending is represented by a plateau in the restoring bending force, i.e. an effectively linear bending energy function for large bend angle on short length scale (5nm $\sim$ 15bp). This function takes over the initially harmonic energy function for small bend angles. The precise energy function upon which stability of sharply bent DNA depends may deviate from the linearly elastic theory, however molecular mechanisms accounting for non linear DNA bending are not known (4). Chapter IV will present some principal results from simulations in molecular dynamics which are an attempt to answer this question.

1.4.3 The phosphodiester backbone

Nucleic acid backbone elastic properties are important for protein/DNA recognition as demonstrated by the various DNA and RNA backbone substates found in recent analysis of X-ray crystal structures of protein / DNA complexes (115). Its fixed length imply specific coupling of helical parameters in response to changes in the stacking geometry and structurally correlated backbone torsions can be used to define several alternative nucleic acid backbone substates (115,116). In particular (ϵ/ζ) in BI (t/g-) and BII (g/t) conformations are frequent in crystal structures of free DNA (115). Analysis of X-ray crystal structures in conjunction with potential energy calculations have shown that the major DNA base stacking mode slide / roll / twist is coupled to the major backbone mode χ -P- δ - ζ and the secondary DNA base stacking mode shift / tilt to the ϵ - ζ crankshaft backbone mode (117). A recent survey of available nucleic acid structures solved by X-ray crystallography and NMR found that mainly GpC and YpR junctions are accessible to the BII state (118). A correlation between the BII state and a negative roll of the adjacent base pair junction was observed

whatever the junction considered (118). This effect is more pronounced when a BII conformation is adopted simultaneously on both strands of the considered base pair junction (119,120). In contrast, free energy calculations found that spontaneous flipping of the other correlated backbone dihedrals alpha and gamma (α/γ flips) is improbable in free B-DNA (116). It was thus a surprise when some α/γ flips were massively found in several MD investigations of free B-DNA systems on time-scale $\gg 10$ ns (121-123).

These results, which have arisen from new opportunities to manage MD simulations up to the microsecond time-scale (124), explain why the series of Amber force field has been re-parameterized for nucleic acid systems with regard to the α/γ torsional terms (parmbsc0, 19). Note that the backbone dihedral angle γ (C4'-C5') in Trans conformation was predicted in (125) and observed in MD (108,126) to be associated with a sharp deflection (90°) of two consecutive base pair in DNA. NMR studies (127) also suggested a role in the A-tract curvature phenomena of the gamma = Trans conformation at the 5' junction of the A-tract block (CpA in 127). The impact of the new parmbsc0 Amber force field parameterisation for nucleic acids backbone dihedrals on base pair kink motifs associated to gamma = Trans in DNA will be demonstrated in chapter IV. Chapter V and VI are devoted to methodological development for improving sampling of epsilon / zeta backbone dihedral angles during molecular dynamics simulations (and in principle any other backbone dihedrals).

1.4.4 DNA mismatches and abasic sites

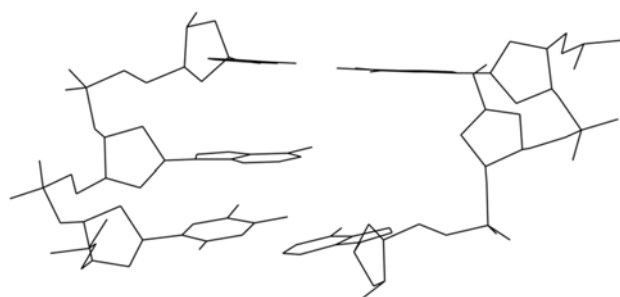


Figure 1.10. Computer generated graphics of an abasic site (apyrimidinic) in sandwich between two common base pairs.

The fidelity of transmission of the genetic code rests on the specific pairings of A with T and G with C bases. Common examples of lesion/mutation in DNA result in mismatches (non standard base pairing) and the formation of abasic sites (unpaired bases) which must be recognised and be repaired by enzymes with high efficiency. The resulting changes in shape at the lesion suggest that they can be recognized specifically by repair enzymes (21). Most damaged DNA can be recognized at checkpoints (128) controlling the passage from the G₁ to the G₂ cell cycle phase to avoid the cell being transformed into a tumour cell. Abasic sites are

of particular importance since, *in vivo*, the first step in the repair of most DNA damages (e.g. oxidative) imply the cleavage of the glycosil bond at the damaged base leading to an apurinic (unpaired pyrimidine) or apyrimidinic (unpaired purine) site (129). Theoretical studies (130,131) indicated that the main perturbation is an enhanced bending flexibility at the position of abasic residues (up to 60° in case of apurinic sites, 131). The amplitude of the bend and the structure taking place at the cavity created by the abasic lesion depend strongly on the sequence. In case of apyrimidinic sites a motion of neighbouring bases into the cavity was suggested (131). An example of DNA apyrimidinic residus as studied in chapter IV and V is illustrated in Fig.1.10.

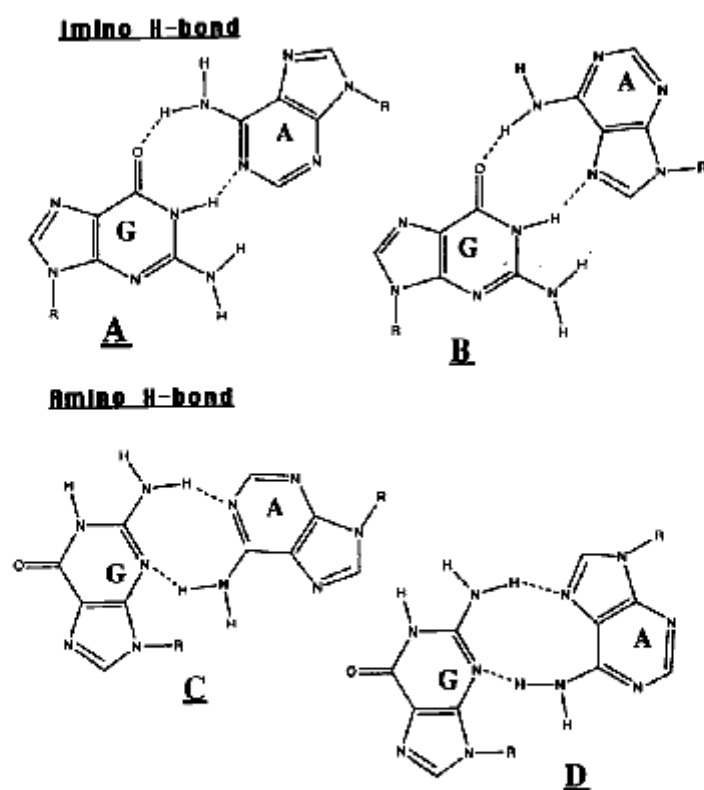


Figure 1.11

Four possible arrangements of G and A in a mismatched base pairing, from (132).

Mismatched base pairs in DNA, are referred to as transition mismatches (pairing of a purine with the wrong pyrimidine) and transversion mismatches (pairing of two purines or two pyrimidines). An important characteristic of mismatched bases in DNA and other nucleic acids is that they may introduce relatively little distortion in the duplex (B-form for DNA) and leave functional group such as imino- or keto groups accessible for tertiary interaction or recognition by repair enzymes (132). Several pairing schemes can be envisioned as illustrated in Fig. 1.11 for the GA transversion mismatch.

1.5 On the conformational flexibility of RNA

Today it is recognized that RNA has greater structural versatility than DNA in the variety of its species, in its diversity of conformations, and in its chemical reactivity (21). In particular tRNA and rRNA have provided a rich source of unusual base pairs and base triplet. Since my PhD has a major focus on the aspect of conformational flexibility in DNA, the following sections present simply an updated overview of RNA versatility.

1.5.1 Non Watson-Crick base pairing

Fig. 1.12 illustrates important examples of unusual (non-Watson-Crick) base pairing observed in RNA molecules. Detailed discussions can be found in several textbooks (21). For the purpose of chapter III, I describe in more details the A-minor interaction which stabilizes the folding of RNA kink-turn motifs (see below) or, as another example, the P4-P6 domain of the self-splicing group I intron which is the first ribozyme* observed, in *Tetrahymena* (81), where an adenine residue belonging to the P6 domain forms hydrogen bonds with the minor groove side of C.G pairs of the P4 domains.

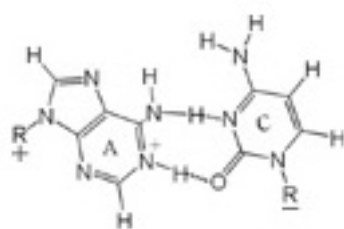
The A-minor motifs (81,133) are represented by single-nucleotide insertions into minor grooves of A-type helices and are more global examples of the ribose-to-ribose structural motif (134). An A-minor motif has been early observed in a few ribozymes structures (81). It involves the smooth N1-C2-N3 edge of an adenine aromatic base inserted into the minor groove of neighbouring helices and has been claimed to be more abundant than Watson-Crick base pairing in RNA tertiary structures (135). The adenines involved in such interactions are often highly conserved and presumably essential for preserving a RNA molecule's functional architecture. A-minor motifs stabilize contacts between RNA helices or between loops and helices. Two types of A-minor motifs have been found to mediate crucial interactions in kink-turn motifs (136). These two types differ in the position of the O2' and N3 atoms of the A-residue relative to O2' atoms in the base pair of the receptor helix (Fig.1.12). The first is the type I A-minor interaction which involves direct trans sugar-edge/sugar-edge (SE/SE) and cis SE/SE base-base interactions and is defined by both the O2' and N3 of the adenine residue inside the minor groove of the receptor helix (observed in kink-turns kt42, kt77 and kt94/99), Fig.1.12. A variant referred to as *water inserted* type I A-minor motif was also found to be important in stabilizing semi-closed/semi-open geometries of some Kturns (kt7, kt23 and

* Ribozymes are RNA molecules that have the potentiality to catalyse specific chemical reactions like Enzymes and thus make the idea of a RNA primitive world conceivable (81).

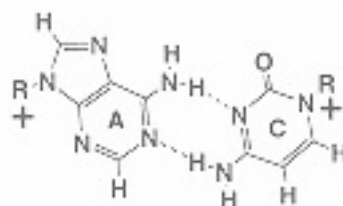
kt38) and involves insertion of a water molecule to mediate the cis SE/SE hydrogen bonds (136). The second type of A-minor motif found in kink-turns (kt15, kt46, kt58, ktU4), referred to as type 0, involves a lateral shift of the adenine residue which brings its N3 atom outside the O2' of the far strand in the receptor helix. Nissen and collaborators (133) have set up a classification of all A-minor motifs identified to date and this classification is reported in Fig.1.12. Note a recent discussion (137) on the difficulty to classify RNA motifs because of intra-variability (small conformational variation) inside each group corresponding to a given version of an RNA motif. This degeneracy leads to a notion of sub-groups (or conformational variants, 137) satisfying a set of definitions for a given version of a RNA motif. It particularly applies here in the context of the A-minor type I interaction, since, notably, we have confirmed (chapter III) that type I divides into three variants: type I “directly hydrogen bonded”, type I “single water molecule inserted” and type I “double water molecule inserted”.

1.5.2. Overview of RNA motifs

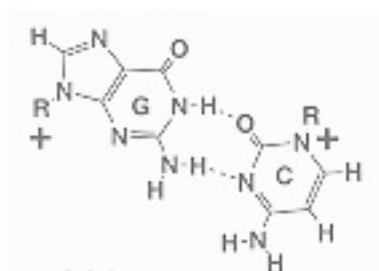
Secondary structures of RNAs can be determined by a combination of techniques including the use of chemical and enzymatic reagents selectives for bases in either single stranded environments or in base pairs and the use of phylogenetic comparisons (21). NMR spectroscopy and X-ray crystallography can help to define the tertiary structure of small RNAs. RNA structural motifs and sub-motifs (137) are classified in the database called SCOR (Structural Classification Of RNA, 138). *Hairpin loops* are widely observed in tRNA and rRNA and appear as single stranded loop (often tetra-loops or tri-loops) at the apex of a double stranded RNA segment. It often forms tertiary interaction between the looped nucleotides and a more distant single stranded region (*pseudoknot*). *Bulges* are formed if there is an excess of residues on one side of a duplex. The extra bases can either stack into the duplex or be looped out. *Internal loops* can occur when there is non Watson Crick apposition of bases. The length of internal loops in RNA is variable and some have been implicated as protein recognition sites. *Junctions* are regions which connect three or more *stems* (= base paired regions); for two stems indeed such junction is an internal loop. Junctions of up to 5 stems have been observed. A junction of three stems forms the catalytic core in the hammerhead structure of the self-cleaving RNA pdb-entry 1HMH (the first ribozyme solved by X-ray crystallography, 139).



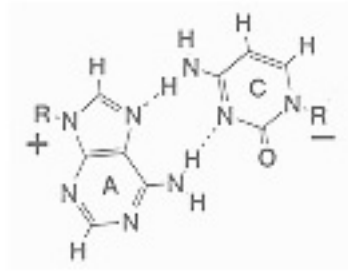
AC Wobble



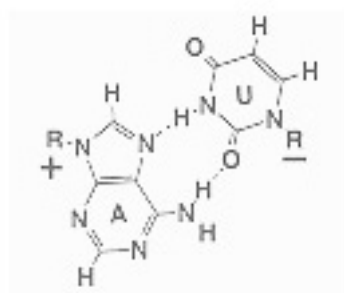
AC Reverse Wobble



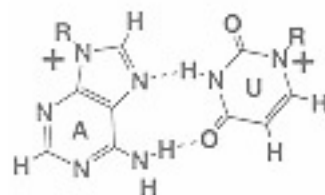
CG Reverse Watson-Crick



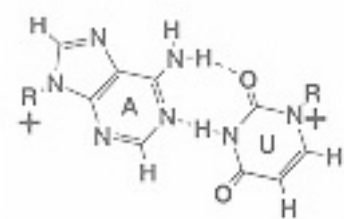
AC Reverse Hoogsteen



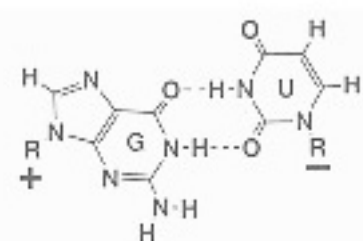
AU Reverse Hoogsteen



AU Hoogsteen

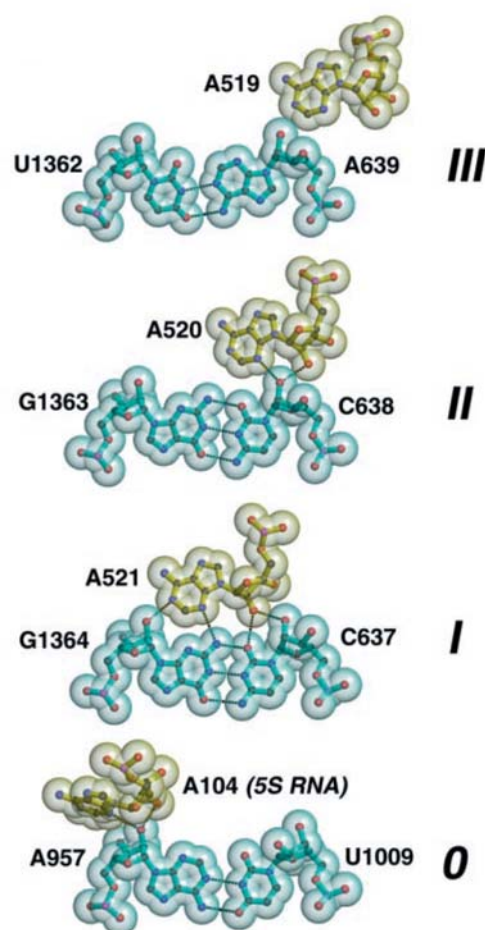


AU Reverse Watson-Crick



GU Wobble

Figure 1.12. Illustration of the variety of frequently encountered non standard base pairing in RNA. The illustration in color shows the four types of A-minor interaction established by Nissen et al. (133), referred to as type 0, type I, type II and type III.



The cloverleaf structure of tRNAs (21,81) is a prime example of four-stems junction and more generally of how RNA motifs organize into structurally complex- biologically functional units. Indeed tRNAs can all be folded into a common cloverleaf secondary structure adopting an invariant L-shaped tertiary structure (Fig.1.13). The prevailing structural feature of tRNA is an A-RNA structure (81) with many bulged bases and short loops that form additional interaction between nucleotides distant in the primary sequence, through unusual types of base pairs such as A-minor interactions that are crucial for the fold. Stems and loops across tRNA species are mostly conserved in size. The topologically complicated elbow region in such invariant L-shaped tRNA contains many non-Watson-Crick base pairs and base triplets. The helix-helix stacking of tRNAs has turned out to be a universal way of building other complex RNAs. Now the simplest RNA motif which exemplifies this feature is the kink-turn motif (chapter III).

Kink-turn RNA motifs

Kink-turn motifs (Fig.1.13) were recently characterized through the analysis of the 2.4Å-resolution crystal structure of the large ribosomal subunit from *Haloarcula marismortui* (141). Kink-turn motifs are characterized by a sharp bend ($\sim 120^\circ$, included angle of $\sim 60^\circ$, also frequently called “V-type” structure) of the phosphodiester backbone between a canonical helix (Watson-Crick base pairs, C-stem) and a helix that contains non-canonical (typically Hoogsteen) base pairs (termed NC stem). The tip of the “V” structure forms a short asymmetric loop comprising nominally unpaired nucleotides between the two flanking rigid stems. Kink (k)-turns have been found in conserved regions of ribosomal (r)RNA and messenger (m)RNA of archaea, prokarya and eukarya (142-144). K-turn motifs can serve as protein recognition site (143,145), mediate long range intra- and intermolecular motions in the ribosome (136,147) and recently was suggested to specify distal targeting in neurons (148).

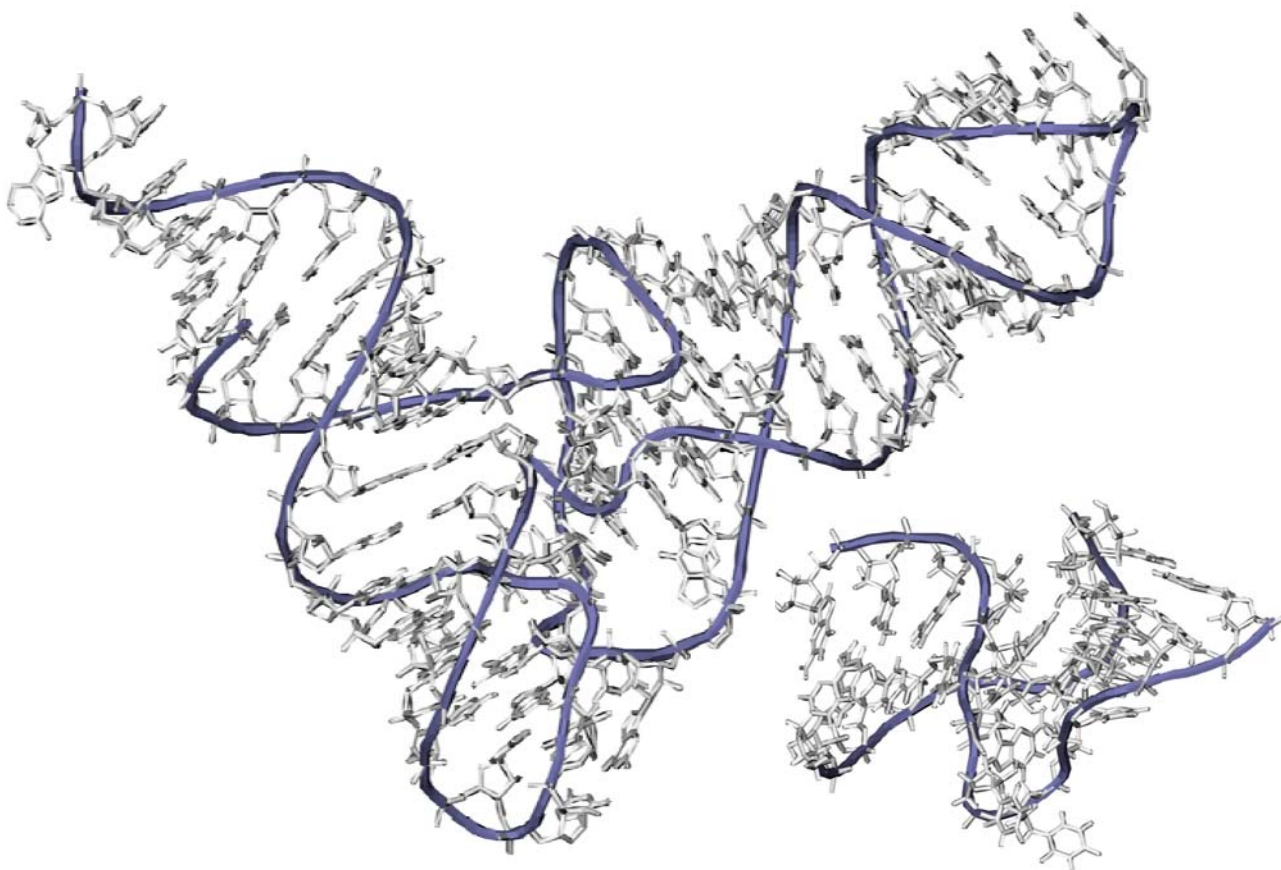


Figure 1.13. Computer generated graphics: the cloverleaf structure of tRNA (PDB-entry: 1EVV, 140) and the structure of a kink-turn (kt38). Both have a V-like topology.

Fluorescence resonance energy transfer (FRET) solution experiments (149) indicated that k-turns are highly flexible and exhibit a dynamic equilibrium between a tightly kinked

conformation and a more open unfolded structure similar to a simple bulged loop. Molecular dynamics (MD) simulation studies on several k-turn structures (136,147) indicated a significant mobility of the angle between the two rigid stems corresponding to alternative kink-turn substates. The biological significance of these findings is that k-turn motifs could mediate large scale hinge-like motions during both protein synthesis and the assembly of ribonucleoproteins. Experimental data on isolated k-turn structures indicate that it indeed can undergo global conformational changes upon metal ion association or protein binding (149-151). Their flexibility is notably thought to be essential for tRNA translocation between the A-, P- and E sites at the interface between the small and large bacterial ribosomal subunits during protein synthesis. For instance in the case of kt38, cryo-electron microscopy shows that the inter-subunit bridge B1a located at the tip of the tRNA acceptor (A) site finger (in helix 38 of the 50S subunit) moves up and down during the global ratchet-like rotation of the ribosomal machinery (152,153). This allows for a coordinated translocation of tRNA from the A to the P site because the A site finger is positioned in between A and P sites in a manner as to permit the passage of tRNA (154,155). It has been suggested that such large scale coordination during elongation is provided by the elbow flexibility of the kink-turn kt38 located near the base of the A-site finger (145). It is thus of significant interest to better understand free energy changes associated with global angular changes of this k-turn motif and this is the purpose of chapter III to do so.

Chapter II

Computer simulations in Molecular Mechanics

2.1 Nature and Implication of complexity when sampling the conformational landscape of biomolecules

*“Whatever complexity means,
most people agree that
Biological systems have it.”(156)*

Molecular models for biomolecules typically involve the jiggling and wiggling of thousands of atoms (157) and this high number of dimensions induces a complex conformational landscape, especially for simulations in explicit water. The Energy Minimisation (EM) technique is by definition challenged by this complexity (since it works at temperature 0 Kelvin). That is also why applications in Molecular Dynamics (MD) simulation to biomolecules had to wait for the advent of high-speed computing (158). The MD sampling problem that comes as a *corollary* is one of the two most urgent issues in the field of MD simulation (the other being the development of more accurate potential energy functions for representing the system) (159). Thus before to present the two techniques, EM and MD, it is of interest to understand what is meant by *complexity* of biomolecules.

Nucleic acids and other biomolecules can fold up in three dimensions into many intricate patterns with tertiary/quaternary interactions difficult to predict *a priori*. Most degrees of freedom are highly discontinuous in relation to hard core van-der-Walls repulsion induced by atomic surface proximity and to the combination of allowable rotamers for dihedral angles in the sugar-phosphate backbone. Moreover the entropy regime associated to each particular state creates a highly anisotropic width in local relief of the conformational landscape (160). Thus, the multidimensional potential energy surface that governs biomolecular structure has many maxima, minima and saddle points, this surface is rough, and highly anisotropic.

Despite the idealized DNA double helix has an axis of symmetry (161), in realistic environment many structural variations dependent on the sequence, solvent and binding agents will occur. Thus real nucleic acids *are not* symmetric molecules in general.

A complex dynamics is brought-about by the superposition of many fundamental motions (normal modes, 162,163) which induce cooperative effects. Thus a unique trajectory may not be within reach (only some representative trajectories) and “butterfly effects” can be expected in such a chaotic molecular system.

Solvent and salt influence structure and dynamics significantly (164) with a particular importance of salt (165) in long range electrostatic interactions for nucleic acids polyelectrolyte (negatively charged).

Finally diverse regimes of time scale characterize the dynamics of biomolecules (156) and this creates a challenge in practice to tackle all of the above complexity. For example the normal mode corresponding to global bending of DNA and RNA (chapter III-IV) has characteristic times in the range 10^{-3} (166) to 10^{-2} s (159) and is a relatively slow conformational change (let ignore evolutionary timescales here). But the patterns of dynamics characterising inter base pair parameters that trigger global bending, i.e. local hinge/bend/kink in DNA, have relaxation times on the order 1-2 ns (167). An example of yet faster metastable state is nucleic acid backbone dihedral conformations (chapter V-VI) with almost instantaneous relaxation time (167). Finally highest frequency vibrations which correspond to bond stretching must also be resolved for numerical stability when integrating the equation of motion in molecular dynamics (section 2.4). Thus a time step $\Delta t = 1$ fs is generally used in molecular dynamics ($\Delta t = 2$ fs if X-H bonds are frozen, see SHAKE (168) in 2.6). Given that at each step the force and energy acting on each particle (based on the high-dimensional *force field* described in 2.2) have to be calculated, the timestep problem (one million iterations for 2 ns of dynamics even when using SHAKE) obviously precludes to capture the many facets of biomolecules complexity from any single MD trajectory.

2.2 Energy function

Force field development for molecular mechanics calculations on biomolecules (EM, MD) goes on since nearly fifty years (169). In molecular mechanics the molecule is represented as a system of particles (atoms) and springs (bonds). In MD *atoms* are actually effective nuclear centres in the sense that they represent a parametric description of an atomic nucleus together

with its electronic cloud *in a given environment*. A “given environment” means that the parametric energy function, to be described below, reads parameter in a list of atom types where for example a carbon is identified by 20-40 entities depending on the chemical groups to which it is attached. Each charged nucleus defines a net atomic charge. This passage from the quantum to the classical world is based on the Born-Oppenheimer approximation (8) and is the basic assumption of any EM / MD molecular mechanics calculation. A second assumption is transferability of a set of parameters developed on a relatively small number of chemical groups to (e.g.) large polymers composed of these chemical groups. Another assumption is the use of pairwise additive potentials to express the potential energy of the system, into a semi-empirical function (force-field) that takes the following typical form:

$$\begin{aligned}
 V(\mathbf{x}) = & \sum_{bonds} \frac{K_i}{2} (l_i - l_{i,0})^2 + \sum_{angles} \frac{H_i}{2} (\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{V_i n}{2} (1 + \cos(n\omega_i - \gamma)) \\
 & + \sum_{i=1}^N \sum_{j=i+1}^N 4\epsilon_{ij} \left[\left(\frac{A_{ij}}{r_{ij}} \right)^{12} - \left(\frac{B_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}
 \end{aligned} \tag{eq. 2.1}$$

The first line of energy terms correspond to the bonded interactions and the second line to the non-bonded interactions. The first three terms (first line) sum over the sets of all bonds, bond angles and dihedral angles respectively. The two last terms sum respectively Lennard-Jones and Coulombic potentials over all (i, j) atom pairs separated by three bonds or more. The symbols l_i , θ_i , ω_i , r_{ij} denote internal variables of bond length, bond angle, dihedral angle and inter-atomic distance respectively. Symbols with sub-script 0 are the corresponding reference values (values of highest probability). The parameters K_i , H_i , $V_i n$, A_{ij} , B_{ij} , q_i , q_j are energy constants (see discussion below).

Covalent bonds and bond-angles are considered harmonic springs hence their energy potential generally follows a simple Hooke's law. Each bonded quartet contributes to the force field by a torsional potential expressed as a cosine series expansion. In this term the multiplicity n gives the number of minima in the cosine function as the central bond of the torsion is rotated through 360°. The phase factor γ determines which dihedral angle value(s) correspond to these minima. For example $n = 3$, $\gamma = 0$ gives a threefold rotational profile with minima at torsion angles of +60°, 180° and 300° and maxima at 0°, 120° and 240°. For van-der-Walls interactions a 12-6 potential is often used to represent the balance between short range repulsive forces and longer range attractive forces (London dispersion interactions).

Electrostatic interactions take the form of a Coulombic potential which includes a possible modification by, e.g. a distance-dependent dielectric function $\epsilon(r)$ (ex.: *Flex* force field in JUMNA (22)) or a more complex function which represents implicitly some properties of a solvent such as screening effects (ex: atomic radii based on a Generalized Born model, 170). The force field parameters (force constant and reference values in each term above of $V(\mathbf{x})$) are obtained from either quantum mechanics, experimental measurement such as vibrational spectra, heats of formation, crystal properties (171) or through empirical trial and error to reproduce observed properties of model systems. A recent illustration of this process is the Amber force field for nucleic acids which recently has been found to overstabilize non standard backbone dihedral conformation (alpha / gamma flips) when first simulations lasting up to one microsecond on relatively large polynucleotide (>10 bp) appeared (124). This has initiated the development of a new force field, parmbsc0 (19) which correctly represents DNA and RNA backbone rotamers in MD simulations up to one microsecond (19,172). This force field development is an example where corrections by quantum mechanics for defining a new atom type (involved in the alpha / gamma dihedral definition and called "CI", 19) and the reproduction of experimental transition kinetics / statistical distribution in x-ray crystal databases of nucleic acid backbone dihedrals, have validated the new set of parameters (172). For this reason, calculations reported in this thesis use the parmbsc0 force field.

On one hand the parametric energy function can be simplified by the introduction of constraint on degrees of freedom irrelevant for the calculations carried-out such as every X-H bond length (SHAKE, 168) in dynamical simulations or every bond length and valence angle in conformational search algorithms such as JUMNA (22, see 2.3). On the other hand the additional inclusion of electronic polarizability of atomic particles in MD is an active field of Research and first polarizable force fields for biomolecules emerge (173,174). The use of continuum representations, e.g. of the solvent, also implies modification of the force field noteworthy the electrostatic term using Generalized Born radii or the Poisson-Boltzmann equation (175). Hybrid representations of the system such as in Quantum Mechanics Molecular Mechanics (QMMM) simulations (176) also requires modification of the force field representation but are beyond the scope of this thesis. Now with a force field given, the favourable regions of the configuration space under its influence (minimisation) or the dynamics of molecular motions (molecular dynamics) can be studied.

2.3 Energy Minimisation (EM)

Minimisation algorithms are used to identify conformations of the molecular system corresponding to minimum points on the potential energy surface. Those minima correspond to the most relatively stable geometrical arrangements for a given force field. Those minima are located using numerical methods which gradually change the coordinates to produce configurations with lower and lower energies until the minimum is reached. The minimisation operation performed at each step when using both quantum mechanics and high order derivatives of a molecular mechanics energy function becomes formidable for a system containing thousands of atoms. As a consequence, most EM algorithms providing reasonable efficiency in the context of biomolecules are first (e.g. steepest descent, conjugate gradients) or second (e.g. Newton-Raphson) order derivatives numerical methods (derivatives are generally calculated analytically). It is important to note that since EM aims to move as fast as possible downhill to the bottom of the nearest minima, the solution to the EM problem can be very different depending on the starting point on the energy surface and is generally not the global minimum.

The method of *conjugate gradients* used in chapter IV is a first order minimisation method that can be briefly described as follows (see (171) for more details about EM methods). Provided an initial 3N component vector (N = number of atoms) $\mathbf{x}_1$ describing the configuration of the system, $\mathbf{x}_2$ is obtained by moving along the direction parallel to the net force acting on each particle $-\frac{\partial V(\mathbf{x}_k)}{\partial x_k^i}$, (where $V(\mathbf{x}_k)$ is the potential energy of $\mathbf{x}_k$, k refers to

the minimisation step and subscript i refers to a given particle) which is the direction of the first derivative of the energy (the gradient). This first step is equivalent to *steepest descent*. The one dimensional minimum along this direction can be obtained by a standard *line search* or an arbitrary step method (see (171) pp. 262-4). Then for every subsequent configurations $\mathbf{x}_k$, the direction along which to move is conjugate to the previous direction and obtained as

$$dir_k = -\nabla V_k + \frac{\nabla V_k \times \nabla V_k}{\nabla V_{k-1} \times \nabla V_{k-1}} \times dir_{k-1} \quad \text{in the original Fletcher-Reeves algorithm (177).}$$

Conjugate gradients methods have notably the property that for a quadratic function of M variables the minimum will be reached in M steps.

As will be done in chapter IV, with EM some conformational feature can be adiabatically separated and fixed in successive EM windows along its range of allowable values while all

the remaining variables are energy relaxed. Such adiabatic mapping is used in two-dimensions in chapter IV for calculating the optimal energy pathways involved in the bending of several DNA oligomers, as a function of the bend magnitude and of the bend direction.

The JUMNA software

JUMNA (*J*unction *M*inimisation of *N*ucleic *A*cids, 22) is a very flexible program for energy optimisation of nucleic-acids and nucleic-acid ligand complexes. It is possible to build virtually any kind of unusual nucleic acid structure within JUMNA and many analysis tools such as an automatic 1D (2D) adiabatic mapping are included.

The application of helical symmetry leads to a reduced set of independent variables (internal helical parameters) for representing the system, typically 100 times less than when working in $3N$ cartesian coordinates (N = number of atoms). JUMNA builds-up nucleic acid systems as a series of (3' monophosphate) nucleotides positioned in space with respect to a common helical axis system using three translations T (X-disp, Y-disp and rise) and three rotations R (inclination, tip and twist) as defined in the 1988 Cambridge workshop (see 1.3.1). The junctions between consecutive nucleotides can be opened at each step of the minimisation procedure to ease transition between conformational states separated by high energy barriers, in particular at the initial build-up of a potentially unusual (high energy) conformation.

In total the independent variables associated to each nucleotide are the three T and three R above, the glycosidic dihedral (χ), three valence angles and two dihedrals within the sugar moiety and two backbone dihedrals (ϵ and ζ). All other variables are dependent and determined by the closure condition between nucleotides. The latter is imposed via harmonic energy penalty terms in the force field. Moreover introducing changes to the energy formulation (additional penalty terms) is straightforward within the algorithmic of JUMNA which incorporates tools for testing the accuracy of any new analytical derivatives. JUMNA uses the conjugate gradients algorithm and analytical derivatives of the potential energy with respect to every independent and dependent variables (combined using the chain rule for derivative of composite functions). JUMNA now offers several choices of continuum solvent representation and force field including the Amber series. We will use (chapter IV) mainly the Amber parm94 function in conjunction with a GB (170) representation of electrostatic interactions. Note also that the program was recently combined to a Monte-Carlo sampling scheme in order to carry-out conformational sampling in the JUMNA framework (178).

2.4 Simulation in Molecular Dynamics:

A finite difference method

Molecular dynamics simulation is a sampling method based on discrete timestepping for integrating the second classical law of motion (Newton's law). The equation of motion takes the form $F = M \times a$ where F is a matrix of forces acting on each particle (minus the gradient of the potential energy function $-\frac{\partial V(\mathbf{x}_k)}{\partial x_k^i}$ as in 2.2), M is a diagonal matrix of atomic masses for every particle and a is the second time-derivative of the position vector in cartesian coordinates. There are many algorithms for integrating the equations of motion using finite difference method but all assume that the stepping can be approximated as Taylor series expansions:

$$\begin{aligned} q(t + \delta t) &= q(t) + \delta t v(t) + \frac{1}{2} \delta t^2 a(t) + \frac{1}{6} \delta t^3 b(t) + \frac{1}{24} \delta t^4 c(t) + \dots \\ v(t + \delta t) &= v(t) + \delta t a(t) + \frac{1}{2} \delta t^2 b(t) + \frac{1}{6} \delta t^3 c(t) + \dots \\ a(t + \delta t) &= a(t) + \delta t b(t) + \frac{1}{2} \delta t^2 c(t) + \dots \\ b(t + \delta t) &= b(t) + \delta t c(t) + \dots \end{aligned} \tag{eq. 2.2}$$

where q is the position in cartesian coordinate, v is the velocity (first time derivative of the position q), a is the acceleration (second time derivative of q), b is the third time derivative, and so on. Finite difference integration algorithms which are widely used have been described generously in some textbooks and in (179) and thus a detailed description is not reproduced here. The program sander of the Amber package with which are performed every molecular dynamics simulations in the following chapters uses the Velocity Verlet integration scheme by default. The latter is expressed through the two following simplified equations:

$$\begin{aligned} q(t + \delta t) &= q(t) + \delta t v(t) + \frac{1}{2} \delta t^2 a(t) \\ v(t + \delta t) &= v(t) + \frac{1}{2} \delta t [a(t) + a(t + \delta t)] \end{aligned} \tag{eq. 2.3}$$

which require first calculating $q(t + \delta t)$, then storing both $v(t)$ and $a(t)$ in the less expensive form $v(t + \frac{1}{2}\delta t) = v(t) + \frac{1}{2}\delta t a(t)$, calculating $a(t + \delta t)$ from new positions $q(t + \delta t)$ and finally $v(t + \delta t) = v(t + \frac{1}{2}\delta t) + \frac{1}{2}\delta t a(t + \delta t)$. Velocity Verlet is a relatively fast and precise algorithm *per se* (see (171) pp. 356-358 for a discussion) but one has to keep in mind that the most demanding part of the calculations (> 95 % !) in a molecular dynamics simulation is the calculation of the force (and interaction energy) for each particle in the system. This means the calculation of the term $a(x)$ at each step in eq. 2.3 and this operation is common to every standard integration algorithm of molecular dynamics.

2.5 Sampling Molecular Dynamics Ensembles

Mathematically the Newton's equation $F = m.a$ is equivalent to an Hamiltonian system of the form $\dot{q} = \nabla_p H$, $\dot{p} = -\nabla_q H$ for some smooth total energy function $H(q,p)$, the Hamiltonian, where q is the particle's position and p the particle's momentum and the dot indicates time derivative. This gives rise to a form of error analysis, that is, how the numerical trajectory deviates from the true trajectory solution to the corresponding Hamiltonian system of differential equations? In particular a symplectic invariant characterizes any Hamiltonian system as well as time reversibility and thus the numerical approximation of molecular dynamics algorithms should be symplectic as well as time reversible.

Symplecticness is a topological property somewhat difficult to state in high dimension (see 179 for a survey). Easier to state is the Liouville theorem which is essentially a conservation law for probability density (180). Clearly the Liouville theorem implies that if q, p represent coordinates in a space with $6N$ dimensions (N = number of particles) then the probability density of any particular point in this phase space loses its *time dependence* on reaching a dynamical equilibrium. This is also stated in the fundamental statistical postulate of equal a priori probabilities: "An isolated system in equilibrium is equally likely to be in any of its accessible states" (181). Such system is then said ergodic. But since the time necessary to attain this state is often of "astronomic" order in high dimension, i.e. when $N \sim 10^3$, molecular dynamics simulators expect at best quasi-ergodic situations when monitoring equilibrium (180,182).

Such quasi-ergodic situation is at the basis of another fundamental postulate when analysing molecular dynamics trajectories, that is, the assimilation of time-average properties to Ensemble-average properties. These postulates are at the basis of statistical mechanics for they are necessary if one is to link the microscopic level (positions, momenta) to the macroscopic state functions such as Temperature or Pressure. This link is encapsulated in the famous entropy relation $S \equiv k \ln \Omega$ where k is Boltzmann constant and Ω the phase space density (partition function). Thereby the *statistical* equilibrium discussed above corresponds in the molecular dynamics simulation to a *thermodynamics* equilibrium (180, 182). With an additional classical property that Newton's equations of motion conserve the total energy (only approximately in numerical simulation) the default sampling of MD is at constant energy (E), simulation box size (volume, V) and number of particles (N).

Other thermodynamics ensembles than the default microcanonical-NVE can be sampled with MD as well. This requires specifying which thermodynamic variable is to be kept constant through the use of additional constraint. The two most common Ensembles are constant NVT (canonical) and constant NPT (isothermal-isobaric). The temperature T can be kept constant by a weak coupling of the system to an external heat bath at the desired temperature through a scaling factor for particle velocities (used in every MD simulation here, see 181 or the Amber manual). Indeed the temperature of the system is related to the time average of the kinetic energy K by

$$\langle K \rangle_{NVT} = \frac{3}{2} N k_B T \quad \text{eq. 2.4}$$

where k_B is the Boltzmann constant and $K = \frac{1}{2} \sum m_i v_i^2$. The weak coupling algorithm (183) uses a coupling constant generally higher than the time step for velocity scaling, which permits to monitor the amplitude of kinetic energy fluctuation about the desired temperature and avoids in particular the formation of local hot spots in the system. Note that the average total energy should remain constant also in NVT.

In order to perform a MD simulation in constant NPT one needs further to specify a constant pressure P and allow the volume of the simulation cell to fluctuate. Ensembles at constant pressure and temperature are of interest because they often better reflect the conditions of experimental measurements (as well as biological compartment *in vivo*, from the mitochondria and cells to the inter-organs' circuitry). In simulations to come, the pressure is also kept constant by an algorithm of the "weak coupling" variety (183). In that case an

external pressure-bath is coupled to the system through a scaling factor which translates every particle coordinate, isotropically in x,y,z directions, and thus changes the volume of the simulation box.

To conclude the section, as mentioned in 2.1 the curse of dimensionality in biomolecular systems ($3N$ with $N \sim 10^3$) involves structural events with multiple time scales. The MD timestep problem, that is Δt on the order of the femtosecond, means that even a “quasi-ergodic” thermodynamic equilibrium as described above is rarely attained (challenge stated in 2.7).

2.6 Some practical aspects and the standard protocol

Here I introduce what I think are other important practical aspects when dealing with the molecular dynamics approach.

Shake

As mentioned the SHAKE (168) algorithm (and the analogue RATTLE (184) developed specifically for velocity-Verlet integration schemes) can increase the time step by freezing every X-H bonds. SHAKE is an iterative procedure to re-evaluate the matrix of inter particle forces and determine the forces directed along given covalent bonds (e.g. X-H) in order to keep these bond length constants. The technique has been extended to handle more general geometrical constraints (e.g. side chains) (185) and to fit the Verlet, velocity Verlet (RATTLE, 184) and higher order predictor-corrector algorithms (186). It has been shown that the introduction of bond length constraints has little effect on structure and dynamics in MD but further application on bond angles seriously affect torsion angles distribution (187). In two words bond length constraints are permissible in MD and SHAKE opens the possibility of sampling with a time step $\Delta t = 2$ fs.

Initial velocities

A second practical aspect of MD is the assignment of initial particle-velocities (at $t = 0$). This can be done using eq. 2.4 which relates the temperature to the average particle velocity and

the equipartition theorem for classical system stating that each degree of freedom has $\frac{1}{2} k_B T$ kinetic energy at thermal equilibrium. This gives rise to the Maxwell-Boltzmann velocity distribution f_V which can be obtained by randomly choosing velocities from a Gaussian distribution with variance $(k_B T / m)$ in each dimension. For the x-coordinate this Gaussian function is: $f_V^x = \sqrt{m_i / 2\pi k_B T} \times \exp\left[\frac{-m_i (v_i^x)^2}{2k_B T}\right]$, and $f_V = f_V^x \times f_V^y \times f_V^z$. A round of MD time steps equilibration (see 2.1) is necessary before thermal equilibrium can be reached.

Periodic boundary conditions

In order to study bulk phenomena from a relatively small number of particles as is done in molecular dynamics simulations, it is necessary to remedy the problem posed by finite boundary condition (surface effects between the box “walls” and the interacting particles). The use of periodic boundary conditions enables to calculate a trajectory as if the system dimensions were infinites. It consists to replicate the simulation cell in all directions to give a periodic array of identical boxes (the size of the array is indeed infinite but in practice only next nearest neighbour cells will be involved in the scheme thus the infinite property is self-implied). If a particle leaves the box then it is replaced by one image particle: the one entering through the opposite side. The implementation of periodic boundary conditions has a great deal of implications both in generation (see the *cutoff* section) and analysis of MD trajectories but this is beyond the scope of this section (see (182) for example).

Cutoffs

The most time consuming part of molecular dynamics simulation is the calculation of non bonded interactions (energies and forces). For a pairwise interaction force field without polarisation, the number of non bonded terms grows as the square of the number of atoms (N^2). The use of cutoff, i.e. a distance beyond which the interaction is set to zero, is a widespread approximation given in particular the r^{-6} distance dependence of the dispersion interaction in the Lennard-Jones potential (first MD simulations were mostly on simple molecule e.g. Argon, under a Lennard-Jones potential, 182). Several techniques (shifted potential, switching function) were developed to switch off as smoothly as possible the

potential (energy/force) and to avoid lack of energy conservation or dynamical instabilities induced by the use of cutoffs.

When using periodic boundary conditions the minimum image convention is applied for calculating interactions between any one particle and at most one image of each other particle, and not with its own image. In order to do so a cutoff distance of no more than half the length of the cell has to be used (171, let take the example of a cubic cell). For long range interactions (e.g. those that decay no faster than r^{-N}) such as electrostatic interactions which are of primary concern when dealing with negatively charged nucleic acids (188), some special methods have been devised noteworthy because their range generally exceeds half the box length. Current MD simulations use the Particle Mesh Ewald (PME, 189) method which is a fast implementation of the Ewald summation method (190) for calculating the full electrostatic energy of a unit cell in a macroscopic lattice of repeating images (see 189).

The standard protocol

Once a force field, an algorithm for integrating the equations of motion, a thermodynamics ensemble and a time step size have been chosen, the indispensable ingredient that remains to choose before to start the MD calculation is an initial set of positions for the solute. This can come from an X-ray or NMR derived conformation or an original build-up, based for example on the standard B-DNA Arnott conformation. An original build-up is used here and can at least pretend unbiased comparison with experimental properties when one attempts to validate the realism of a MD trajectory.

On introducing the solvent molecules (in explicit environment), unrealistic overlaps between particle's Lennard-Jones interaction-distances may lead to strong repulsion which would immediately create instability upon dynamics (e.g. rupture of covalent bonds or explosion). Thus the first step (Fig.1.14) is to carry-out a stage of EM to relax strain contacts. Then a multi-stage equilibration is performed where the simulation cell is gradually heated up to the desired temperature with strong harmonic restraints which freeze the solute in its minimized conformation. In current application we always performed three subsequent heating dynamics: 200 ps at 100 K, 200 ps at 200 K and 200 ps at 300 K. At the temperature of interest (300K), the harmonic restraints on the solute atoms are reduced down to zero again in a multi-stage simulation (200 ps each), such that the solute adapt to the bath of explicit water molecules (we use the non polarizable TIP3P three-sites-model for H₂O, 191). A longer (1-5 ns) equilibration step is also often performed at this point since it corresponds to the final

condition of the simulation. At the end of this equilibration period, all memory of the initial configuration should have been lost, which can be checked by testing the protocol on a different starting conformation (or e.g. a different seed for the random number generator used when assigning the initial velocity distribution upon heating).

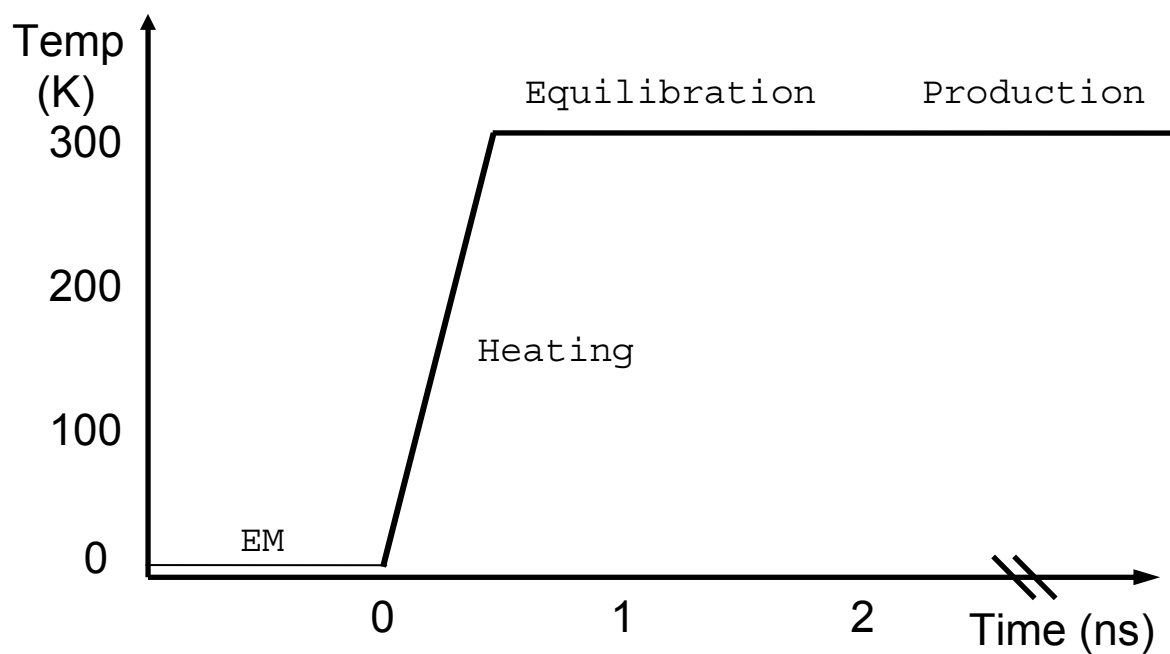


Figure 1.14. The four stages of the molecular dynamics recipe: Energy Minimisation (EM), Heating, Equilibration and Production.

More generally system equilibration is monitored by recording instantaneous values of state functions such as potential energy or pressure or more detailed conformational parameters such as the root mean square deviation in atomic positions. These parameters should oscillate about steady mean values without further systematic drift before the production sampling phase is started. Then, during the production some specific parameters that evaluate the quality of sampling (the ergodic measure (192), the statistical inefficiency (180) and other space/time auto-correlation functions (193)) can be recorded in real time; or afterward during the analysis of the trajectory.

2.7 The challenge

In the canonical Ensemble, the probability density of conformations $\rho(q, p)$ is:

$$\rho(q, p) = \exp\left(\frac{-H(q, p)}{k_B T}\right) / Q, \quad \text{eq. 2.5}$$

($Q = \frac{1}{N!} \frac{1}{h^{3N}} \iint dp dq \exp\left(\frac{-H(q, p)}{k_B T}\right)$ is the partition function / normalising factor; h is

Planck constant and ensures consistency with quantum mechanics; the other symbols have been described). Thus $\rho(q, p)$ is proportional to the so-called *Boltzmann factor* (the numerator). From the Boltzmann factor it appears that conformations with high energy are exponentially unlikely and this property creates an additional sampling problem illustrated on the drawing of Fig.1.15 (simplified from Fig.2.1 in ref. 180).

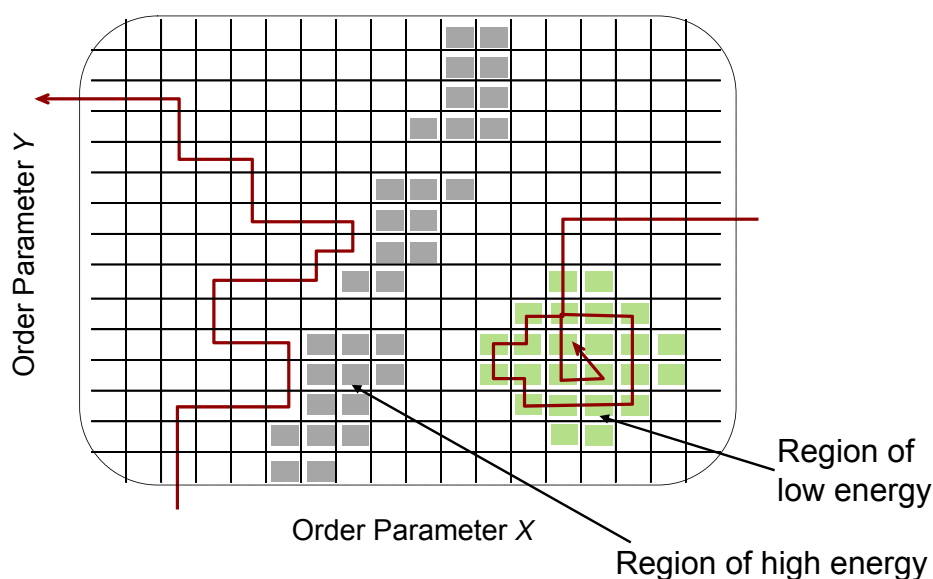


Figure 1.15. Schematic representation of the molecular dynamics phase space and two trajectories

The chip-like array represents a section of the $6N$ -dimension phase space (N = number of particles) in just two dimensions and each square corresponds to a point in this phase space. The region shaded in grey, a region of high energy, is referred to as a bottleneck in statistical sampling. Since these grey squares are exponentially unlikely to be sampled, the region shaded in green, a region of low energy and thus crucial for an accurate understanding of the conformational landscape, may remain inaccessible during the entire length of a trajectory sampled by MD (the red path on the left). Another trajectory initiated in the right hand side of

the phase space of Fig.1.15, by using a different starting conformation for example, may quickly get kinetically trapped in the green region. This alternative trajectory thus also may never reach the whole ensemble of conformations on the left hand side of the phase space. In both cases, the presence of high energy bottlenecks or low energy kinetic trapping, none of these two MD trajectories would be representative of the conformational landscape at equilibrium. This sampling problem in molecular dynamics is referred to as the broken ergodicity problem.

2.8 Umbrella Sampling simulations

A standard “staging” sampling method aimed at improving the sampling quality in rugged energy landscape is umbrella sampling developed by Torrie et Valleau (11). Combined with weighted histogram analysis method (WHAM, 194) to reconstruct probability distributions from explicitly biased histograms, umbrella sampling attempts to calculate the free energy pathway (195) associated to a given reaction coordinate (potential of mean force, PMF, 196). The umbrella coordinate may be a simple distance between two groups of atoms as in chapter III or a more complicated generalized coordinate as in chapter IV.

Umbrella sampling consists to perform a series of independent simulation ‘windows’ along a given reaction coordinate (order parameter X on Fig.1.15) with consecutive values of the reaction coordinate restrained by a penalty term in the consecutive windows. Thus for each window i , the value x_i of reaction coordinate x is modified by a small amount and restrained to the new value x_{i+1} in window $i+1$ using an harmonic biasing potential $v_i(x)$. These sampling windows can be simulated either consecutively, or in parallel if a pre-equilibrated path is already available.

The WHAM method (194) is then used to calculate a potential of mean force (196) or free energy associated with the reaction coordinate. The WHAM equations express the optimal estimate for the unbiased probability distribution $P(x)$ using all data point available over each individual biased probability histogram $P_i^*(x)$,

$$P(x) = \frac{\sum n_i P_i^*(x)}{\sum n_i \exp\left(\frac{F_i - v_i(x)}{k_B T}\right)} \quad \text{eq. 2.6}$$

where n_i is the number of data points used to construct the biased distribution function in window i , $k_B T$ is the thermal energy per particle and each sum runs over the total number N^w of windows sampled (be careful $N^w \neq N$). The free energy constant F_i is itself determined using the expression of the optimal estimate for the probability distribution function $P(x)$,

$$F_i = -k_B T \times \int \ln P(x) \times \exp\left(\frac{-v_i(x)}{k_B T}\right) dx \quad \text{eq. 2.7}$$

The two equations eq2.6 and eq.2.7 have to be solved self-consistently: starting from an initial guess for the N^w free energy constants F_i to estimate the unbiased distribution function $P(x)$ via eq. 2.6, the latter is subsequently used in eq. 2.7 to generate a new estimates for the N^w constants F_i , these are used in turn in eq. 2.6, and so on. The iteration cycle is repeated until both equations are satisfied, i.e. the largest change in the set of values F_i on two consecutive iterations is below a tolerance index (here 10^{-3}). The relative free energy (PMF) at a given bend angle is defined by $W(x) = -k_B T \times \ln P(x)$. Checking the convergence of the resulting PMF can notably be made with a backward simulation using the same protocol of extraction with WHAM. The backward simulation consists to use the final structure of the sampling window restrained with the maximum value of the reaction coordinate. Additionally, one also extracts the PMF from the cumulative sampling after different time-increments in each umbrella window and compare the overlap between these curves.

2.9 Replica Exchange sampling methods

Parallel tempering (Replica Exchange, 13) is commonly used to deal with the problem of quasi-ergodicity (2.7) during molecular dynamics simulation of biomolecules. As illustrated in Fig.1.16, the main difference between replica exchange and umbrella sampling is that the independent trajectories in each window need to be sampled simultaneously. Indeed replica exchange is a sampling technique whereby an exchange of information between sampling windows (*replicas*) procure reciprocal advantages where hence an enhanced sampling for each of the replicas. In the original version of replica exchange the reaction coordinate was the Temperature (scaling of the whole system) and thus the Temperature replica exchange method (TREM, 13) is first described. More detailed introduction to replica exchange approaches and some extensions will be found in (197).

2.9.1 Temperature REM

Temperature replica exchange molecular dynamics can be seen as an improved update of some independent canonical MD simulations at different temperatures (replicas) through an allowed diffusion (exchange) of conformations across the different replicas. In this method some independent trajectories evolve simultaneously with different temperature and a criterion (Detailed Balance) is defined so that at a given period (every picosecond for example) the sampled conformations can be exchanged between some pairs of replicas. The Markov chain used to swap conformations operates alternatively on every odd and even pairs of replica so that as time progresses, conformations can diffuse across the full ladder of replicas. As a result, the high variety of conformations sampled in the replicas with high temperature periodically takes over any trapped conformation in the replicas that sample the low temperature regime (regime close to 300 K). In TREM the dynamics is lost in each of the replica, however each replica and in particular the replica at physiological temperature, as justified below, can be considered to be a canonical Ensemble of conformations.

Because the replicas are run independently and simultaneously, each set of simultaneous conformations for each replica $C = \{C_1, C_2, \dots, C_n\}$ (n : number of replicas) can be seen as a state in a generalized Ensemble for which assigned weight is simply the product of every C_i conformation Boltzmann weights. Hence,

$$W_{GE}(C) = \exp \sum_i^n \beta_i H(C_i) \quad \text{eq. 2.8}$$

where GE stands for generalized Ensemble. The non-Boltzmann weight factor eq.2.8 given to a “transverse” state in the set of replicas leads to realize a Markov chain in the generalized Ensemble of the compound system. Conformation exchanges inside the transverse state conserve a canonical distribution in each replica if the detailed balance condition is imposed on the exchange criterion. This condition takes the form of the equation of microreversibility:

$$W_{GE}(C) \omega(C \rightarrow C') = W_{GE}(C') \omega(C' \rightarrow C) \quad \text{eq. 2.9}$$

Such an exchange can be accepted or rejected based on the usual Metropolis criterion (198). Noting i, j two different replicas corresponding to two different temperatures and q_i, q_j their N

atoms coordinate vector sets in C , the simplified criterion becomes

$$\omega(C \rightarrow C') = \min(1, \exp(-\Delta)),$$

$$\Delta = \beta_i(U_i(q_j) - U_i(q_i)) - \beta_j(U_j(q_j) - U_j(q_i))$$
eq. 2.10

where U is the potential energy of the system and β the inverse thermal energy $1/k_B T$. From eq. 2.10 we see that the acceptance ratio decreases exponentially with $(\beta_i - \beta_j)$. This is why exchanges are attempted only between pairs of consecutive replicas, for each pair must show overlapping potential energy histograms for the acceptance probability to be significant.

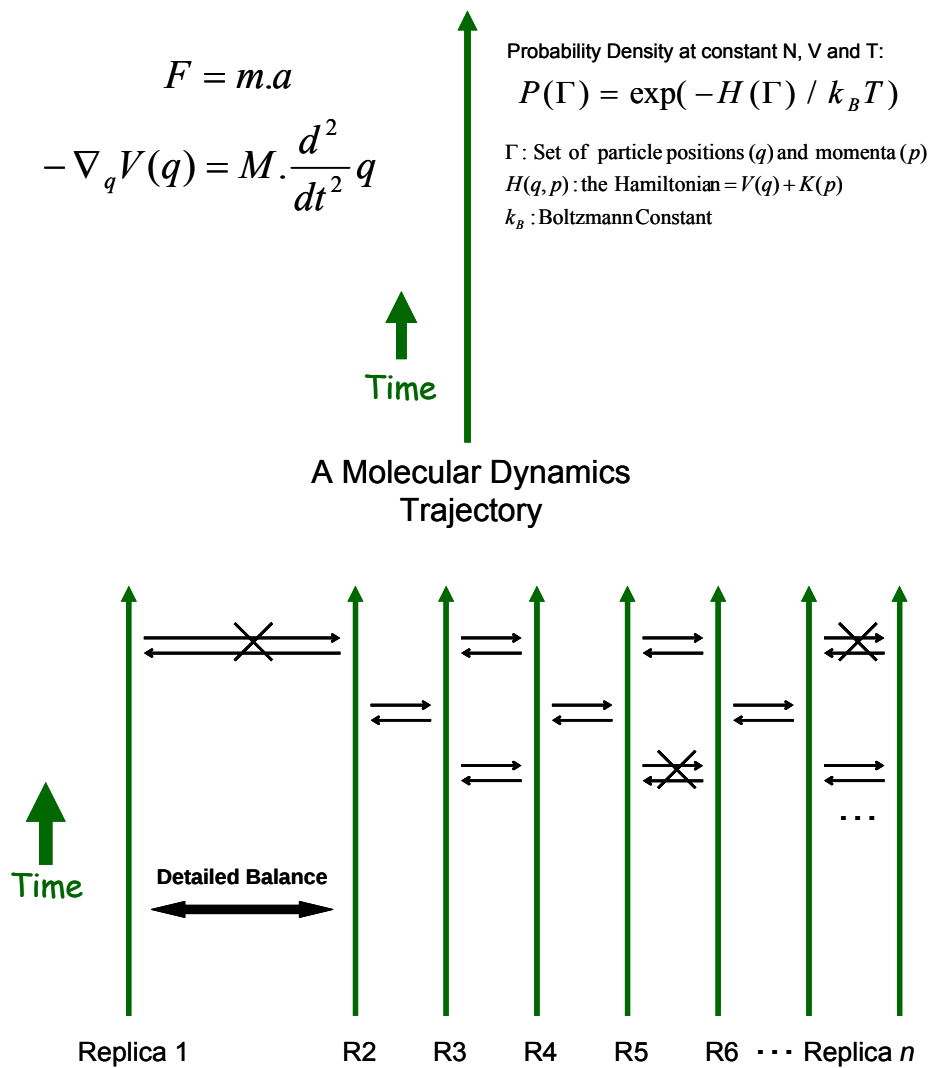


Figure 1.16. Time evolution in MD (up) and in replica exchange (down) simulations. Green arrows are some MD trajectories and black arrows indicate a swap of conformation.

2.9.2 Hamiltonian REM

Expressions eq. 2.8 and eq. 2.10 for $w(C_{old} \rightarrow C_{new})$ have made TREM simulations attractive compared to other generalized Ensemble approaches because the probability weight factors are known *a priori*. This owes to the fact that replicas are non-interacting. The non interaction of the replicas induces a second advantage in that it also allows to use different Hamiltonians for the different replicas, as such defining a multi-dimensional (temperature + Hamiltonian parameters) replica exchange process (15). In that case a random walk both in temperature space and in a predefined Hamiltonian parameter space is realized for each of the replicas. Thereby writing the potential energy function U as a sum of replica dependent term v and replica independent term ($U - v$) and taking the same temperature ($\beta_i = \beta_j$) for every replica, one defines a “parameter” space of replica exchanges for which the acceptance ratio is greatly increased since it scales only with the number of atoms’ coordinates implied in the definition of the term v that differ between the replicas:

$$\Delta = \beta (v_i(q_j) - v_i(q_i) + v_j(q_i) - v_j(q_j)) \quad \text{eq. 2.10'}$$

The Hamiltonian replica exchange (14,15,199) also called Model Hopping (200), Multi-Self-Overlap Ensembles (201) or Biasing Potential replica exchange (for extended Hamiltonians, 202) focuses the enhanced thermalization process of the system on some particular degrees of freedom. Compared to standard TREM, HREM can thus be used to work with increased system size or to increase the statistical efficiency on a given system as has been shown on polypeptide systems (200). In case where v is an additional penalty term (extended Hamiltonian) HREM is homologue to the umbrella sampling scheme but where sampling windows are simulated simultaneously and conformation exchanges can improve the sampling in the coordinates perpendicular to the reaction coordinate in each window.

Chapter III

RNA bending and the flexibility of ribosomal kink turn motifs

. **Keywords:** free energy simulations; kt38 motif; RNA flexibility; kink turn meta-stable states.

. **Reference publications:** Curuksu, J., Sponer, J., Zacharias, M. (2009) *RNA* (submitted).

. **Recommended introductions:** 1.5.1, 1.5.2, 2.8 and 2.9.

3.1 Abstract

Kink-turns (k-turns) are common structural motifs that can introduce sharp kinks into double stranded RNA and have been proposed to mediate large scale motions in the ribosome. K-turns consist of a bulge loop region flanked by trans sugar-Hoogsteen G:A pairs and the sharp kink conformation is stabilized by A-minor interactions. Umbrella sampling molecular dynamics simulations were used to disrupt an A-minor interaction in the ribosomal kt38 turn and to calculate the associated free energy change. Coupling of umbrella sampling with replica exchanges between neighboring umbrella sampling intervals could further improve the convergence of the free energy calculations. The simulations revealed a coupled A-minor disruption and global opening of the K-turn motif and allowed to characterize several intermediate A-minor conformations. The calculated free energy profile indicated a meta-stable semi-open structure of slightly higher free energy (~ 1 kcal mol⁻¹) and separated by a small free energy barrier (~ 1.5 kcal mol⁻¹) from the closed (highly kinked) form. Both k-turn states are stabilized by distinct variants of the A-minor interaction. Further opening of the k-turn structure required significantly larger free energy changes. The semi-open form had a reduced kink angle compatible with experimental data on k-turn solution structures and opening was coupled to a continuous global unwinding of the kink-turn motif. The range of free energy changes associated with kt38 opening and unwinding are compatible with the idea that k-turns may trigger biologically relevant motions during large scale ribosome dynamics.

3.2 Introduction

The A-minor interaction and Kink-turn motifs in RNA were introduced in section 1.5.1 and 1.5.2 respectively. In this study we have analyzed both the A-minor fine structure and free energy costs for disruption of the motif and the associated conformational changes in the kt38 structure. Molecular dynamics umbrella sampling calculations of kink-turn opening/closing motion were performed using the disruption/formation of the A-minor interaction as a reaction coordinate. The results indicate a nearly bimodal free energy curve for kt38 elbow motion with a semi-open state of slightly higher free energy compared to the fully kinked state. Several variants of the A-minor motif were observed during the k-turn opening/closing simulations some of which are new or have already been observed experimentally in other k-turn structures. Interestingly, the simulations suggest a coupling of global angular kt38 motion and the relative helical orientation (twist) of the flanking stem regions. A similar effect was characterized recently in gel electrophoresis experiments on a related k-turn motif. More generally, this study confirms a possible role of kink-turn elbow flexibility in coordinating large-scale motion in the ribosome.

3.3 Method

Kt38 RNA oligomer

The starting conformation of the kink-turn kt38 was taken from the X-ray structure of the 50S subunit of *Haloaricula marismortui* (141) (residues 934-942 and 1024-1036 of the 23S ribosomal RNA, pdb-entry: 1JJ2, 141). The kt38 kink-turn motif contained 22 nucleotides from which four nominally unpaired nucleotides form a tightly bent loop (kink element) linking two rigid helical stems, referred to as elements C (canonical base pairs) and NC (non canonical base pairs). The NC-stem comprises two *trans* Hoogsteen SE base pairs, A939/G1031 and A1032/G938, involved in A-minor tertiary interactions between C and NC stems stabilizing the global structure (Fig. 3.1). The first A-minor interaction involves the *trans* SE/SE A939/G1027 base pairing (for the sake of simplicity we call this interaction as A-minor throughout this paper despite that strict definition of an A-minor interaction requires

presence of a complete Watson-Crick base pair as receptor for the adenine). The position of the adenine residue in the NC stem is conserved in all kink-turn motifs reported so far. The second A-minor interaction involves both the *trans* SE/SE A1032/G940 and the *cis* SE/SE A1032/C1026 base pairing (Fig. 3.1).

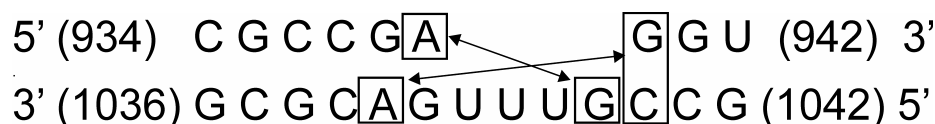


Figure 3.1. Sequence, base pairing and A-minor interactions of the kt38 structure (pdb-entry: 1JJ2, 141). Numbering is given with respect to the 23S rRNA of *Haloaricula marismortui*. The first A-minor interaction is indicated by a short double arrow and the second A-minor interaction is formed by a single adenine interacting with a G:C base pair in the RNA minor groove. The A-minor interactions stabilize a “V” shaped geometry of the C(canonical paired)-stem and the NC(non-canonical paired) stem flanking the K(kink)-element.

Umbrella Sampling approach

The Umbrella coordinate was defined as the distance between the center of mass of the nucleoside A1032 and of the pair of nucleosides G940 and C1026. Umbrella Sampling consists of a series of simulation ‘windows’ for steadily increasing or decreasing values of the reaction coordinate δ . The reaction coordinate was modified in 1 Å steps between adjacent windows and restrained to a window specific reference value using an harmonic biasing potential $v_i(\delta) = k (\delta - \delta_{\text{ref}})^2$ with a force constant $k = 2.0 \text{ kcal mol}^{-1} \text{Å}^{-2}$. As $\delta \sim 6 \text{ Å}$ corresponds to the crystal structure and $\delta \sim 13 \text{ Å}$ to a quasi-linear RNA helix, the range of δ values for the Umbrella Sampling was chosen to be from 6 to 13Å.

Umbrella Sampling was also carried out using the replica-exchange umbrella sampling (REUS, 15,16) method. The latter was used because of a significant hysteresis between free energy pathways obtained by conventional forward and backward umbrella sampling as discussed in the Results section. REUS (15,16) is an enhanced sampling technique which consists of frequent swaps of conformations between adjacent replicas i and j (every pair i, j is tested in turn) according to an exchange probability which satisfies detailed balance. This exchange criterion takes the form of an equation of micro-reversibility and takes into account the probability density of conformations in the two replicas i, j (i.e. Boltzmann factors based on the biased potential energy function of replica i and replica j). The periodic exchanges can

significantly improve the sampling in each window because conformations trapped in a low energy region of the conformational landscape easily diffuse to neighboring replicas which are under the control of a different umbrella potential. The alternative replicas provide an enhanced variety of conformations, accessible to every replica in the limit of a large number of exchanges between the replicas (long-time limit). Details of the Hamiltonian Replica Exchange method and its application for free energy calculations (Replica Exchange Umbrella Sampling) can be found in (15,16,17).

Simulation details

The starting conformation was neutralized by 20 Na⁺ counter ions and solvated with 8408 TIP3P water molecules within a truncated octahedral box using periodic boundary conditions. Long range electrostatic interactions were treated using PME (189) with a 9 Å direct space sum cutoff. Simulations were performed at constant temperature and pressure using the Berendsen algorithm (183). Bond lengths involving hydrogen atoms were constrained using SHAKE (168) and the equations of motion were integrated using the Verlet algorithm with a 2 fs time step. Molecular dynamics (MD) simulations were performed using the *sander* module of Amber 8.0 (203) with the most recent force field developed for nucleic acids (parmbsc0, 19). The solvent and counter ions were first relaxed by energy minimization (2000 steps) and then allowed to equilibrate around a harmonically restrained kt38 structure during three simulations of 0.2 ns at constant pressure and 100 K, 200 K and 300 K simulation temperatures, respectively. The kt38 oligomer was then relaxed by progressively decreasing the harmonic restraints applied to each atom from 15 to 0 kcal mol⁻¹ Å⁻² over a total period of 1.2 ns at constant volume and temperature.

Further equilibration of 0.2 ns was carried out in each window and the structure obtained at the end of this sampling was used as a starting point for the following window. Production runs (Umbrella Sampling and REUS) were then performed in parallel using the final structure of each equilibration step and the same force constant of $k = 2.0 \text{ kcal mol}^{-1} \text{ Å}^{-2}$ in each umbrella sampling window. MD simulations in each independent window were carried out for 2 ns. Both backward (same protocol for equilibration in each window but starting from the final conformation obtained in the forward sampling window at $\delta = 13 \text{ Å}$) and extended simulations (additional 2ns forward per window) were carried-out to better assess the convergence of the sampling. The weighted histogram analysis method (WHAM, 194) was

used to calculate the potential of mean force (PMF) or free energy along the reaction coordinate δ . Helical and global parameters were calculated using Curves (33,34).

3.4 Results

Global flexibility of the kink-turn kt38

Free Energy change of kt38 elbow motion and A-minor disruption

The potential of mean force (PMF) for the kt38 opening was calculated using umbrella sampling by gradually increasing the distance δ between geometric centers of A1032 and the G940:C1026 base pair of the second A-minor interaction (Fig. 3.2). Free energy simulations were performed in both forward (starting from the equilibrated experimental kt38 structure) and as a control also in backward direction (started from the final structure of the forward simulation after re-equilibration). To further improve the sampling quality and convergence the REUS method of using umbrella sampling coupled with replica exchanges between neighboring umbrella sampling windows was also used (backward control REUS simulation restricted to $\delta = 6-9.5$ Å).

In each simulation the calculated global free energy minimum was close to the experimental highly kinked kt38 structure with a reaction coordinate distance δ of 6-6.5 Å. In addition, a second local energy minima at $\delta = 9$ Å was observed that corresponded to a semi-open conformation. Comparing forward and backward umbrella sampling simulations revealed a free energy hysteresis with the calculated free energies shifted to lower values by ~ 1 kcal mol⁻¹ around $\delta = 7-11$ Å in the backward simulation. By allowing exchanges of conformations between the umbrella windows according to the REUS formalism (Fig. 3.2 bold line), the forward and backward PMF curves showed a much smaller hysteresis effect indicating better convergence. Two secondary minima are found from the backward simulation, however when applying REUS on the starting conformations, for each window, obtained after the backward equilibration phase, these two local minima merged into a broad plateau (from $\delta = 7$ to $\delta = 9$ Å). Thus, using the REUS enhanced sampling technique, PMF results from forward and backward simulations are in agreement on predicting a meta-stable conformation (secondary minimum) at $\delta \sim 9$ Å for kt38, characterized by a free energy difference of ~ 1 kcal mol⁻¹ and a small barrier with respect to the closed form of ~ 1.5 kcal mol⁻¹ (Fig. 3.2). The intermediate A-

minor conformations contain water molecules inserted between A1032 and the G:C base pair. Similar A-minor states with inserted water molecules have been described in previous unrestrained k-turn MD simulation studies (136,147).

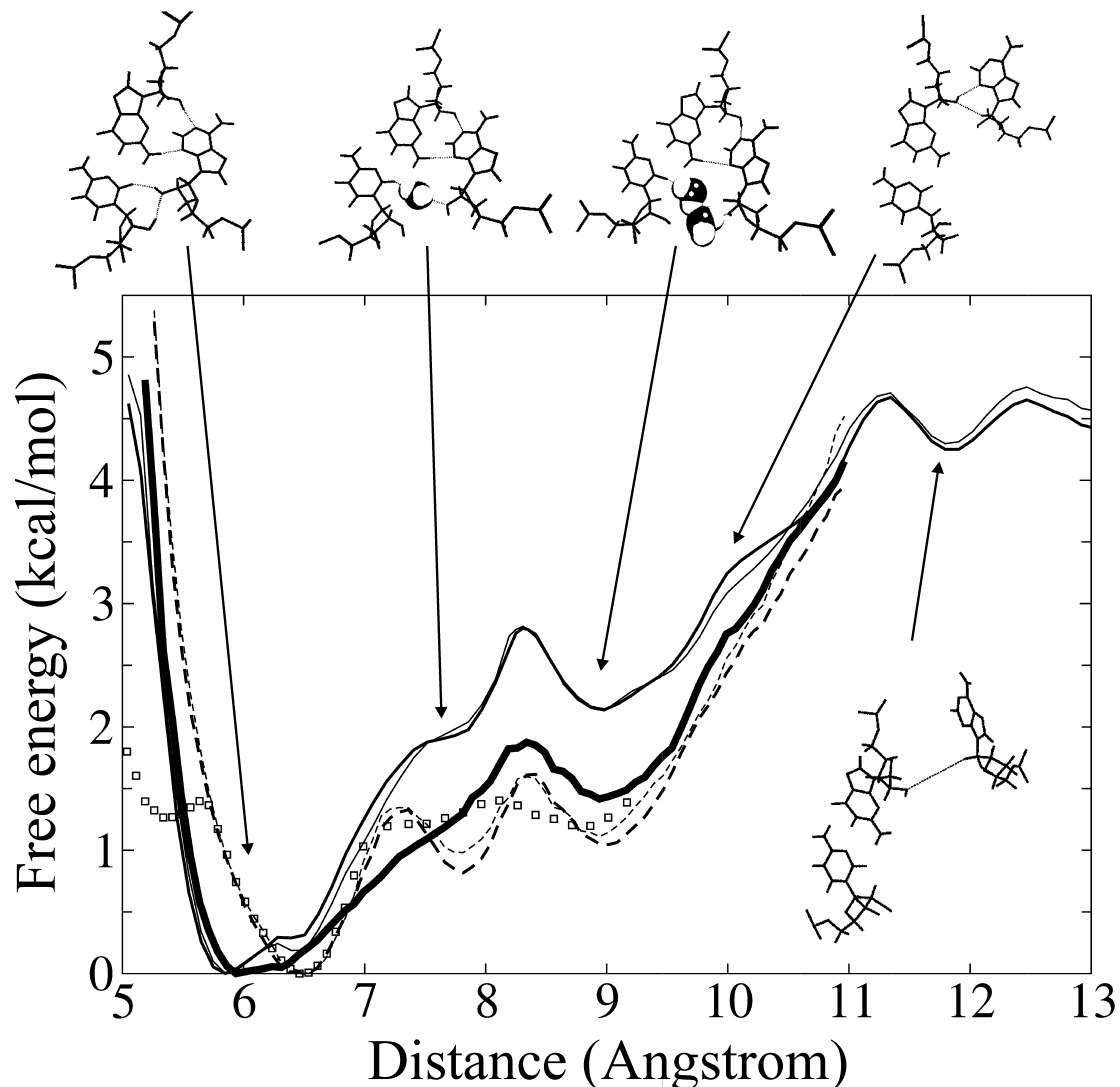


Figure 3.2. Calculated potential of mean force (PMF) for the opening of the kt38 kink-turn motif along the reaction coordinate δ . The reaction coordinate δ corresponded to the distance between the geometric center of the adenine (A1032) and the guanine (G940): cytosine (C1026) base pair of the second A-minor motif. Thin lines correspond to PMFs calculated in the forward (full line) and backward (dashed line) simulations, respectively. Hair lines correspond to a data gathering time of 1 ns per umbrella sampling interval. Thin lines indicate the same simulations but extended to 2 ns per interval. The bold line is the PMF obtained when using the REUS method (Replica Exchange Umbrella Sampling) with starting conformations in each window as obtained after the forward (full line) or backward (open squares) umbrella sampling production runs. Representative simulation

snapshots of the second A-minor motif (G940:C1026 base pair and A1032) for each phase of the umbrella sampling simulation are shown as stick models (inserted water molecules in van der Waals representation).

Further opening of the kt38 structure (i.e. $\delta > 11\text{\AA}$) resulted in a disruption of all stabilizing tertiary interactions (complete loss of A-minor motif) and a free energy level of $\sim 4\text{ kcal mol}^{-1}$ higher than the folded motif. This corresponds to globally almost co-linear arrangements of the kink stem helices. Details of molecular interactions characterizing the successive kt38 substates are described in the following sections.

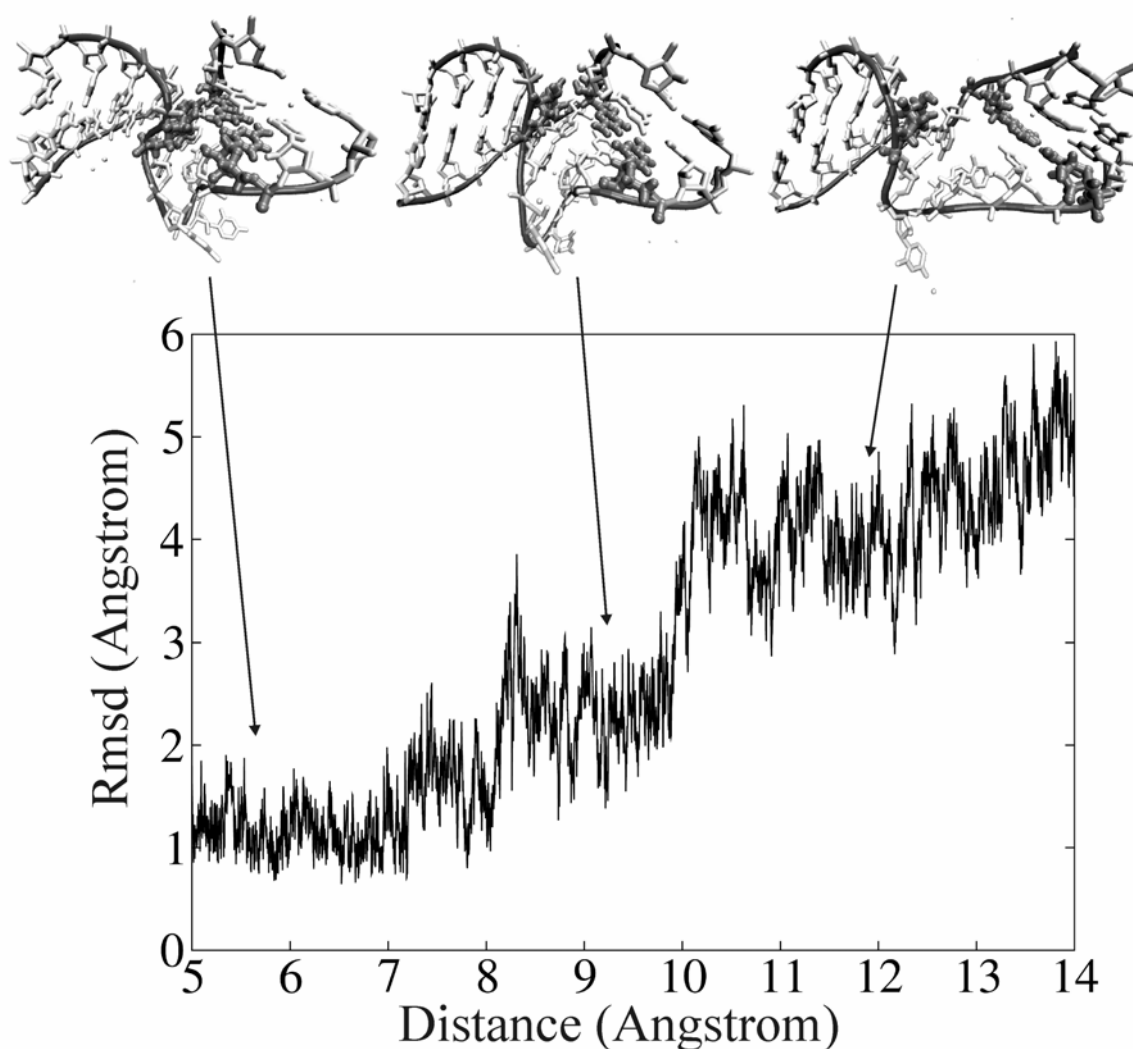


Figure 3.3. Root mean square deviation (based on the nucleic acid backbone heavy atoms) with respect to the equilibrated starting structure ($\delta = 6\text{ \AA}$) during umbrella sampling of the kt38 opening. Representative conformational snapshots for each simulation stage are indicated (arrows). Nucleotides

belonging to the second A-minor interaction are shown as bold stick model (dark tube representation of nucleotide backbone).

Global deformability of kt38

The phosphodiester backbone Rmsd (Fig. 3.3) based on heavy atoms with respect to the initial conformation revealed consecutive shifts in the range 1-5 Å during the umbrella sampling simulations. One can distinguish four different Rmsd levels corresponding to $\delta < 7$ Å (global minima), $7 \text{ Å} < \delta < 8 \text{ Å}$ (saddle region), $8 \text{ Å} < \delta < 10 \text{ Å}$ (local minimum, semi-open state) and $\delta > 10 \text{ Å}$ (plateau regime, fully open state). These states closely resemble the different A-minor states identified in the previous paragraph. The amplitude of Rmsd fluctuations increased from one regime to the next. Rmsd fluctuations were $> 1 \text{ Å}$ on the plateau, which indicates a significantly increased flexibility of the open kt38 structures compared to the closed highly kinked form.

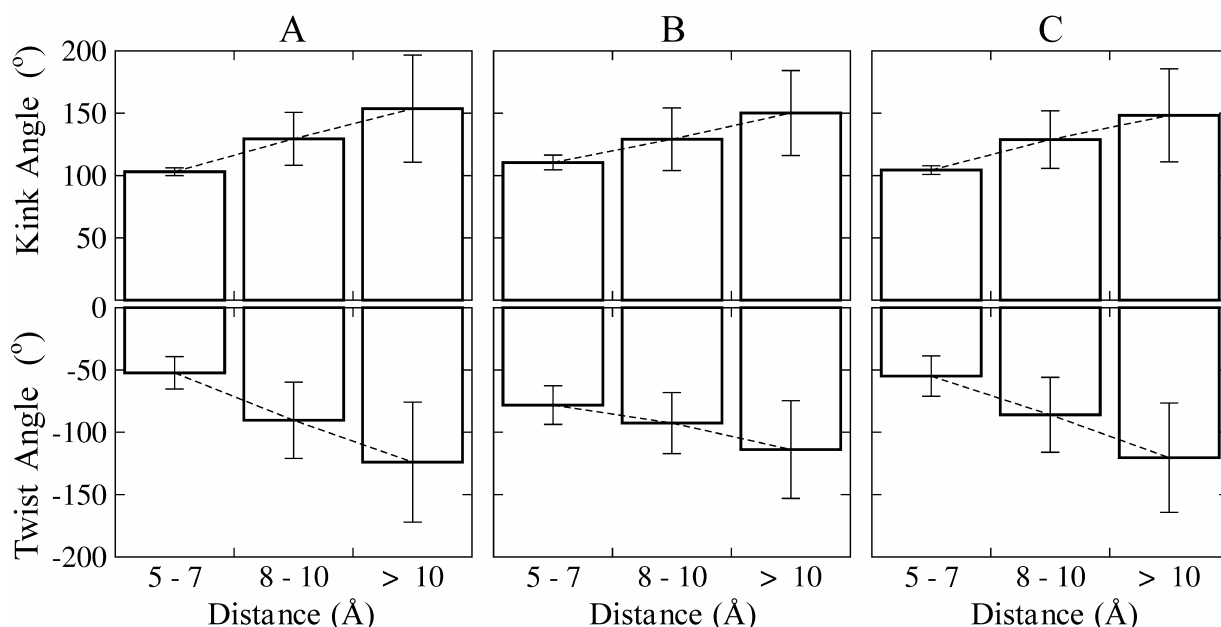


Figure 3.4. Kink angle and Twist angle between C and NC stems averaged in successive block of three sampling windows. A: Forward Umbrella Sampling simulation. B: Backward Umbrella Sampling simulation. C: Replica Exchange Umbrella Sampling simulation (starting from final conformations of the forward Umbrella Sampling).

Global kink and twist deformation associated with different regimes of the PMF curves are illustrated in Fig. 3.4. The kink angle between the two flanking stems remained close to 100° for the closed conformation (see error bars), while the semi-open geometry corresponded to

an average value of 125° . The average inter-stem kink angle in the fully open geometry was 150° . For the semi-open and open conformations, much higher fluctuations of the kink angle were observed with standard deviations of approximately 25° and 50° , respectively. Very good agreement was observed between averages obtained from forward, backward and REUS production sampling for each regime.

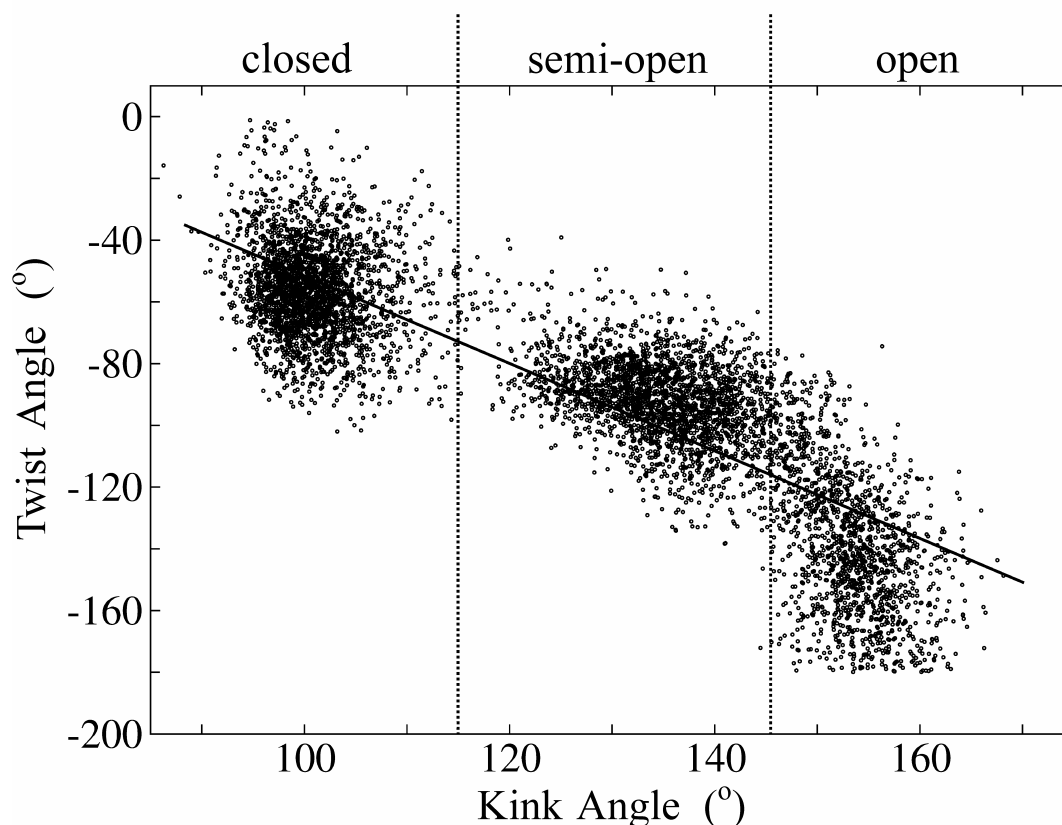


Figure 3.5. Correlation of the inter-stem kink and twist angles. Data points were recorded every 2 ps during umbrella sampling simulations. The kink/twist correlation is also indicated by least-square fitted line.

The simulation studies predict that the opening motion of kt38 is correlated with a twist rotation of the two flanking stems (Fig. 3.4 lower panels and Fig. 3.5). The calculated average inter-stem twist angle is -50° for the closed conformation, -75° for the semi-open conformation and $\sim -100^\circ$ for the open conformation in kt38. The correlation coefficient with global kink angle extracted from Fig. 3.5 is -0.87. Note, that larger global twist fluctuations were also observed for more open k-turn structures. This increase was, however, less pronounced than for the kink angle fluctuations. Interestingly, recent gel electrophoresis studies of kt42 motif in RNA constructs that contained a second kinked RNA motif indicated also a significant twist change upon k-turn formation (149). The kt42 motif is similar to the

kt38 element and undergoes a metal ion induced transition between an open and closed form with an associated change in bending direction of $\sim 75^\circ$ (149) in qualitative agreement with the above calculated twist change upon transition from the closed to the semi-open kink turn state.

Dynamics of the second tertiary A-minor interaction reveals three local substates engineering the hinge like motion of kt38: type I, type 0 and disrupted.

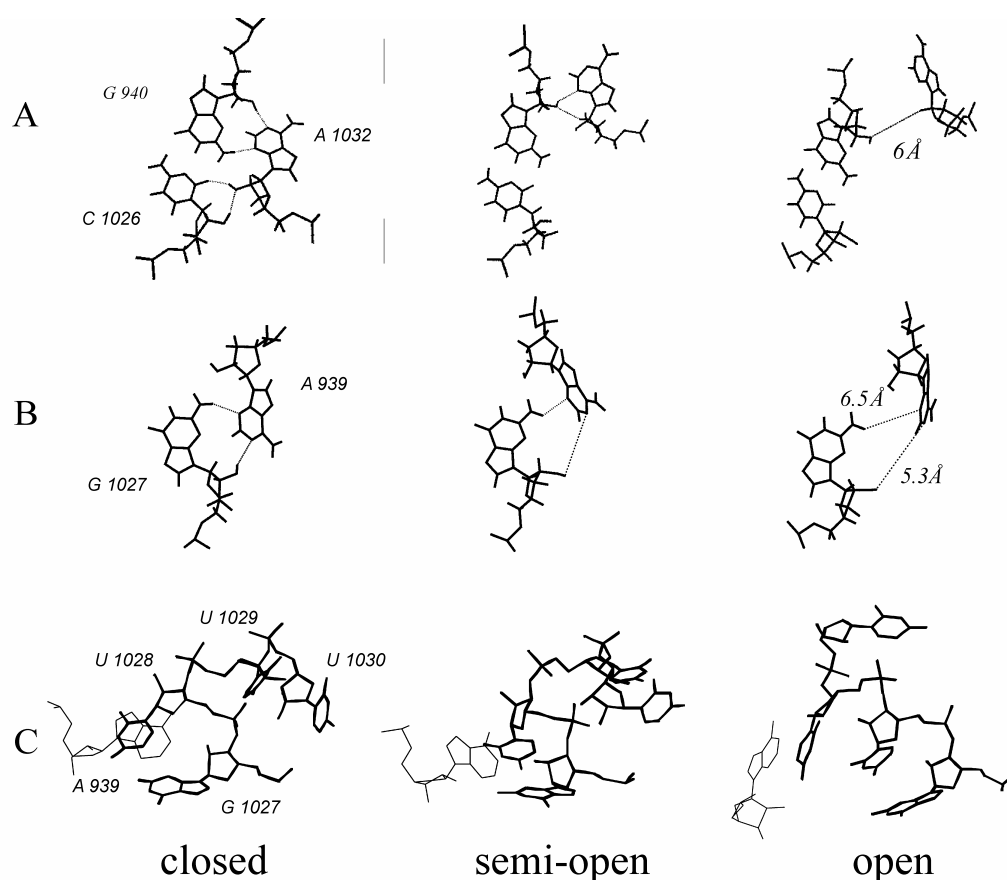


Figure 3.6. Structural motifs characterizing the closed, semi-open and open conformations of the kt38 kink-turn structure. Some key hydrogen bond donor-acceptor distances are included as dotted lines. (A) Second A-minor interaction of kt38, from left to right: A-minor type I cis SE/SE A/C, A-minor type 0 and disruption of the second A-minor motif. (B) Representative snapshots of the first A-minor interaction during kt38 opening (C) Central loop (element K) between C and NC stems (A939 is shown as thin line). From left to right: U1028 stacks on A939, U1028 alternatively stacks on A939 and U1027 and in the open conformation: U1029 inserts into the central cavity inducing stable stacking between A939, U1029, A1028 and G1027, respectively.

The global dynamics of kt38 depended on the local variability of its second A-minor interaction linking the C and NC stems (136). Upon k-turn opening the A-minor motif undergoes transition to a number of conformational variants (Fig. 3.2) including conformations that involve one or two water molecules partially filling the space between A1032 and the G940/C1026 base pair. The type I A-minor interaction in the kinked structure included the trans SE/SE A1032/G940 and the cis SE/SE A1032/C1026 pairing in accordance with the classification of Nissen et al. (133). In the closed state of kt38 in current free energy simulations ($\delta = 6-6.5$ Å), both the N3 and O2' atoms of A1032 are inside the minor groove of the receptor helix formed by the G940:C1026 base pair (Fig. 3.6a, left panel), and thus correspond to a type I A-minor interaction. Both interactions between A1032 (N3, N5) and G940 (O4, O2' respectively) and between A1032 (O2') and C1026 (O2', O2 simultaneously) are hydrogen bonded.

At the saddle point ($\delta = 8$ Å, $\Delta G = 1.5$ kcal mol⁻¹), opening the A-minor motif induces an increase of the distance between A1032 and C1026, i.e. these two nucleotides loose their direct hydrogen bonds. The disruption of the A-minor interaction does not occur simultaneously for all hydrogen bonds but in a stepwise fashion involving a translational shift but also a rotation of the A1032 relative to the G940:C1026 base pair. The free energy saddle is indeed associated with a second A-minor type I water inserted *cis* SE/SE A/C motif, observed also in unrestrained MD simulations (136,147). We observed both the insertion of one out of plane water molecule mediating hydrogen bonds between A(O2') and C(O2) and insertion of two water molecules mediating hydrogen bonds between A(O2') and both C(O2)/C(O2') (corresponding to semi-open interaction in 136). We observed multiple water molecule exchange events at this site and very frequently an absence of such water molecule mediating H-bonds, i.e. no interaction between A1032 and C1026. An interpretation is that the loss of entropy associated with water inserted A-minor type I variant is relatively high and the reason for the peak in the free energy surface (~ 1.5 kcal mol⁻¹) corresponding to the semi-closed state.

In the semi-open conformation of kt38 (local energy minima at $\delta = 9$ Å) the distance between A1032 and G940/C1026 base pairs further increases. This induces a transition of the second A-minor motif toward a type 0 illustrated in Fig.3.6 (upper middle panel). The type 0 is characterized by a lateral shift of the A1032 with respect to the G/C receptor minor groove: the N3 of A1032 is located outside of the G940:C1026 minor groove but can still form a hydrogen bond with the ribose of G940. The local lateral shift when going to a type 0 variant in the second A-minor interaction is a major reason for the global untwisting of the structure.

During the simulations this tertiary interaction is disrupted in the fully open substate (onset of plateau, Fig. 3.6 upper-right) in agreement with conformational search results reported in (147). The first A-minor interaction in the kt38 motif involves a A939:G1027 base pairing interaction. In contrast to the second A-minor motif this interaction was not completely disrupted throughout the induced kt38 opening (Fig.3.6b). A significant change was a progressive inclination of A939 with respect to G1027 upon kt38 opening (Fig. 3.7).

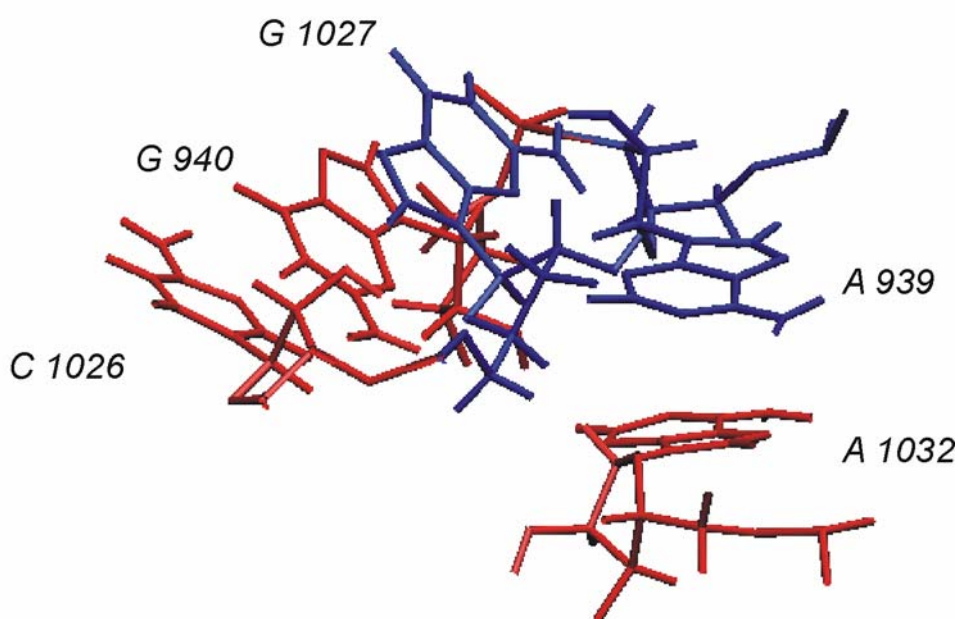


Figure 3.7. The two A-minor interactions, i.e. involving A939 (first) and A1032 (second), which stabilize kt38 in the semi-open conformation ($\delta = 9\text{\AA}$). First A-minor interaction (blue bond-sticks) and second A-minor interaction in the type 0 (red sticks) are “interlocked” by *trans* SE/SE hydrogen bonds between A939 (in blue) and G940 (in red).

Rearrangement of nominally unpaired nucleotides

The long strand of the inter-stem loop (termed element K) contains four nominally unpaired nucleotides (G1027, U1028, U1029 and U1030). G1027 has been described as part of the first A-minor interaction. The aromatic base of U1028 remained stacked inside the motif during all stages of the simulations. In the closed state U1028 is approximately in the plane of G1027 with which it forms a side-by-side *cis* Hoogsteen SE pair (Fig.3.6c left panel). In the semi-open state U1028 stacked preferentially on G1027 (Fig. 3.6c middle) and this was favored by the lateral shift of A939 (NC stem) with respect to the C stem.

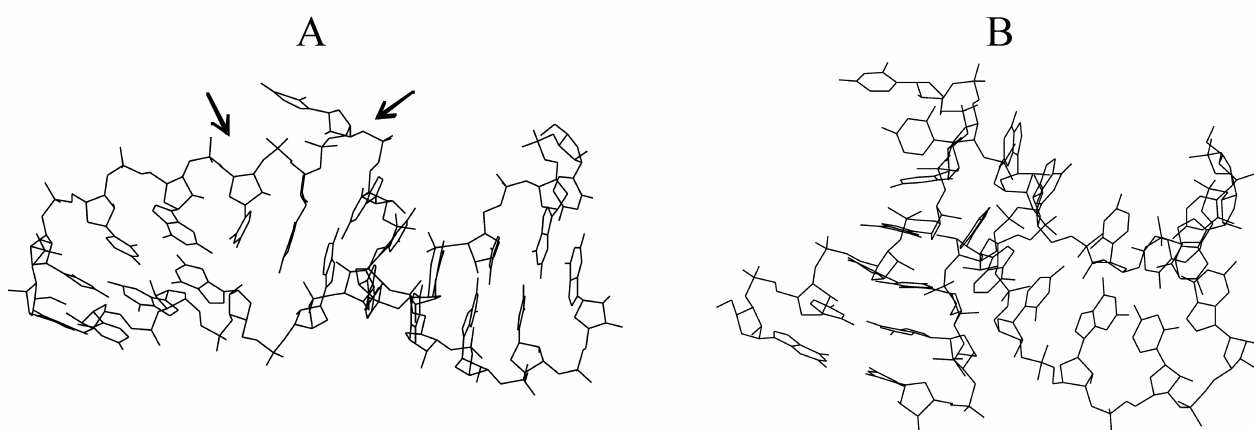


Figure 3.8. Final structure ($\delta = 13 \text{ \AA}$, A) corresponding to a fully open geometry of kt38 compared to the initial start structure ($\delta = 6 \text{ \AA}$, B). The central loop (element K) in the open state contains a single bulged-out nucleotide (U1030). The two nucleotides with largest associated changes in backbone structure upon k-turn opening are indicated by arrows (in A).

Global opening of the kt38 motif induced the extension of a free space below the phosphodiester backbone containing unpaired and bulged-out nucleotides (i.e. K element, Fig. 3.1). This cavity was absent in the closed conformation because of the sharp kink in the element K. During the smooth elbow dynamics wherein $\delta < 11 \text{ \AA}$ the cavity was accommodated by a lateral shift between C and NC stems (see above). At $\delta = 11 \text{ \AA}$ the larger cavity was *filled* by either solvent molecules or partial stacking of U1028 between G1027 and A939 (Fig. 3.6c, middle). For $\delta > 11 \text{ \AA}$ the U1029 base inserted into the cavity and stabilized the stacking of U1028 on top of G1027 (Fig. 3.6c, right panel). U1030 remained looped-out in all simulations. In the most extended form of the kt38 ($\delta = 13 \text{ \AA}$) most of the bulge loop bases except U1030 remained in an intra-helical stacked conformation (Fig. 3.8). It is possible that further opening could result in a looped out conformation for all bulge bases, however, the current reaction coordinate may not be appropriate to achieve a complete co-linear arrangement of the stem helices. Most of the backbone conformational changes that accompanied the transition of kt38 to a fully open form was localized to the U1027 and U1030 bulge nucleotides (Fig. 3.8).

3.5 Discussion

Umbrella sampling simulations on the kt38 using the opening (or closing) of the key A-minor interaction as a reaction coordinate were used to study the free energy change associated with

k-turn opening and to characterize associated structural changes. In case of kt38, the elbow motion involved free energy changes of $\sim 1.5 \text{ kcal mol}^{-1}$ in favor of the closed form involving only a small energy barrier. The relatively small calculated free energy change ($\sim 3 \text{ RT}$, R: gas constant, T: temperature) may trigger biologically relevant motions during large scale ribosome dynamics and protein synthesis. It was demonstrated that coupling of umbrella sampling simulations with a Hamiltonian replica exchange methodology significantly improved the convergence of calculated free energies and resulted in a closer match of forward and backward simulations compared to “standard” umbrella sampling.

The simulations revealed a dynamic equilibrium between type I and several different type I A-minor subtypes with one or more inserted water molecules characterizing the “saddle point” regime of sampled conformations. The small calculated free energy difference is consistent with the frequent sampling of k-turn states with inserted water molecules during unrestraint MD simulations (147). The meta-stable semi-open conformation required a shift of the second A-minor motif into the type 0 form.

In addition to the local conformational changes, the simulations indicated a global change in bending and twist of the C stem with respect to the NC stem correlated with the opening of the kt38 structure and controlled by the conformation of the second A-minor motif. The molecular origin of the bending and twist motion is at least in part due to the shift induced locally in the second A-minor motif from type I to type 0. It has been reported for the related kt42 element that in solution (in the absence of Mg^{2+} ions) the motif adopts an open conformation with a bend angle representing that of a 3 or 4-base bulge structure (149). A 3-base bulge in RNA adopts a bending angle of $\sim 130^\circ$ (204, in the current nomenclature where 180° corresponds to no bending). Such global geometry closely matches what has been defined as semi-open in the current study whereas the fully open form that requires higher free energy changes ($\sim 4 \text{ kcal mol}^{-1}$) approaches already a global geometry with a smaller residual overall bending. This latter form is characterized by a complete disruption of the second A-minor motif.

In our simulation study the semi-open state is predicted to be of slightly higher free energy than the closed highly-kinked state. In contrast, experiments on a related kt42 motif indicate that in free solution the kt42 prefers a form that matches our proposed semi-open structure (149). It is important to keep in mind that the experimental system differs from the simulation system in that it contained longer negatively charged RNA stem regions. Such negatively charged stem regions may add a significant electrostatic repulsion of the stem helices. This electrostatic contribution can shift the free energy profile such that the semi-open form is

more favorable at low salt concentrations (because of the larger distance and smaller repulsion of the stem regions). In turn, at high salt concentration or in the presence of divalent cations which are known to very efficiently neutralize the negative phosphate charges in RNA the highly kinked k-turn form (closed form) becomes the most stable form. Hence, the present simulation study resembles more experimental conditions at which the RNA stem charges are completely neutralized (absent in the simulations). The relatively small calculated free energy changes for k-turn opening/closing support the idea that k-turn motifs can act as flexible elements to mediate global changes of helical stem elements in large RNA containing molecules possibly triggered by ion or protein binding.

Acknowledgements

This work was performed using the computational resources of the CLAMV (Computer Laboratories for Animation, Modeling and Visualization) at Jacobs University Bremen and supercomputer resources of the EMSL (Environmental Molecular Science Laboratories) at the PNNL (Pacific Northwest National Laboratories). J.C. is supported by a grant (I/80485) from the VolkswagenStiftung to M.Z. and by a grant from the Universite Franco-Allemande (UFA). JS acknowledges support by grants GA203/09/1476, Grant Agency of the Czech Republic, IAA400040802 by The Grant Agency of the Academy of Sciences of the Czech Republic and grants grants no. AV0Z50040507 and AV0Z50040702 by Academy of Sciences of the Czech Republic.

Chapter IV

DNA bending, curvature and the spontaneous occurrence of a Kink between two base pairs

- . **Keywords:** DNA fine structure; indirect readouts; damaged DNA; A-tracts; DNA bending free energy; base pair kink motifs; DNA bending stiffness.
- . **Reference publications:** Curuksu, J., Zakrzewska, K., Zacharias, M. (2008) *Nucleic Acid Research*. **36**: 2268-83; Curuksu, J., Zakrzewska, K., Lavery, R., Zacharias, M. (2009) *Nucleic Acid Research* (submitted).
- . **Recommended introductions:** 1.3, 1.4, 2.2, 2.8 and 2.9.

4.1 Abstract

DNA bending flexibility is central for its many biological functions. A new bending restraining method for use in molecular mechanics calculations and molecular dynamics simulations was developed. It is based on an average screw rotation axis definition for DNA segments and allows inducing continuous and smooth bending deformations of a DNA oligonucleotide. In addition to controlling the magnitude of induced bending it is also possible to control the bending direction so that the calculation of a complete (2-dimensional) directional DNA bending map is now possible. The method was applied to several DNA oligonucleotides including A(adenine)-tract containing sequences known to form stable bent structures and to DNA containing mismatches or an abasic site. In case of G:A and C:C mismatches a greater variety of conformations bent in various directions compared to regular B-DNA was found. We then use this approach to calculate the free energy of DNA bending in function of its base sequence by MD simulations with the parmbsc0 force field and WHAM algorithm. Because this project was initiated in the context of a large scale effort to characterize the sequence dependence of DNA flexibility (“ABC” initiative, 51, 52) and before the apparition of the parmbsc0 force field (see 2.2), an independent study using a

previous Amber force field series, parm94 was carried-out. These results were compared with the adiabatic mapping EM study in details.

With parmbsc0, the four oligomers studied are alternating [AT] and [CG] and two oligomers with A-tract, differently positioned, [Atract-1] with its central position and [Atract-2] for the off centre A-tract. Bending up to 90° requires a total free energy change of 5., 7., 4.5 and 5.5 kcal/mol for the [GC], [AT] [Atract-1] and [Atract-2], respectively. For higher bending, up to 150°, we obtain the respective values of 12.5, 12., 10. and 13.5 kcal/mol. We find also, in agreement with experimental data that for higher bending the free energy increases linearly with bending angle. All four oligomers bend rather smoothly without sharp kinks but, with parmbsc0 important differences are observed in the regime of strong bending. [GC] and [Atract-1] form type II kink, [AT] dissociates and [Atract-2] forms a series of mini kinks. The flexibility of individual base pair junctions is also analysed.

4.2 Introduction

The curvature and bending flexibility of DNA is important for its many biological functions including recognition by proteins, DNA repair, packaging and transient melting during transcription and replication (23,67,76,77,114,205-208). A-tracts (see 1.4.2) and other sequences can adopt intrinsically curved structures or may also exhibit a higher bending flexibility of DNA (4,107,108,109,112). Furthermore, the bending deformability of mismatches and abasic sites may play a decisive role for distinguishing damaged and undamaged DNA during DNA repair processes. In addition to biophysical and structural studies, molecular modeling and simulation methods have been used to explain sequence-dependent curvature of DNA. These include unrestrained molecular dynamics (MD) simulations of A-tract containing DNA (209-214). Studies based on unrestrained MD simulations suffer, however, from two drawbacks: Firstly, currently accessible timescales may not be sufficient to sample all relevant states compatible with bent DNA structures even around the equilibrium state. Secondly, many interesting DNA bending deformations seen, for example, in protein-DNA complexes go beyond equilibrium fluctuations of isolated DNA. Hence, it is unlikely to observe and characterize such deformations using conventional MD simulations. One possibility to study DNA bending deformations systematically is to employ energy minimization including constraints to bend DNA (215-216). For example Sanghani et

al. (216) used the JUMNA (Junction Minimization of Nucleic Acids, 22) program to compare induced DNA bending on the Hagerman sequences $(dA_4T_4CG)_n$ and $(dT_4A_4CG)_n$. The authors defined a curved superhelical axis that constrained the DNA to follow a superhelical pathway of defined radius of curvature. Constrained energy minimization using different radii of curvature for the superhelical path was used to calculate the bending properties of the Hagerman sequences (216). A disadvantage of such approach is that the constraint applied to the complete DNA segment may restrict the possibilities of the structure to relax the bending induced strain.

In order to systematically study the microscopic effects giving rise to DNA curvature and the associated energetic costs, we have developed a new restraint to induce continuous bending deformations of DNA. The method is based on a screw axis of double helical segments and a restraint for the angle between them. We implemented this restraint in the JUMNA program for bending a DNA fragment during energy minimization. The bending angle variable has been defined in such a way that all DNA conformational parameters can distribute optimally according to the specific dependence on the sequence without disturbing the helix locally. An additional advantage of the method is the possibility to restrain also the bending direction. This allows a systematic mapping of the global bendability of a given DNA sequence as a function of every bending direction and representing it on a polar plot. The approach has been applied to several DNA molecules including examples of A-tract, mismatch and abasic site containing sequences. The results are in good agreement with experiment data for every Atract-like sequences investigated. However the calculations indicate a significantly altered bending deformability with disparate stable states at abasic sites and some mismatches compared to regular DNA that could play a role during DNA damage recognition. A comparison with a MD version of the approach (parm94 force field, 217) indicate a softer penalty for bending but good qualitative agreement with adiabatic mapping on the mechanism of A-tract DNA curvature. The coupling of DNA bending to other global variables such as twisting, stretching and groove width was also investigated.

The second project carried-out with the generalized DNA bending coordinate focuses on DNA bending propensities far beyond equilibrium using Umbrella Sampling MD. DNA molecules can undergo such strong bending in many protein/DNA complexes (23,76,89,114,207,208,218,219), in looped DNA (90,92,220,221) and nucleosomal complexes (52,55,56,60,61), see the general introduction sections 1.4.1 and 1.4.2. Recent cyclisation experiments of short DNA fragments indicated that significantly stronger bending or kinking than suggested by a simple elastic rod model of DNA can occur spontaneously in DNA (12).

Other experimental techniques based on molecular force sensors (5), fluorescence energy transfer (FRET, 6), or atomic force microscopy (AFM, 4) also suggested that strong bending of DNA is easier than expected and theoretical models have been developed that attempted to reproduce this property (71-73). At the same time, the molecular mechanism of DNA bending is of great interest, in particular the question if strong bending results in sharp kinks or rather involves a smooth and distributed deformation of DNA.

Sharp kinking of DNA was proposed by Klug, A. and Crick, F. (125) already in the 1970th. Smooth helix bending has also been proposed: 45° opening of the minor groove, (69) and 22.5° in both grooves alternating with the helix phase, (70). Other propositions for enhanced DNA bending propensity (222,223) considered base flip-out or formation of local bubbles (224,225). The Klug and Crick type of kink was found in several recent simulations (108,126) and was called type I kink. It is characterised by a high roll (of the order of 90°) at a particular junction leading to an unstacking of just one base pair, with relatively low disturbance of the neighbourhood. In addition, the type II kink has been described as due to disruption of hydrogen bonds of the central base pair and stacking of these bases on their 5' neighbours. This second motif reveals very large propeller (roughly 120°) and stagger but intact flanking base pairs.

Bent and kinked DNA molecules correspond to non-equilibrium conformations of DNA that may occur only transiently and are therefore difficult to study experimentally. Our Umbrella Sampling Molecular dynamics simulations are in principle well suited to study bent and kinked DNA at high spatial and time resolution. This project has two objectives, firstly, we use our restraining approach to obtain the bending free energy of short DNA fragments as a function of base sequence and secondly, we aim to characterize DNA bending dynamics at the base pair (bp) level. The four oligomers studied are alternating [AT] and [CG] and two oligomers with A-tract, differently positioned, [Atract-1] with its central position and [Atract-2] for the off centre A-tract. Interestingly, in the regime of weak bending (up to 50°) the bending free energy follows closely a quadratic curve consistent with the experimentally measured DNA persistence length. However, for larger bend angles the slope of the free energy increase becomes smaller consistent with recent experimental studies using AFM experiments (4). This regime of bending is accompanied by bending induced kinks in the DNA (strong bending localized to individual base pair steps). Our findings may have important implications for understanding the mechanics of global conformational deformability of DNA molecules. For this second project using the parmbse0 force field, a full analyse of the four trajectories is given in Annexe I.

4.3 Method

Geometric variable associated with DNA bending

The definition of a bending angle for a double stranded oligomer is based on the single parametric description introduced in section 1.3.2 (model II). The general framework is the theory of screw motion derived from the theorem of Chasles and applied to helical transformation independently on each strand of DNA. The bending parameter will be defined as an angle between two screw axes of the adjacent few base pairs (referred to as handles below). A restraining potential of the form $v(\theta) = k.(\theta - \theta_{\text{ref}})^2$ is then added to the force field where θ is an effective angle between two vectors corresponding to the two handles.

Screw axis bending variable

The orientation of one nucleotide is defined by a reference axis system located on the base with unit vectors (e_1, e_2, e_3) where e_1 is the bond N9-C8 for purines or N1-C6 for pyrimidines and thus approximately parallel to the direction of the central dyad axis of the double helix. The e_2 vector is defined by the cross product between e_1 and the glycosidic bond C1'-N (opposite sign for the complementary strand) and thus perpendicular to the base plane, and e_3 is the normal with respect to e_1 and e_2 and points approximately in the direction of the long base-pair axis. Space-invariant rotation vectors between adjacent base reference axis systems (Fig.4.1a, 47-49, 226) can be obtained from the "rotation formula" (41,227). Eq. 4.1 expresses the rotation matrix as a function of the rotation angle θ and the rotation vector $\vec{u}$ (u_x, u_y, u_z):

$$q = I + \sin \theta \cdot S(\vec{u}) + (1 - \cos \theta) \cdot S^2(\vec{u}) \quad \text{eq. 4.1}$$

I is the identity matrix and $S(\vec{u})$ the anti-symmetric matrix associated with vector $\vec{u}$. The matrix expression in eq. 4.1 can be used to define the vector $\vec{u}$ as a function of elements of the rotation matrix q in the following way (47):

$$u_x = \frac{(q_{32} - q_{23})}{\varepsilon}, \quad u_y = \frac{(q_{13} - q_{31})}{\varepsilon} \quad \text{and} \quad u_z = \frac{(q_{21} - q_{12})}{\varepsilon} \quad \text{eq. 4.2}$$

ε is a constant ($= 2.\sin\theta$) and can thus be set as $\|\vec{u}\|$. Since the coordinates of the base-reference axis system are known, the rotation matrix q can be obtained by:

$$q = B' \times B^{-1} = B' \times B^T \quad \text{eq. 4.3}$$

Hence, eq. 4.2 can now be solved. As illustrated in Figure 4.1b, space-fixed internucleotide rotation vectors are calculated for several neighbouring di-nucleotide steps inside a fragment of n base pairs and their sum defines a vector “handle” (analogue to a local axoid in ref. 49) oriented as a function of $(2n-2)$ specific inter-nucleotide single rotations.

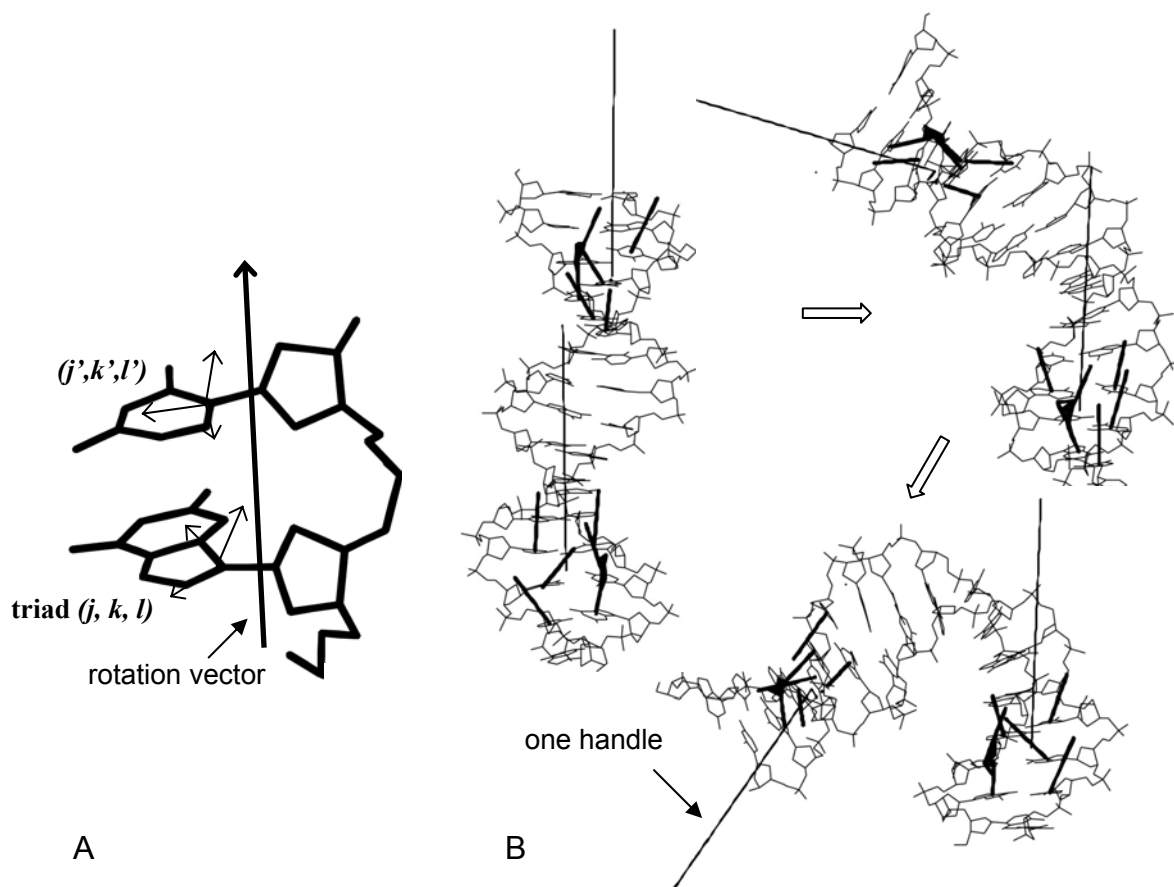


Figure 4.1. Definition of bending angle restraining coordinates for adiabatic mapping and umbrella sampling calculations. (A) Illustration of the local base associated coordinate systems (three orthogonal vectors at each base) of two consecutive nucleotides and the associated rotation vector (each dinucleotide step defines one rotation vector). (B) Definition of the bending angle: Each handle vector (long thin sticks) is the sum of six rotation vectors (short sticks) for 4 bp terminal fragments. The global bend angle is given by the angle between the two handles.

Typically, the 4 terminal base pairs at each end of a DNA molecule were included to define two (terminal) handles. The global bend angle is given by the angle between the two handle vectors.

Directional bending variable

A specific direction of bending can be imposed by the following construction. The two handles are kept perpendicular to a vector $\vec{r}$, obtained by rotating the long axis of a chosen base pair (we take the vector between the C1'-atoms of the two complementary bases) by the angle α around the average vector of the two handles $\vec{m}$. A direction $\alpha=0$ corresponds to bending in the direction perpendicular to the reference base pair C1'-C1' direction (approximately in the directions of the base pair dyad axis), pointing towards the minor groove. An angle $\alpha = 90^\circ$ corresponds to bending toward the phosphates of the base pair, in the direction of the leading 5'-3' strand.

During directional bending two quadratic restraining potentials of the form $V(\Omega_j) = k(\Omega_j - 90)^2$ keep the two handles perpendicular to $\vec{r}$. The angle Ω_j between handle H_j and reference vector $\vec{r}$ obtained from the scalar product has no sign. To control the unidirectional nature of bending inside the plane normal to vector $\vec{r}$, one can restrain the 'rotation' of a handle with respect to the other around a vector $\vec{r}$, which means using eq. 4.1 to restrain the sine of the bending angle θ . The bending angle is thus $\theta_i = \tan^{-1}(\sin\theta_i / \cos\theta_i)$ and for computational simplicity when calculating analytic derivative of θ , independent harmonic biasing potential are used to restrain $\cos\theta_i$ and $\sin\theta_i$.

$$\cos\theta = \vec{H}_1 \cdot \vec{H}_2 \quad \text{scalar product} \quad \text{eq. 4.4}$$

$$\sin\theta = \frac{1}{\vec{r} \cdot \vec{H}_1} [\vec{H}_2 - \vec{H}_1 - (1 - \cos\theta) \cdot [\vec{r} \cdot (\vec{r} \cdot \vec{H}_1)]] \quad \text{Rotation formula} \quad \text{eq. 4.5}$$

Energy minimization and adiabatic mapping along the bending coordinate

The bending angle restraint was implemented in the program JUMNA (22). JUMNA employs a combination of helicoidal and internal variables to describe and energy minimize the structure of DNA oligonucleotides. In the current calculations the Amber Parm98 force field

(228) and a Generalized Born (GB) model based on a pair-wise descreening approximation was used to implicitly account for solvent effects (170,229). Bending energy curves were generated by a series of simulation “windows” for stepwise increasing values of the bending restraining angle θ_{ref} (5° -steps; force constant: $k=400 \text{ kcal mol}^{-1} \text{ rad}^{-2}$) in the harmonic restraining energy term. Backward simulations starting from the final structure of the last windows were also performed in each case. Two-dimensional (2D)-bendability maps that included the bending direction were obtained by a series of bending minimizations for each possible bending direction. The bending direction, α , was changed in 20° steps from 0° to 360° . Structures along the DNA bending pathways were analysed in terms of local helical parameters using the Curves program (33, 34).

Energy minimization studies were performed on several AT-rich doubled stranded (ds)DNA dodecamers. This included a dodecamer with a central A-tract $(5'\text{-dGGCA}_6\text{CGG})_2$ (230), an alternating pyrimidine-purine sequence $(5'\text{-d(AT)}_6)_2$ and a pair of sequences, $\text{d}(5'\text{-CGA}_4\text{T}_4\text{CG})_2$ and $(5'\text{-dCGT}_4\text{A}_4\text{CG})_2$ studied experimentally by Hagerman (97) and referred to as Hagerman H1/H2 sequences (98), respectively. In addition, the recognition sequences of the human and bovine papillomaviruses E2 proteins, $(5'\text{-dACCGAATTCGGT})_2$: HPV sequence) and $(5'\text{-dACCGACGTCGGT})_2$: BPV sequence), respectively, were also studied (referred to as HPV/BPV sequences, 231-234). In addition to regular dsDNA, the approach was applied to modified or damaged DNA oligonucleotides with sequences $5'\text{-dCGTAC-CATGC}/5'\text{-dGCATGAGTACG}$ (central abasic site, 235), $5'\text{-dGCTTCAGTCGT}/5'\text{-dACGACGGAAGC}$ (central G:A mismatch, 236), $5'\text{-dGCCACCAGCTC}/5'\text{-dGAGCTCGTGGC}$ (central C:C mismatch, 237) and $5'\text{-dCCATGCGTGG}/5'\text{-dCCATGCGTGG}$ (tandem G:T mismatches, 238). For all these sequences experimental structures are available. However, to allow for an unbiased comparison of experiment and calculation all calculations (on all DNA molecules) were started from canonical B-DNA structures.

Molecular Dynamics Umbrella Sampling

MD simulations were first performed with the parm94 Amber force field (217) series on an oligonucleotide with a central A-tract $(5'\text{-dCGCGCA}_5\text{CGCGC})_2$ as explained in the introduction. The standard B-DNA starting structures was generated using the nucgen program of the Amber8 package (203). The molecule was neutralized by 16 K^+ counterions and solvated with 3519 TIP3P water molecules (191), corresponding to a solvent layer ≥ 10

Å, within a truncated octahedral box which had a face to face dimension of 63 Å. Simulations were performed at constant temperature and pressure applying periodic boundary conditions and the particle-mesh Ewald approach (188) with a 9 Å direct space sum cutoff. Bond lengths involving hydrogen atoms were constrained using Shake (168); the equations of motion were integrated using the Verlet algorithm and a 2 fs time step.

After an initial stage of energy minimization, the solvent and counterions were allowed to equilibrate during 0.6 ns at constant volume while progressively increasing the temperature up to 300 K and restraining the solute to the start structure (standard B-DNA). The solute was then partially relaxed by progressively decreasing the harmonic restraints applied to each atom from 15 to 0 kcal mol⁻¹ Å⁻² over a total period of 1.2 ns at constant pressure and temperature, including a final stage of 0.2 ns without any restraint.

After equilibration of the relaxed oligomer, a series of simulation ‘windows’ for steadily increasing or decreasing values of θ_{ref} were carried-out for Umbrella Sampling (11,196,239). For each window i , θ_{ref} was modified by 5 degrees and restrained to the new value using the harmonic biasing potential $V_i(\theta)$ with a force constant of $k=0.01$ kcal.mol⁻¹.degrees⁻² and equilibrated within 0.2 ns. The final structure obtained at the end of this sampling was used as a starting point for the next window generating the starting conformations for the entire bending pathway. Individual window’s production runs were then performed simultaneously with a force constant of $k=0.2$ kcal.mol⁻¹ degree⁻² for at least 1.0 ns. Conformation were recorded every 2 ps. Backward simulations starting from the final structure (100° bent) were carried-out using the same protocol. Potential of Mean Force (PMF) calculations using the weighted histogram analysis method (WHAM, 194) described in section 2.8 were carried-out to extract the free energy associated to DNA bending.

Then for all simulation with the Parmbsc0 force field (19), the MD protocol using Amber 8.0 (203) and Potential of Mean Force (PMF) calculations using WHAM are *identical* to the above study. The Parmbsc0 Umbrella Sampling has been applied to four B-DNA 15-mers, d(CGCGCGCGCGCGCGC), d(CATATATATATATAC), d(CGCGCAAAAACGCGC) and d(CGCGCGCGC AAAAAC) referred to as [GC], [AT], [Atract-1] and [Atract-2] oligomers, respectively. In each case standard B-DNA start structures were also used for the simulations. In order to obtain better statistics at the base pair level, production simulation time was extended to 3 ns per umbrella sampling window in forward sampling series (1ns/windows in backward series) and global bending range increased from 0° up to 150°, see table I. This sums-up to a total production sampling time of 124 ns for each of the four oligomers studied (producing 248 000 DNA conformation snapshots). Conformations were analysed in term of

helical parameters using the program Curves (33,34). The conformations were scanned for important values of local roll, propeller, opening and stagger and selected conformations were analysed in more details.

In order to characterize changes in stiffness of base pair junctions with respect to increasing bending constraint we calculated the force constant $k_{i,j}$ for every base pair junction i , in each umbrella window j . $k_{i,j}$ was obtained from a harmonic analysis (112,114,240,241) of the local bending angles $a_{i,j}$ defined as:

$$a_{i,j} = \sqrt{(\text{roll}_{i,j}^2 + \text{tilt}_{i,j}^2)} \quad (53,54) \quad \text{eq. 4.6}$$

$$k_{i,j} = 1/2 \times k_B T \times (1 / \langle (a_{i,j} - a0_{i,j})^2 \rangle) \quad (7) \quad \text{eq. 4.7}$$

Thus $a_{i,j}$ corresponds to the bending angle for the junction i over the window j and generally represents unimodal (Gaussian) distributions as discussed in the results section. $a0_{i,j}$ corresponds to its most probable value. Angular brackets denote averaging over a given sampling window.

4.4 Results

Induced DNA bending during restrained energy minimization

Correlation with conventional parametric descriptions (1.3.2, model 1)

The current definition of a DNA bending coordinate based on two screw axis (handles) associated with the two ends of a DNA oligonucleotide (Figure 4.1b) allows continuous bending of DNA and at the same time full conformational flexibility of the molecule to relax towards a stable conformational state. The induced bending angle based on the angular orientation of the two handles correlates with the global axis curvature as defined by the program Curves (values with the latter are yet systematically lower; Figure 4.2). Note, that in Curves the global bending is given by the angle between the first and last segments of the curved helical axis. Even better correlation can be seen for the angle between two linear axis obtained from a Curves analysis of the two terminal 4 bp fragments (the same nucleotides that define the two handles; Figure 4.2, circles).

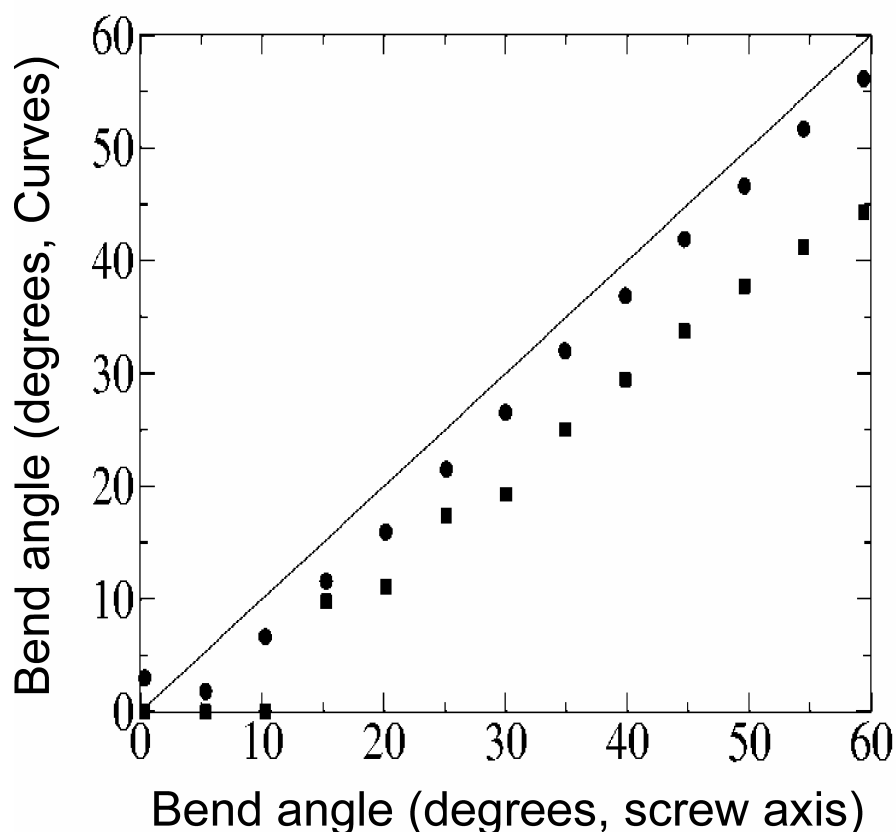


Figure 4.2. Correlation of DNA bend angles (15 bp polyCG) based on the screw axis description vs. global bending calculated using the program Curves5.0 (squares, ref. 34). In this case the global bending angle is defined by the angle between the helical axis direction at the first base pair vs. axis direction at the last base pair. The circles correspond to a Curves based bend angle between the linear helical axes of the terminal 4bp fragments of the DNA oligonucleotide. The line indicates perfect correlation.

Application to DNA oligonucleotides

The application of the restrained bending minimization method (energy minimization in 5° steps) for the Hagerman sequences (H1: (5'-dA₄T₄CG)_n and H2: (5'-dT₄A₄CG)_n with n=1,2) resulted in an energy minimum near zero degrees for the H2 sequences whereas energy minima at bending angles of ~15° (n=1) and 25° (n=2), respectively, were found in case of the H1 sequence (Figure 4.3a, onset). Only a small stabilization energy ~-0.3 kcal/mol per A₄T₄-tract compared to a straight structure was obtained. Such a small stabilizing energy is not surprising since experimentally already small changes in the temperature can significantly affect the curvature of DNA [“melting” of bent DNA structure (101)]. Based on the energetic

stabilization it is possible to calculate a probability distribution that takes into account the intrinsic geometric probabilities for bend angles (Figure 4.3). The probability distribution includes a Jacobian factor ($\sin(\text{bend angle})$) so that the probability distribution reflects the underlying free energy of bending which includes the bending energy but also the bending entropy (higher intrinsic probability of bending angles near 90° vs. bending angles around 0° or 180°). This results in distinct probability differences between H1(n=1), H1(n=2) and H2 sequences and also a slight shift of the free energy minimum (maximum of the probability density) compared to the potential energy minimum. Probability maxima of $\sim 17^\circ$ for H2 (n=1), 21° for H1 (n=1) and 31° for the H1 sequence (n=2) cases, respectively, with longer tails towards larger bend angles in the probability curves in case of the H1 sequences were obtained (Figure 4.3). The calculated bend angle maximum for the H1 sequence is close to experimental A-tract bending angle estimates of $17\text{-}21^\circ$ per A-tract (based on gel electrophoresis, 94). The results differ quantitatively from the study of Sanghani et al. (216) who employed a superhelical bending variable, a different force field (Flex force field) and a sigmoidal distance-dependent dielectric function representing solvent effects. Using the same sigmoidal distance-dependent dielectric function with the parm98 force field we found an energy minimum for H1 at a bending angle of 25° compared to $\sim 10^\circ$ for H2 (Figure 4.4). A distance dependent dielectric model appears to stabilize the curved H1 structure more strongly ($> 1 \text{ kcal mol}^{-1}$ compared to a straight structure) than using the present GB model.

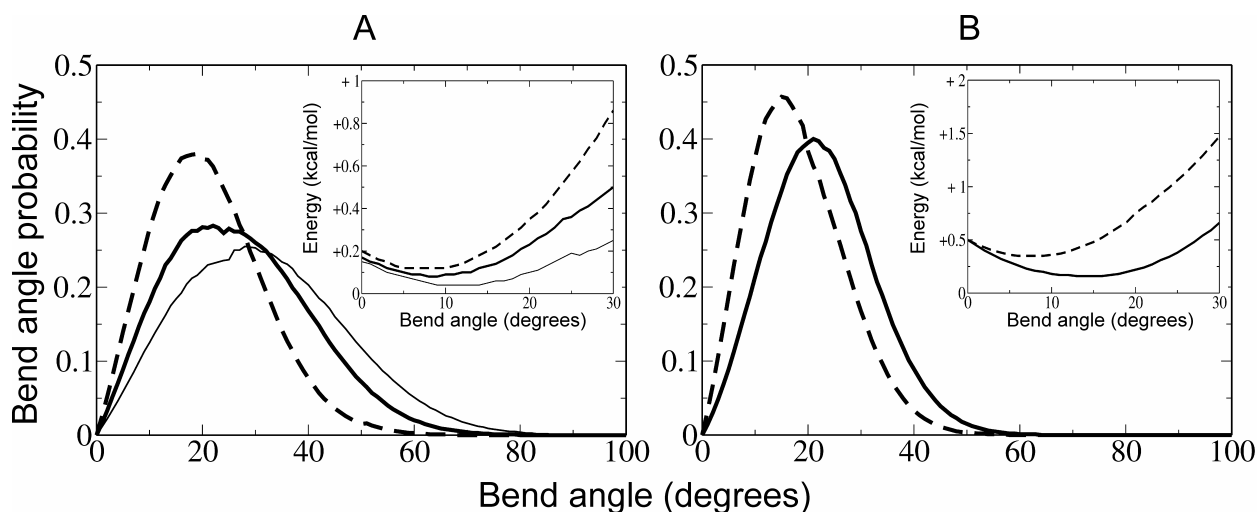


Figure 4.3. Bending angle probability distributions obtained from the potential energy curve as defined by, $p(\theta) = \sin\theta \exp[-E(\theta) / RT]$, vs. bending angles. (A) Bend angle probability for the Hagerman 1 sequence (bold line, $(5'\text{-dCGA}_4\text{T}_4\text{CG})_2$), two H1 sequences in phase with helical repeat (thin line, $(5'\text{-dCGA}_4\text{T}_4\text{CGA}_4\text{T}_4\text{CG})_2$) and Hagerman 2 (H2) sequence (dashed line, $(5'\text{-dCGT}_4\text{A}_4\text{CG})_2$). (B) Same for the HPV-E2-recognition sequence (bold line, $(5'\text{-}$

dACCGAATTCGGT)₂) and BPV-E2 recognition sequence (dashed line, (5'-dACCGACGTCGGT)₂). The corresponding potential energy vs. bend angle plots calculated using restraint energy minimization within Jumna (22) are shown as panel insets (same line types as for probability distributions).

The papilloma virus E2 protein binds to two sequence elements on DNA that are separated by a 4 bp linker not in contact with the protein (see ref. 242). Efficient interaction requires the DNA to bend towards the protein in the direction of the central minor groove (231, 242). It has been shown that the E2 recognition element in human DNA (HPV) adopts already a (pre)bent conformation (towards the minor groove) in the absence of the E2 protein whereas the corresponding bovine DNA recognition sequence is straight in solution influencing the binding affinities of HPV- and BPV-DNA for the E2 protein (83,231,243).

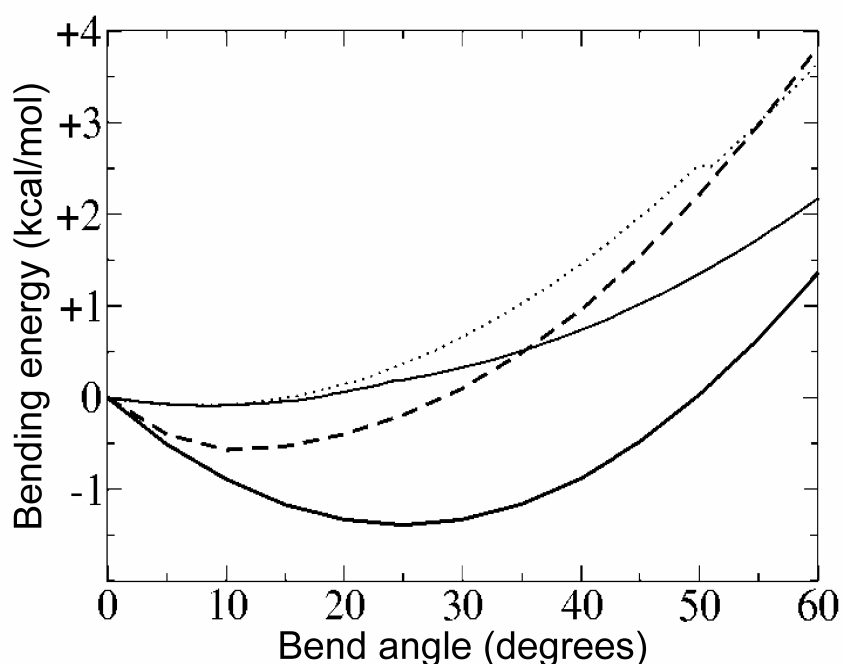


Figure 4.4. Potential energy vs. global bending: comparison of the Generalized Born (GB) model and a sigmoidal distance-dependent dielectric (sddd) function (see text and ref. 216). Calculations were performed on the H1-(5'-dCGA₄T₄CG)₂-sequence (thin line: GB-model, bold line: sddd-model) and the H2-(5'-dCGT₄A₄CG)₂-sequence (dotted line: GB-model, dashed line: sddd-model).

Indeed, restrained energy minimisations along the bending coordinate resulted in an energy minimum with a bending angle of ~18° for the HPV sequence (probability maximum at ~21°) and a steeper rise of the bending energy with a minimum near zero degree for the BPV sequence (probability maximum at ~15°, Figure 4.3b). The curvature estimated with Curves

was $\sim 15^\circ$ for HPV which is in good agreement with previous modelling studies on human and bovine E2 protein papillomavirus binding sites (83,242,243). The current bending angle calculated for the X-ray structures (231, 232) is, however, smaller for both HPV and BPV (7° and 4° respectively). A possible reason for the disagreement might be crystal packing effects that enforce a more straight conformation of the DNA in the crystal compared to free solution (as has been observed for A-tract DNA, 105).

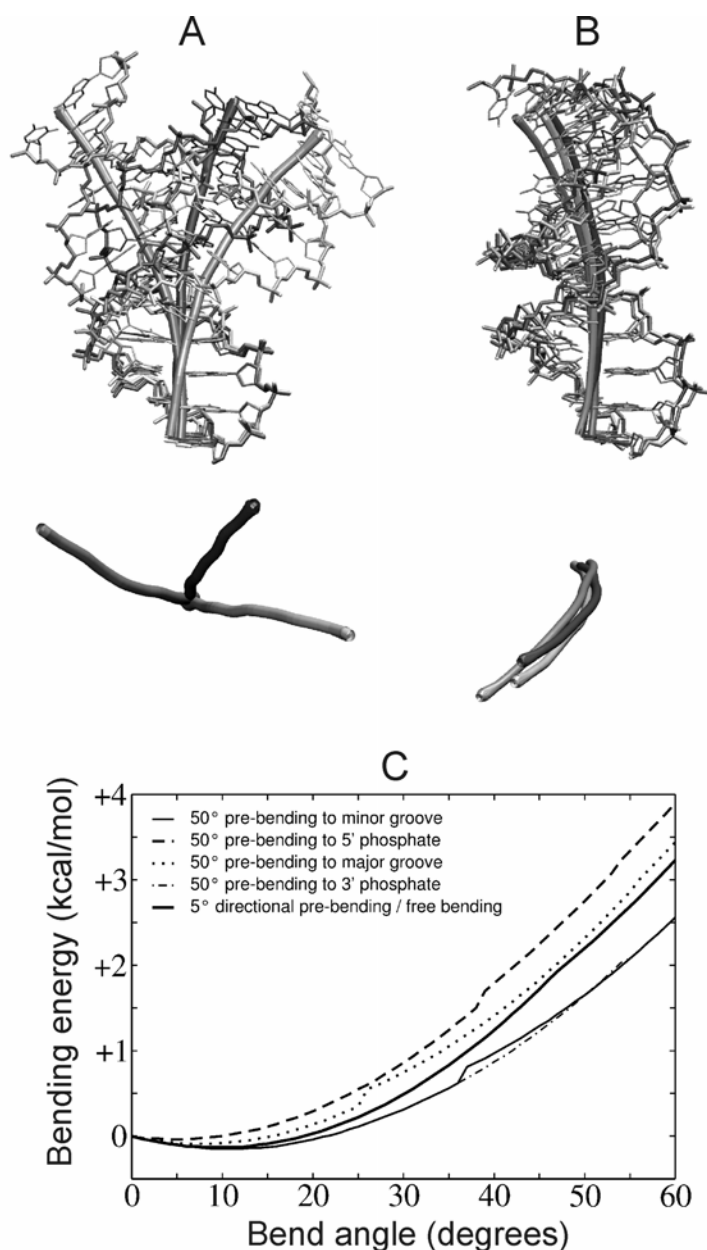


Figure 4.5. Influence of initial DNA bending direction (H1-sequence) on results of restrained bending minimization including only a restrain on bending magnitude. (A) Superposition of three DNA start structures (stick models at different grey levels) bent in different directions by 50° . Bending directions were towards the central major groove (darkest grey) and towards the phosphate groups closest to the

5'- or 3'-ends of one DNA end, respectively (bending direction approximately perpendicular to bending towards central minor or major grooves). The helical axis of each DNA calculated using the program Curves is indicated (bold lines). The view in the upper panel is orthogonal to the helical axis whereas in the lower panel a perpendicular view approximately along the helical axis is given (only the helical axis for each DNA is shown).

(B) Same for the energy minimized structures (starting from the structures shown in A) after removal of the restrain on bending direction but keeping a restrain on bending magnitude.

(C) Calculated potential energy vs. bending magnitude starting from various DNA structures pre-bend in different directions and by different magnitudes (according to line type indicated in the inset of the Figure). Free bending corresponds to the bending energy curve obtained by starting from ideal B-DNA (same curves as obtained by starting from any of the 5° pre-bend structures; the 5° pre-bend structures were generated in 10° steps of the bending direction).

The restrained bending minimization studies described above were all performed starting from ideal B-DNA with zero initial bend angle and without any restriction on the initial bending direction. It is possible that the initial bending direction has a significant influence on the subsequent bending steps and the resulting structures and energies. In order to control the influence of the initial bending direction on the restrained bending minimization results the start structures (H1-sequence) were pre-bend in various possible directions (see Methods and paragraph on directional DNA bending). The structures were then subjected to restrained bending minimization without any restrains on the bending direction. Interestingly, for small initial bending angles (5° bending magnitude and all possible bending directions in 10° steps) the same bending energy curve and identical minimized structures were obtained as for the bending minimization starting from ideal B-DNA (no pre-bending, Figure 4.5). At a pre-bending of 50° slightly different bending energy curves depending on the direction of the pre-bending were obtained (Figure 4.5). However, even at a large pre-bending of 50° in various directions the structures undergo large axial rotations such that the final bending direction (upon restraining only the bending magnitude not the direction) was very similar (directed towards the central minor groove, Figure 4.5a,b). This result indicates that there is a significant directional preference for DNA bending (analysed in more detail in the paragraph on directional DNA bending) and the energy surface for DNA bending is relatively smooth such that relatively large axial rotations are possible upon removal of a restrain on the bending direction. The differences in the final structures (and energies) upon removal of the directional restraint are mainly due to differences in the conformational DNA backbone substates in the structures obtained after the initial directional bending. It is of interest to note,

that if one restrains the dihedral backbone of the DNA during initial directional bending (50° case) close to standard B-DNA and then removes the directional restraint (only keep the restrain on the bending magnitude) all start structures relax to the same structure (bend towards the central minor groove).

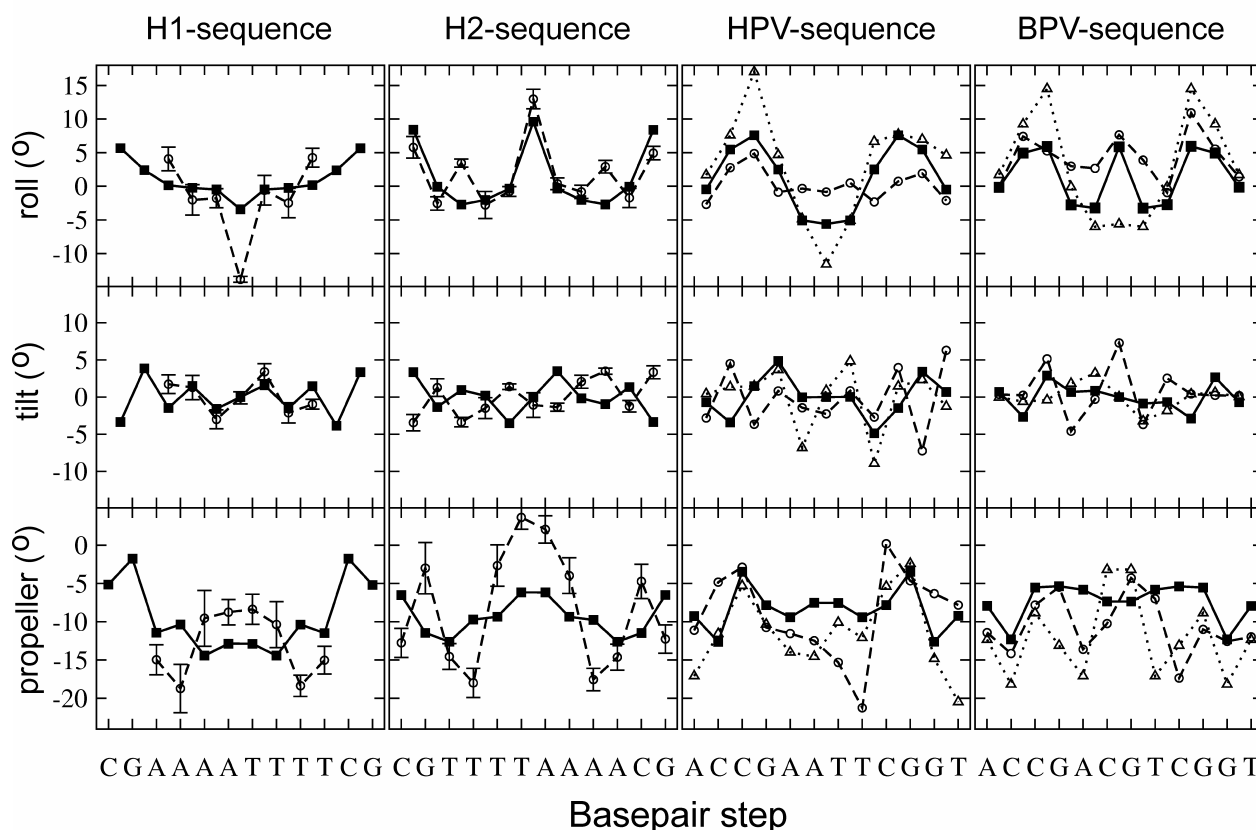


Figure 4.6. Helical parameters roll, tilt and propeller twist (filled squares in each plot) versus base sequence for the most stable bent conformations (probability maxima in Figure 4.3). All calculations were performed on palindromic (single repeat) sequences. Experimental data for NMR derived structures of the H1 (5'-dCGA₄T₄CG)₂) and H2-(5'-dCGT₄A₄CG)₂) sequences and X-ray structures of the unbound HPV-E2 and BPV-E2 recognition sequences are indicated as circles (dashed lines). Peripheral base-pair are not reported for the NMR structure of H1 which contains GC instead of CG flanking steps. Error bars have been derived from the analysis of all published NMR structures in each pdb-entry. For the HPV-E2 and BPV-E2 recognition sequences experimental data on the DNA molecules in complex with E2-proteins are plotted as triangles (dotted line).

Helical parameters of bent DNA structures

We analysed the molecular origin of bending in terms of local helical parameters of the optimally bent DNA structures calculated with the program Curves (Figure 4.6). Several studies have already been reported to highlight the structural origin of curvature for DNA

fragments containing runs of adenines (A-tract, see Introduction). A major finding of X-ray crystallography is that A-tracts and the H1 sequence adopt an exceptionally narrow minor groove and the A:T base-pairs are highly propeller twisted (244). In agreement with experiment (98,244,245) the bending energy minimum observed for the H1 sequence showed significant (negative) propeller twisting of the complete central segment (Figure 4.6). The calculated bending direction was, indeed, towards the central minor groove. An important point concerning the H1 sequence is that the central AT step does not disrupt the structural uniformity of A-tract stacking (negative propeller twist pattern, 244, 245). In addition, for the H1 sequence the positive roll angles at the CG steps flanking the central A-tract add up in phase with a small negative roll at the central AT (half a turn away) resulting in the observed global bending (Figure 4.6). This bending mechanism is in good qualitative agreement with experiment (NMR structure of the H1-sequence, 98) that also shows continuous negative propeller twisting of the central segment. However, the magnitude of the negative roll at the central AT step is larger than for the calculated structure.

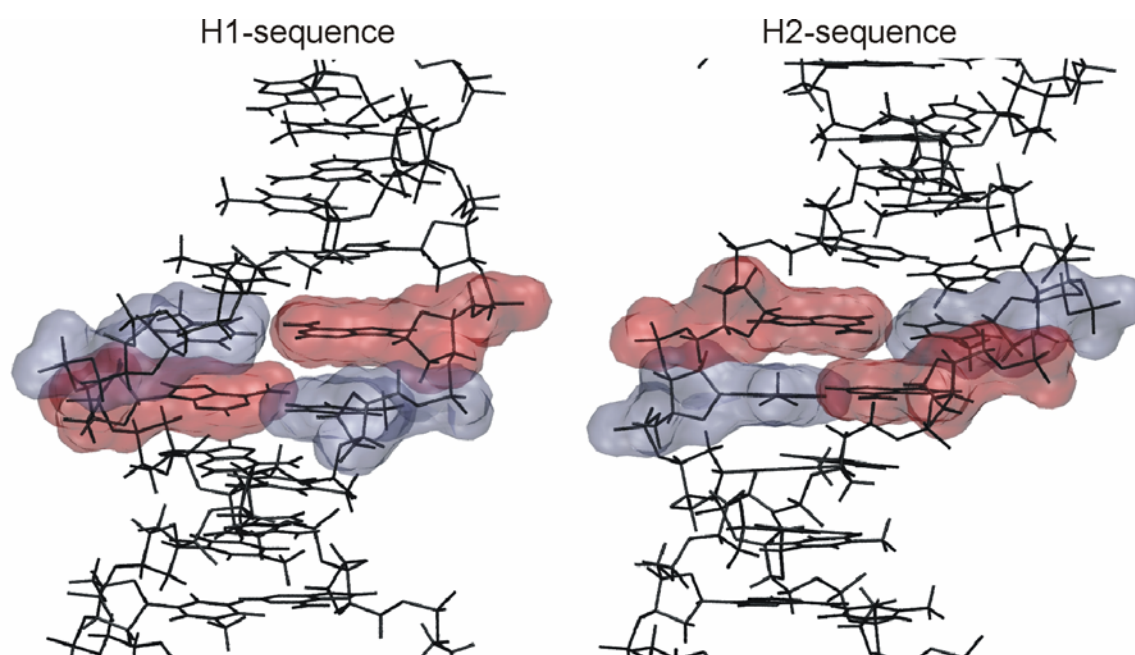


Figure 4.7. View into the minor grooves of the central AT and TA bp steps of the Hagerman H1 and H2 sequences, respectively (bond stick and opaque molecular surface representation). The central adenine/thymine nucleotides are indicated in red/blue, respectively. In case of the H2 sequence a central bending towards the minor groove interferes with a potential steric clash of the central adenine bases (partial cross stacking; not the case for the H1-sequence).

For H2, both the calculations and the analysis of the experimental NMR structure (98) indicate on average more positive propeller twist at the central segment (smaller fluctuation in the calculated structure). The calculated roll angle pattern agrees almost quantitatively with experiment. The positive roll angles at flanking CG steps are out-balanced by a large central positive roll. Consequently, significantly larger local roll changes are necessary in order to add up to achieve global bending compared to H1 (energetically unfavourable). In addition, the central negative propeller twist pattern (necessary to narrow the minor groove) is interrupted at the central TA step of H2 by a reduced magnitude of negative propeller twisting (almost zero for the experimental structure). In Figure 4.7 we illustrate the dependence between negative propeller twist of H1/H2 sequences and the minor groove narrowing. In the case of A tract and H1-sequences the negative propeller represents a sterically favouring mechanism, whereas for H2 the reduced magnitude of propeller twisting can be explained by the steric hindrance of the two central adenine amino groups upon negative propeller twisting, in agreement with classical Calladine rules (31).

When looking at the HPV/BPV pair, where the unique difference between the two sequences is a substitution of a central AT by a central CG, one can draw up basically the same interpretation of sequence effects as detailed above for H1/H2. Indeed, positive rolls at CG steps add up with a small negative roll at the central AT in HPV, but are opposed by a large positive roll at the central CG in BPV. Together with the smaller variation in tilt this results in a net curvature in the former sequence and confirms the structural uniformity of AT step in different sequence contexts. The results on the calculated structure are in qualitative agreement with the experimental free DNA structure (231,232) and in even better agreement with the DNA bound to the E2-protein (233,234). This concerns the pattern of roll and tilt angles as well as the on average negative propeller twist angles of the central segment (Figure 4.6). For both the HPV and BPV sequences the propeller twisting in the experimental structure shows larger fluctuations compared to the calculated structures.

Directional bending of B-DNA

In addition to free relaxation upon induced DNA bending it is also possible to add an additional restraint for restricting the bending in a pre-selected direction. A two-dimensional (2D) directional bendability map provides an instantaneous picture of the propensity of a DNA to bend in any possible direction. Bending directions were systematically checked using both Curves (version 5) (34) and Madbend (209) (data not shown). Instead of the potential

energy, we use the probability density for bending with respect to the bending amplitude θ for a given direction α , expressed as $p_\theta = \exp [-E_\theta / k_B T] / \sum \exp [-E_\theta / k_B T]$ where the sum runs over all value of θ corresponding to one particular direction of bending. This variable enhances contrasts and highlights more selectively global bending properties compared to a plot of the bending energy. In addition, the probability maximum (most likely conformation) with respect to the total 2D-probability distribution is also reported (squares in Figs. 4.8, 4.9). For the oligonucleotides with a central A₆-tract and for the H1 sequence the directional 2D-bendability map indicate bending probability contours that extend to large values for bending towards the minor groove (Figure 4.8). The result agrees with unrestrained MD simulation studies and NMR spectroscopy of the A-tract containing DNA fragments that also indicated curvature towards the DNA minor groove (210-213). Other bending directions are significantly less favourable. Note, that only a directional bending map as calculated in the present study can give an ultimate indication of a preferred DNA bending preference since both bending magnitude and direction are treated as variables that can be changed separately. Almost quantitative agreement between experiment and most probable DNA conformation in terms of bending magnitude and direction have been obtained (square and cross in the plot for the A-tract and other DNA structures, Figure 4.8). The H1 sequence exhibits a slightly different bending direction and slightly reduced bendability compared to the A-tract with six consecutive adenines. Experimental studies have indeed reported maximum electrophoretic anomalies for A-tracts of length 6 (96). In case of the H2 sequence and the d(5'-(AT)₆)₂ oligonucleotide a smaller overall bending tendency towards the minor groove was obtained. The $p=0.1$ contour was in both cases $< 25^\circ$ whereas in case of the A6-tract and H1 sequence the $p=0.1$ contour reached bend angles of $> 35^\circ$ (Figure 4.8).

Noticeably, for bending towards the central major groove (direction angle $\sim 180^\circ$) the $p=0.1$ contour for the alternating AT sequence was slightly larger than for the H1 or A-tract case. The energetic cost was similar (~ 5 kcal/mol) for a 60° bending in the direction of the central minor or major groove for the alternating AT sequence but roughly doubled (~ 10 kcal/mol) for the latter direction in case of A-tract, H1 and H2 sequences. Hence, the directional bendability map predicts a much easier bending of alternating TA sequences towards the major groove. Indeed, alternating AT sequences are often found as a part of recognition elements for DNA minor-groove binding proteins. Binding to the DNA minor groove requires minor groove opening and bending towards the major groove. Hence, the easier bending deformability of short alternating AT sequences toward the major groove may contribute

indirectly to the recognition by the minor groove binding proteins as also suggested by other studies (209).

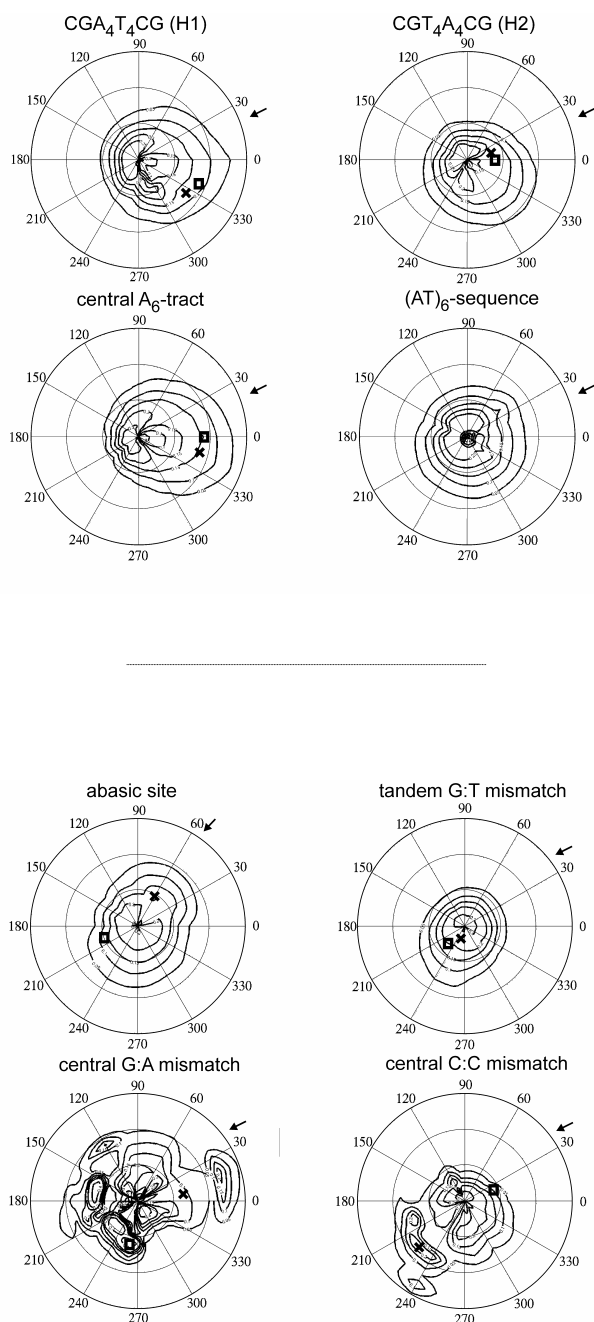


Figure 4.8. Iso-probability surfaces for directional DNA bending. Bend angles up to 45° have been considered (largest circle; inner circles correspond to 15° and 30° bending angle magnitude, respectively). Meridians correspond to bending directions (0 to 360°; ticks on the outer circle). The direction of bending towards the central minor groove has been marked by an arrow. The contours (in steps of 0.05 with a maximum of 0.25) are constant Boltzmann probabilities normalized independently in each bending direction as discussed in Methods and Results sections. The squares in each plot correspond to the probability maximum of the complete 2D-probability distribution (most stable bent conformation). The cross represents the directional bending state of the experimental conformation for each case (except for the (AT)₆-sequence for which no experimental data is available).

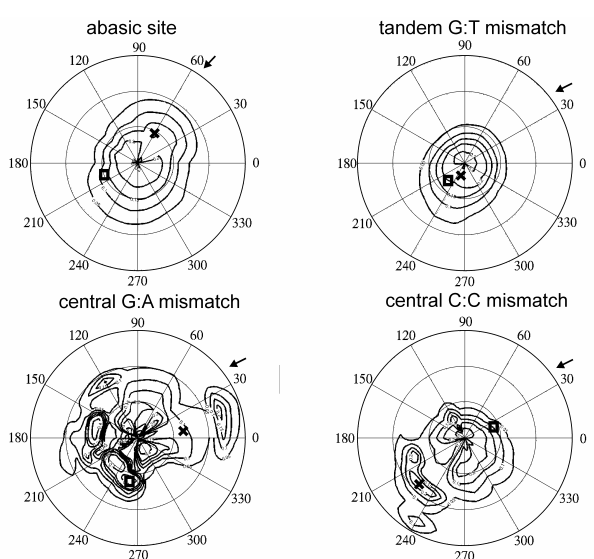


Figure 4.9. Iso-probability surfaces for directional bending of abasic DNA (5'-dCGTAC-CATGC /dGCATGAGTACG, 235), a tandem G:T mismatch ((5'-dCCATGCGTGG)₂, 238), a central G:A mismatch (5'-dGCTTCAGTCGT /5'-dACGACGGAAGC, 236) and a C:C mismatch (5'-dGCCACCAGCTC /5'-dGAGCTCGTGGC, 237).

Directional bending of damaged and mismatched DNA

Common examples of lesions in DNA include mismatches and the formation of single abasic sites. Changes in the global structure and bending deformability may play a role for the recognition of damaged DNA by repair enzymes (128,129). The presence of abasic sites in DNA can also affect structural and dynamic properties of the adjacent duplex DNA. MD (77,

130) and EM (131) studies indicate significant bending associated with a decrease in inter-strand interaction energy at the site of the lesion. However, a systematic analysis of the bendability of abasic sites and mismatches in all possible bending directions has so far not been performed. In order to do such an analysis we chose DNA oligomers for which experimental structures are available with either a central abasic site, a single C:C mismatch, a single G:A mismatch or a tandem G:T mismatch. In the case of the abasic site the experimental structure indicates a bend conformation with similar bending magnitude as the most probable conformation of our 2D bending map but with opposite bending direction (square and cross in Figure 4.9). However, the experimental conformation lies within a region of high calculated bending probability. This result indicates that the abasic site might be compatible with multiple possible conformations that are overall bent in different directions (131).

The directional bendability map for the tandem G:T mismatch indicates little difference to regular B-DNA and a slight preference for bending towards the central major groove. The predicted most probable bending magnitude and direction agree very well with the experimental results (Figure 4.9). On the contrary, the 2D bendability maps for both a single central G:A and a single C:C mismatch differ significantly from regular B-DNA, showing much more complex behaviour. In both cases several probability submaxima are observed indicating several possible stable substates with different global bend angles and bending directions. The calculated most probable bending magnitude and direction differ from experiment, which is not surprising, since the balance between different minima may easily be changed. Hence, our results show that a greater variety of alternative stable bending directions can be accommodated by these two sequences if external factors (here a virtual bending coordinate) act on these structures. This property may play a decisive role for the recognition of these elements by repair enzymes.

Comparison with Umbrella Sampling Molecular Dynamics (Atract, parm94 force field)

The adiabatic mapping of DNA bending employing an implicit continuum solvent model allows the systematic investigation of DNA deformability as a function of bending direction and magnitude. In order to investigate the effect of a more realistic representation of surrounding aqueous solvent and ions we performed dynamics simulations of DNA bending. In order to calculate free energies of bending we used the umbrella sampling technique. This approach allows in principle also the sampling of several conformational DNA substates

compatible with a given global bend angle (not only energy minima as during the adiabatic mapping calculations). The MD/umbrella sampling approach is, however, computationally much more demanding than the adiabatic mapping EM method. As a first step therefore, MD simulations were limited to one A-tract containing DNA and results were directly compared to adiabatic mapping using the same molecule (5'-dCGCGCA₅CGCGC)₂. To control the convergence of the simulations, forward simulations starting from B-DNA and backward umbrella sampling simulations starting from the final structure of the forward simulation (but otherwise independent) were performed. In addition, the free energy curves were calculated for two simulation times of 0.5 and 1 ns per umbrella sampling window (spacing of windows: 5°), respectively. The force constant to restrain the bending angle to a window reference angle was small enough ($k=0.01 \text{ kcal mol}^{-1} \text{ degrees}^{-2}$) to allow for sufficient overlap between neighbouring angle windows. The resulting free energy curves were very similar with a variation of $< 0.5 \text{ kcal/mol}$ at each data point and a free energy minimum close to $\theta = 20^\circ$ (Figure 4.10). The average bend angle calculated with Curves was also $\sim 20^\circ$ (average of conformations from the corresponding bend angle restraining window). The optimal bend angle from the MD umbrella sampling calculations shows a good agreement with the global minimum of bending estimated for this sequence when using the EM approach (dashed line in Figure 4.10) and is also in good agreement with experimental observations (94).

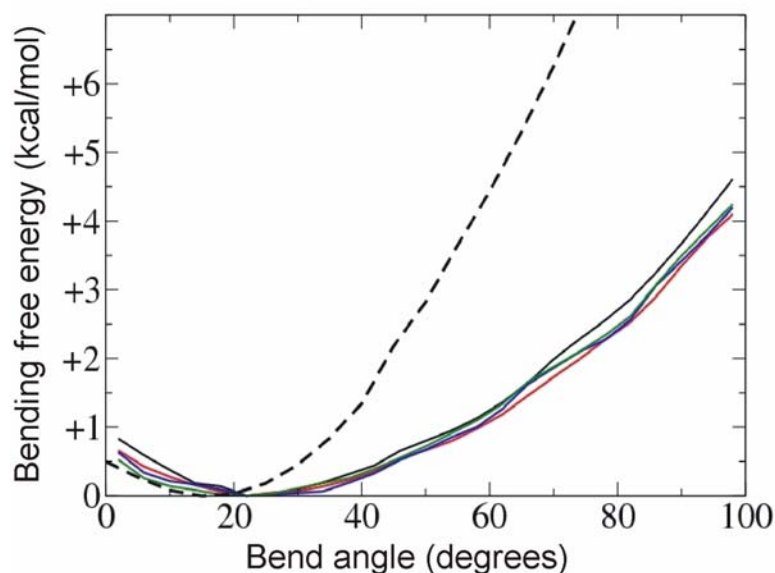


Figure 4.10. Potential of mean force for bending of an A-tract DNA oligomer with the sequence (5'-dCGCGCA₅CGCGC)₂ using molecular dynamics (parm94) umbrella sampling simulations and a bend angle restraining step size of 5° (see Methods for details). Calculated free energy curves are plotted for forward simulations after 0.5 ns (black) and 1 ns (red) of data collecting time for each umbrella

sampling window. Backward simulation were started from the fully bent structures and free energy curves were calculated independently from the forward simulations (blue and green curves for 0.5 ns and 1 ns sampling time per window, respectively). The variation of each independently calculated free energy at each bend angle was $< \pm 0.5$ kcal/mol. The dashed line corresponds to the potential bending energy as obtained from restrained energy minimization (adiabatic mapping) of the same DNA molecule.

Bending up to $\theta = 100^\circ$ required a total free energy change of 4 kcal/mol which is significantly smaller than the calculated bending energy change from the adiabatic mapping with a GB solvent model. This result indicates that either the explicit solvent representation significantly softens the bending deformability of DNA or the MD simulations (in Cartesian coordinates instead of internal coordinates in JUMNA) provides many more degrees of freedom to relax the bending induced strain on the DNA conformation. It is possible to fit a quadratic function to the bending free energy curve near the free energy minimum. From the associated curvature it is possible to calculate an estimate of the persistence length P of the DNA molecule (using ref. 67 i.e. $P = 2 \cdot \text{contour length (in bp steps)} / \langle \text{bend angle variance at room temperature} \rangle$). In case of the MD simulations a persistence length between 170-180 bp was obtained in good agreement with experiment (~ 150 bp, 67). The adiabatic mapping predicts a stiffer DNA with a persistence length of 350 bp, more than twice the experimental value for DNA (67). During both, the free energy simulation and the adiabatic mapping, bending was strongly anisotropic, directed toward the minor groove of the central A-tract.

Helical conformational changes induced upon A-tract bending (MD, parm94)

Helical DNA parameters of conformations near the optimal bending angles ($20-30^\circ$) and at higher induced bend angles of $\sim 80-100^\circ$ were analysed (Figure 4.11). For comparison the helical parameters of the structures corresponding to 80° and 20° obtained by adiabatic mapping for the same oligonucleotide were also calculated. An experimental structure of exactly the same sequence is not available, however, the structure of a close sequence, with a central A6 instead of an A5 tract has been determined by NMR spectroscopy (230). The helical parameters for this structure (except for the central AA step) have also been calculated for comparison with experiment. Both types of calculations (EM adiabatic mapping and MD simulations) suggest qualitatively the same bending mechanism in line with the analysis of the Hagerman sequences (see above). Both in the regime of optimal bending as well as in the

regime of strongly induced bending the results from the EM calculations are close or within the error bars obtained from the analysis of the MD Umbrella Sampling simulations (Figure 4.11).

At the optimal bending angle both the adiabatic EM mapping and the Umbrella Sampling/MD simulations indicate significant negative propeller twisting of the central basepairs ($\sim -12^\circ$) and a significant narrowing of the minor groove relative to the flanking sequences. This is also seen for the experimental structure (Figure 4.11). The average propeller twist in the central segment indicates that interstrand bifurcated hydrogen bonds linking the adenine residue N6-H amino group to adjacent thymine O4 group across the major groove are at most transient, since they require a propeller close to -20° . This supports the notion that bifurcated hydrogen bonds are not a driving force in the A-tract induced bending phenomenon (212, 247). Slightly negative roll angles within the A-tract and on average slightly decreasing tilt angles along the A-tract were observed. However, large positive rolls were seen at both peripheral CG and CA steps, respectively, associated with a reduction of twist. Almost quantitative agreement between the experimentally observed pattern of roll and tilt angles and the calculations at the optimal bending angle (for both EM / MD) was obtained (Figure 4.11). In the regime of high curvature ($80-100^\circ$), the above picture of local changes in the conformation does not change dramatically (no completely new or different bending mechanism is seen). It evolves continuously so that trends observed at optimal bending can be seen even more clearly. The pattern of roll angles along the sequence shows significantly increased positive values on both sides of the A-tract except for the 3' purine-pyrimidine junction, and negative values ($\sim -5^\circ$) at each AA step (for both adiabatic mapping and free energy MD simulation). The positive roll on both sides of the A-tract is coupled to an unwinding of the helix especially at the CG steps flanking the A-tract (Figure 4.11). Also, the tilt profile indicates slightly positive values at the beginning of the A-tract (CA step) and slightly more negative values at its 3' side. The minor groove narrowing and the negative propeller-twist angles along the A-tract are more pronounced than for the small bending regime.

Overall, the trends observed during both adiabatic mapping and during MD free energy simulations agree qualitatively very well with experimental results on A-tract bending (106, 230) and free MD simulation studies (211-213). Solution studies of oligonucleotides containing A₄ tract (106) and A₆ tract (230) by NMR spectroscopy reported a very similar roll pattern along the DNA molecules with significant positive rolls at base pair steps flanking the A-tract,

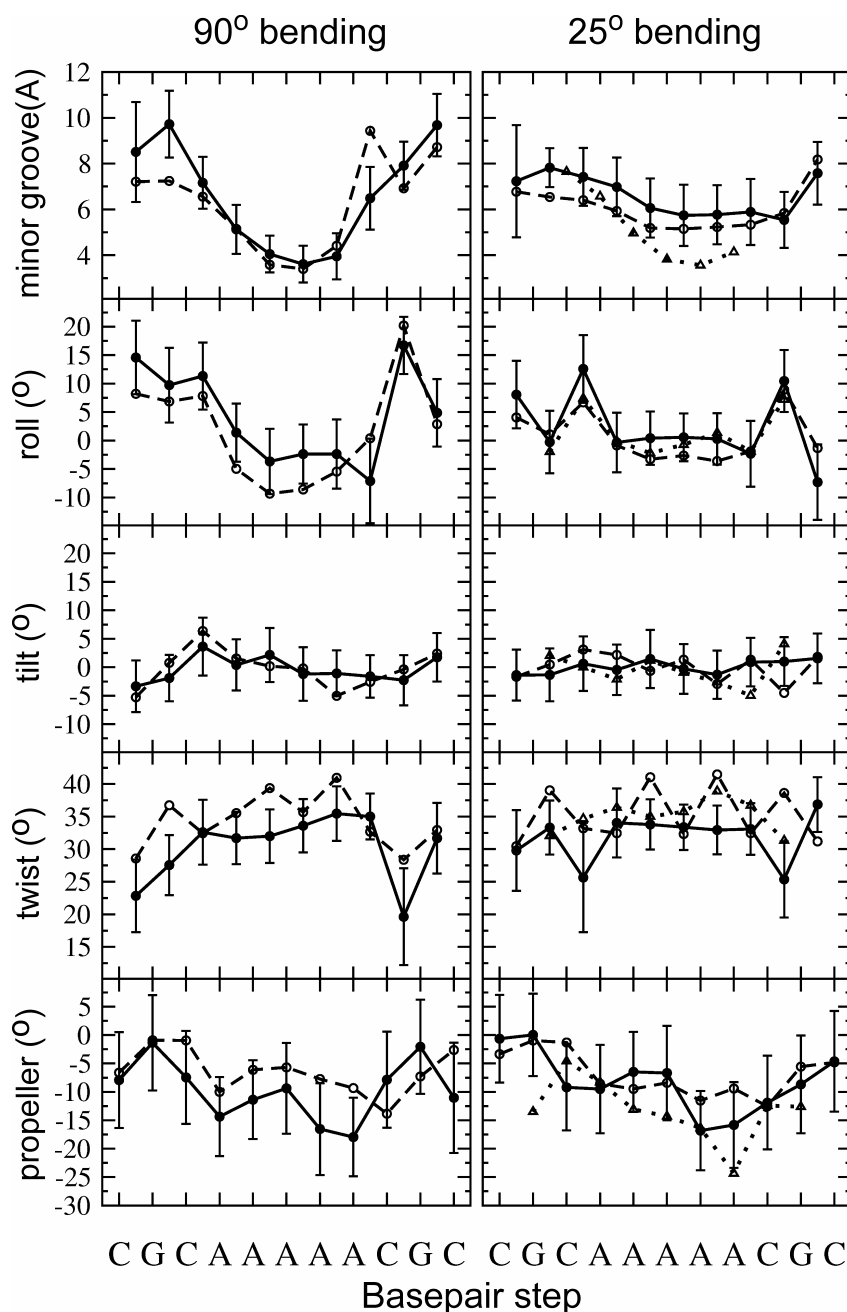


Figure 4.11. Average helical parameters in the regime of high and small curvature for the central sequence (5'-dCGCA₅CGC)₂ during MD free energy simulations (filled circles and full line) and during adiabatic mapping using energy minimization (open circles and dashed line). Averages and standard deviations (indicated as error bars) were taken from DNA conformations sampled in the umbrella sampling windows in the range of 20° to 40° for small (near optimal) curvature and in the range of 80° to 100° for tight bending. Parameters obtained from the adiabatic mapping approach were calculated for energy minimised structures corresponding to 25° and 80° bending, respectively. For comparison, helical parameters extracted from the experimental structure of a similar A-tract DNA ((5'-dCGCA₆GCG)₂, 230) that adopts a bent conformation in solution (~20°) was also included (triangles and dots curve). To allow direct comparison the central helical step for the experimental structure is not shown.

coupled to an unwinding at those steps and small negative roll angles within the A₆ tract (230). In addition, the trend observed for tilt (positive near the 5'-end of the A-tract and negative at the 3'-end) was also observed in both NMR studies (106, 230).

The above analysis of the helical parameters is largely compatible with the junction model and a recent extension, termed Delocalized Bend (DB) Model (106) which combines elements of the wedge (103,104) and junction models (94). The junction model assigns the main cause of bending to elements at the junction of A-tracts. In the DB model significant contributions to the overall bend arise also from roll and tilt of the dinucleotides inside the A-tract. Such contributions can be seen in the present simulations especially in the regime of strong DNA curvature.

Coupling of bending deformations to other global parameters of DNA (Atract, MD, parm94)

In contrast to free MD simulations or other restraining approaches the present methodology allows the continuous smooth global DNA bending much beyond equilibrium levels as it can for example occur upon protein-DNA binding. It is also possible to investigate systematically the coupled changes of other global parameters upon inducing DNA curvature. Unexpectedly, the average minor groove width of the A₅-tract containing DNA fragment that was studied by EM and Umbrella Sampling simulations showed only an insignificant reduction with increasing bend angle. However, this was largely due to an opposing effect of the central A-tract (narrowing of the minor groove) and the flanking sequences. From the analysis of complexes of proteins that bind in the minor groove of DNA it is known that minor groove opening results in unwinding (untwisting) of DNA. This coupling was also seen for the central A-tract segment that showed an increase of the average twist with increasing bend angle (and decreasing minor groove width, Figure 4.12). In contrast, the average twist and average rise (path length / number of base-pair steps) per bp of the flanking sequences started to decrease upon bending beyond 40°. Opposing trends seen for the central and the flanking segments appear to reduce the overall coupling of bending to other global conformational parameters of the DNA molecule. The specific trends for A-tract and flanking sequences concerning the stretching, winding and average minor groove widths were found in both EM and MD calculations (Figure 4.12). Interestingly, the fluctuation of twist, rise and minor groove width of the central A-tract (error bars in Figure 4.12) do not change with increasing bend angle and are always significantly smaller for the central A-tract compared to the

flanking sequences. This result indicates that the central A-tract forms a relatively rigid segment that largely keeps its rigidity even under significant bending stress.

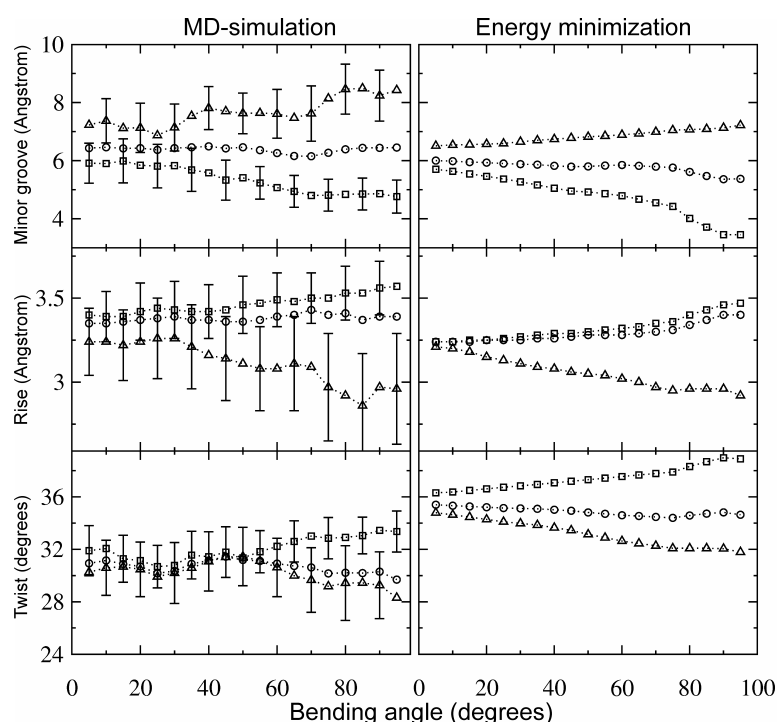


Figure 4.12. Correlated changes of global conformational variables upon induced DNA bending of the (5'-dCGCA₅CGC)₂ oligonucleotide. The average minor groove width, rise per base-pair step and average DNA twist are plotted vs. induced global bend angle (circles) for both the MD umbrella sampling simulation and adiabatic mapping EM (using *Curves* global parameters). In addition to the average along the complete DNA, corresponding averages for the central (5bp, squares) and terminal parts (triangles) of the DNA are also shown.

DNA Bending Free Energies simulations with parmbsc0

Free Energies global pathways

The free energy change associated with global bending of each DNA duplex was calculated for forward simulation times of 2 and 3 ns, respectively, per umbrella sampling window spaced by 5° (Fig. 4.13). In the case of the dCA(TA)₆G duplex [AT] a clear transition is observed at 80° bending and for higher bend angles the duplex started to dissociate. The curves for 2 ns and 3 ns simulation times are very similar indicating good convergence of conformational ensembles for forward sampling. Backward simulations starting from the final structure of the forward simulation were also performed (1 ns/window).

For all cases with the exception of [Atract-2] a significant hysteresis was observed (not shown) presumably due to insufficient sampling of all relevant states on the present time scale. However, when starting from the structure bent to 100° in the forward direction very similar free energy curves were obtained with the exception of the [AT] sequence (Fig. 4.13).

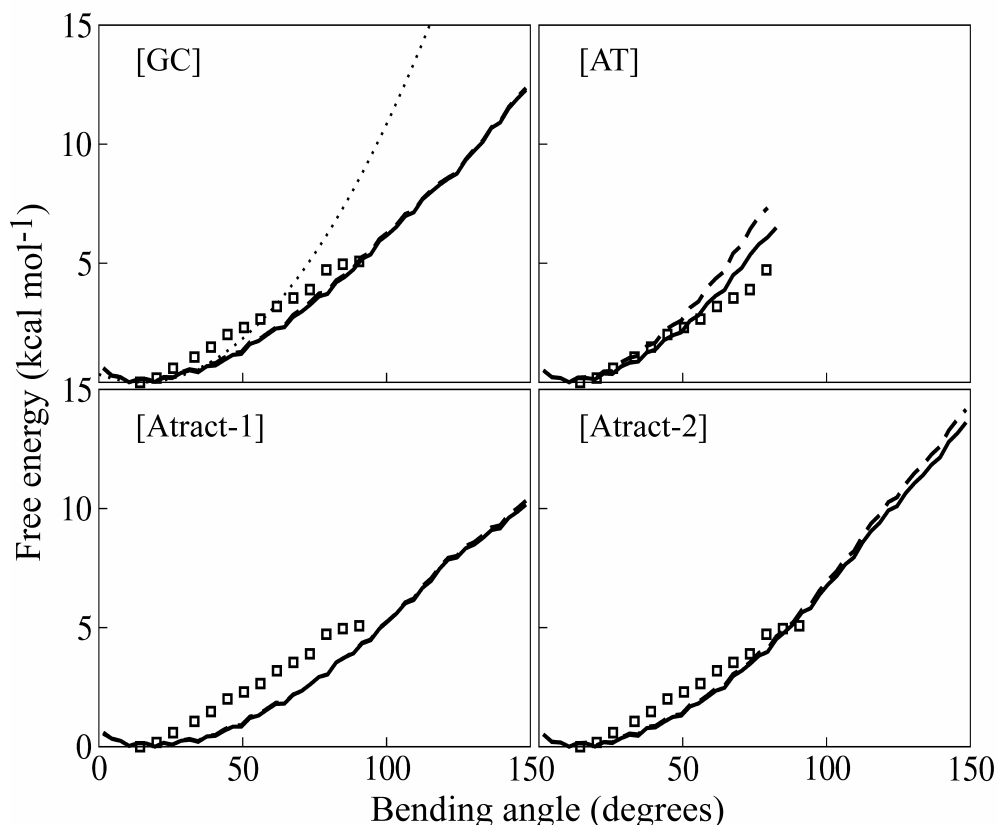


Figure 4.13. Calculated free energy for bending DNA oligomers are plotted for 2 ns (bold dashed line) and 3 ns (bold full line) sampling per 5° window. The bending free energy deduced from AFM experiments (4) is shown by open squares. A quadratic energy function based on the equilibrium persistence length is drawn for the [GC] oligomer (dotted line) to help visualize the departure of the calculated bending free energy from the harmonic regime. For [AT] free energy curves are not shown beyond 80° because the two strands dissociate and clearly break down the reaction coordinate for bending (this is demonstrated in Fig.6 of Annexe I).

This indicates good convergence of the free energy curves up to 100° but the presence of metastable states for bending higher than 100°, with probably significant energy barriers that prevent spontaneous relaxation under reverse bending strain. For the [GC] and [AT] sequences the global free energy minimum corresponds to 10° and for [Atract-1] and [Atract-2] it is shifted to larger values near 20°, in agreement with experiment that indicates stable bending of such sequence in solution (94). The overall shape of the bending free energy curves is overall similar for the four sequences. For small bending angles (up to 50°) the free energy curves follow closely a quadratic curve based on the equilibrium persistence length of DNA (indicated for [GC] as dotted line in Fig. 4.13).

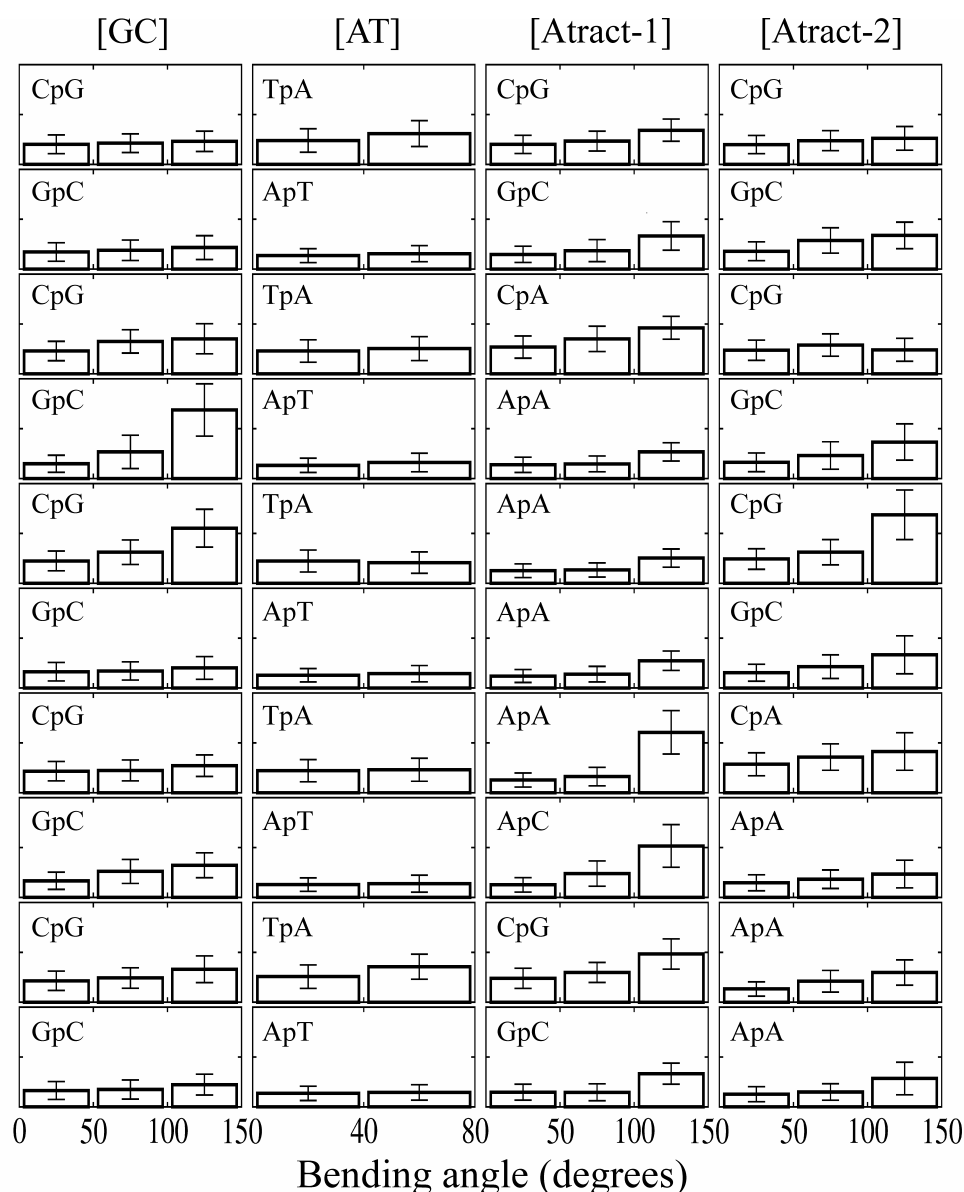


Figure 4.14. Mean hinge angle α (vertical axis, scale 0°-50°) and its standard deviation (cross-bars) for the 10 central junctions of the four oligomers and for three bending regimes (weak: 0-50°, medium: 50-100°, strong: 100-150°). For [AT] only two bending regimes are shown, 0-40° and 40°-80°.

However, global bending beyond this “small” bending regime is clearly not quadratic and flattens out to an almost linear increase of the free energy with bending. It is interesting to note that such almost linear dependence for strong bending of DNA has recently been measured experimentally on random individual DNA molecules on same length scale (5 nm) by high resolution Atomic Force Microscopy (4). The experimentally obtained free energy changes follow nicely the trend seen in current calculations (squares in Fig. 4.13.). Bending up to 90° requires a total free energy change of 5. +/-0.5 kcal/mol for A-tract sequences and 6.

+/-1 kcal/mol for alternating [GC] and [AT] sequences (compared to 5 kcal/mol in (4)). For higher bending, up to 150°, which was not probed by AFM experiments we obtained the values of 13. +/-0.5 kcal/mol for [GC] and [Atract-2] and ~10 kcal/mol for [Atract-1]. A structural explanation for this difference in bending free energy between the two A-tract containing oligomers is given in the next section. It is possible to fit a linear regression to the PMF for the bending angles above 60°, with the slope value in kcal.mol⁻¹radian⁻¹ of 7.4 [Atract-2], 6.9 [GC] and 5.7 ([AT] and [Atract-1]). These values might be useful as sequence specific bending elasticity for DNA Sub-Elastic Chain (SEC) models recently developed (4,72). They are indeed in agreement with the coefficient of 6.8 found experimentally (4) averaged over a large number of AFM samples (diverse sequences).

Weak bending of DNA

For each DNA oligonucleotide, bending up to 50-100° can be considered as “smooth” since it involves mostly roll angles with amplitudes up to 20°. This is illustrated in Fig. 4.14 which shows the local bend angle at each base pair step. The local bend angle at each step is mostly due to a change in the flexible roll compared to tilt angle. In the regime of global bend angles of up to 50° one can also see a clear pattern of larger contributions at pyrimidine-purine junctions (CG, TA, CA steps) compared to purine-pyrimidine steps (e.g. GC, AT). This pattern can also be clearly seen at the central region of the [Atract-2] duplex which contains CG and GC steps at the center but not for the [Atract-1] duplex. Even for bending restraining angles in the 50-100° regime the AA steps at the center of the [Atract-1] show only small local bend angles (Fig. 4.14). This results points towards a principally different bending mechanism for the [GC] or [TA] case vs. the [Atract-1] bending. In fact, weak bending of the [Atract-1] is directed towards the minor groove at the center with contributions towards the major groove mainly at the junction to the flanking GC-rich sequences (Fig. 4.15). As can also be seen in Fig. 4.15, in case of representative simulation snapshots (i.e. time conformations) of weakly bent [GC], [AT] and [Atract-2] duplexes the bending occurs preferentially towards the major groove. It agrees with previous observations of an intrinsic bending of A-tract sequences towards the minor groove (99,213).

Strong bending and analysis of base pair kinking

At larger restraining bend angles (80-150°) stronger localized kinks at individual base pair steps were observed. In recent molecular dynamics simulations of short DNA minicircles (108,126) two types of double helix kinking were observed. Type I kink consists of complete unstacking of two consecutive base pairs creating a local bend of more than 90° and opening the major groove. It is characterised by a high negative roll between the two base pairs involved. Type II kink is due to disruption of the hydrogen bonds of the central base pair and stacking of these bases on their 5' neighbours. This second motif reveals very large propeller (roughly 120°) and stagger but intact flanking base pairs (illustrated in Fig. 4.16).

In our simulations, which use the recent parmbsc0 force field a full kinking according to the type I kink never occurred. In order to verify that the difference comes from the force field we analysed simulation of the [GC] case employing the parm94 force field and found that indeed in those simulations type I kinks occurred. It was shown (122,123) that the parm94 force field used in previous simulations (108,126) promotes non-canonical transitions of the α/γ backbone dihedral angles (towards the trans states) and this could enhance sampling of conformations with type I kink (since the angle γ is in conformation trans in this motif (108,125)). With parmbsc0 force-field, incomplete and reversible base pair unstacking frequently occurs (shown in Fig. 4.17). In particular a kink with mean roll amplitude in the range of 20°-50° and directed toward the major groove (opening the minor groove) emerges as a frequent and slowly relaxing inter base pair degree of freedom.

Type II kinks, however, can be seen in the current study at base pairs G₆C₇G₈ in [GC] for bending higher than 130° and for A₉A₁₀C₁₁ in [Atract-1] beyond 120° bending (illustrated in Fig. 4.17, see also Fig. 4.14). In the case of the GCG motif the average propeller is +60°, the kink closes the local major groove and often induces bifurcated hydrogen bonds in the triplet (Fig. 4.17b). In the AAC triplet (Fig. 4.17a), where the minor groove is closing, the propeller value reaches +120°, similarly to that found in (108) at AGG. Type II kinking induces both roll and tilt rotations of neighbouring base pairs. A linear axis fitted with Curves to four base pairs on each side of the motif indicates the local bend for type II kink, 91.3° in (GCG) and 88.6° for (AAC). Fig.4.17 shows also other alternative kink motifs observed in our study, notably, an adenine flipped into the major groove at A₁₂T₁₃ in [AT], (Fig. 4.17c), and a small kink towards the major groove, corresponding to roll of 45° at C₅G₆ in [GC] (Fig 4.17d).

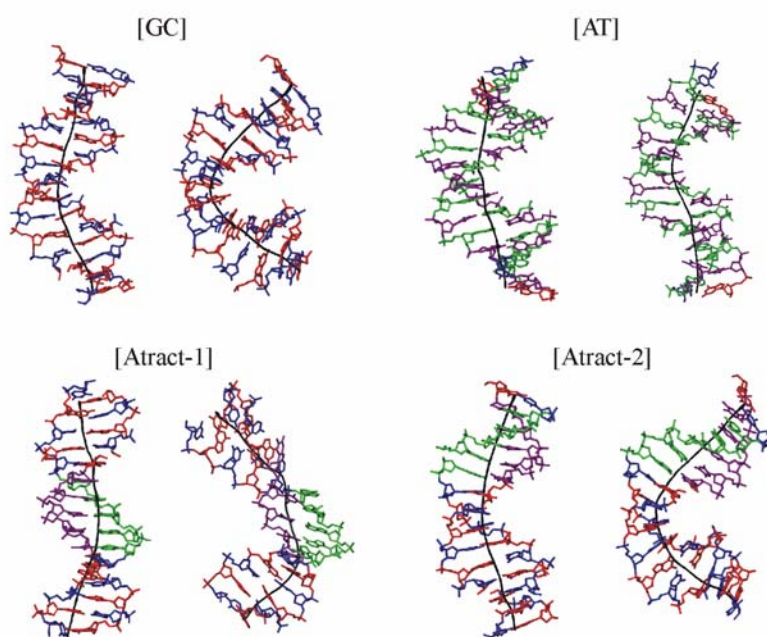


Figure 4.15. Representative conformational snapshots of the four duplexes for weak bending ($\sim 50^\circ$, left panels for each duplex) and strong induced bending ($>100^\circ$ except 80° for [AT], right panels for each duplex). The nucleotides are colour coded (cytosine in blue, guanine in red, thymine in purple and adenine in green). The global helical axis calculated using Curves (33, 34) is shown in black.

A-tract bending with parmbsc0

As already indicated above in the analysis of weak DNA bending AA dinucleotides located inside the A-tract show particularly low fluctuations in junction bending (mainly roll) up to 120° bending for [Atract-1] and up to 150° for [Atract-2], as was already observed in our simulation with parm94 force field. At 120° bending for [Atract-1] a transition occurs, which fundamentally alters the pattern of helical parameters. The first three AA junctions of the A-tract become rolled by roughly 25° toward the major groove. We find thus a threshold in bending amplitude beyond which the classical mechanism of A-tract bending, corresponding to smooth and phased local bends at specific junctions (as described before in the current chapter), rearranges into a different pattern. It consists in soft kinks directed toward major groove on the 5' side of the A-tract and a type II kink at its 3' junction, closing the minor groove.

It is of interest to note that with the A-tract block shifted away from the oligomer center, [Atract-2], bending arises from rolls of small amplitude directed toward major groove at the central junctions (GC type), and toward minor groove at the 5' side of the duplex. Comparing the central sequence of [GC] and [Atract-2] shows that an important difference is one junction CG in the former being replaced by CA in the latter. The CA junction should be easier to unstack since the stacking energy for CA is relatively high compared to CG (26). Therefore

no kink II is created and bending of [Atract-2] remains delocalized along the sequence and energetically less costly.

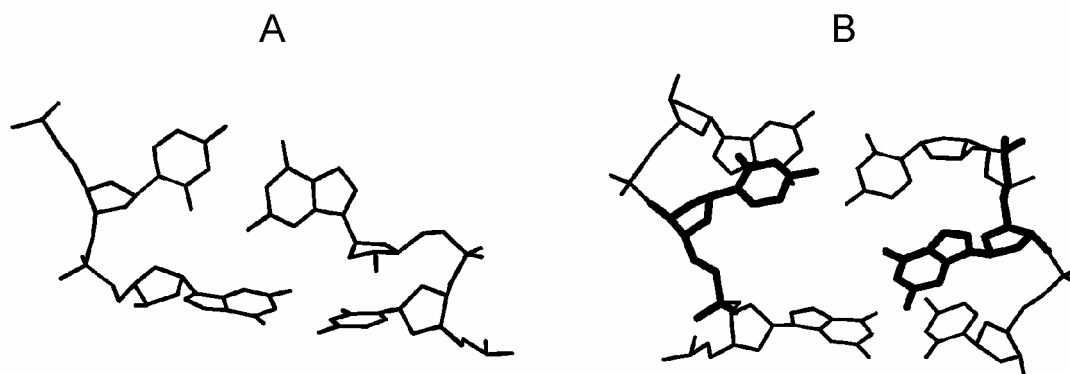


Figure 4.16. Illustration of type I (A) and type II (B) kinks in DNA as defined in (108). (A): A *direct* unstacking of two consecutive base pairs produces a bend of ~ 90 degrees and opens the major groove. (B): A local bend angle with approximately the same amplitude as in (A) is produced by a trinucleotide motif involving disruption of hydrogen bonds in the central pair of complementary bases (in bold). Shown is an example of type II kink motif opening the DNA *minor* groove.

Helical parameters and phosphodiester backbone

It is interesting to see that the kind of soft kink (local bend angle between 20 - 50°) occurred not only at pyrimidine-purine junctions (CG, TA, CA) but also at some purine-pyrimidine ones (GC, AT) as well as AA. An inverse correlation of roll vs. twist (70,248,249) was systematically observed for kinked base pairs steps. Anisotropic DNA bending toward the grooves (250) was also systematic, i.e. tilt parameter close to zero except for type II kink junctions. Hysteresis in helical parameters can explain hysteresis of the free energy curves (as mentioned about Fig.4.13 despite not shown). Most important are the opening of base pairs in the first three base pairs of the [AT] oligomer (causing local strand separation) as well as (roll/tilt/propeller) parameters at junctions making-up type II kink motifs in [GC] and [Atract-1] oligomers. This explain why the only oligonucleotide not showing energetic hysteresis is [Atract-2] for which high bending does not induce type II kinks or strand separation.

We found no coupling between strong base pair bends and BI/BII backbone transition suggested in (116,120,251). We observed, however, some sequence effect, with a higher population of BII states for the junctions GC. Backbone dihedral angle γ (C4'-C5') in trans

conformation were predicted in (125) and observed in (108) to be associated with type I kink and in (127) with A-tract curvature at its 5' CA junction. Such transition is not observed in our study, in agreement with the absence of type I kink.

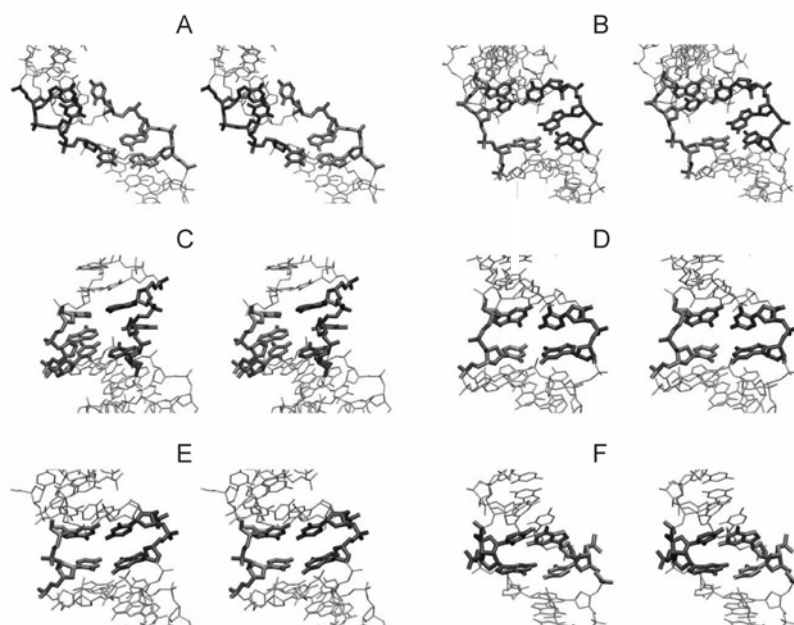


Figure 4.17. Stereo view of kinks in DNA observed during MD simulations. The central nucleotides around the kink site are enlarged. (A) Type II kink at A₉A₁₀C₁₁ in [Atract-1], directed towards the minor groove, (B) Type II kink at C₆C₇G₈ in [GC], directed toward the major groove, (C). Flipped base at A₁₂T₁₃ in [AT] (towards the major groove), (D) Small kink (roll of 45°) towards the major groove at C₅G₆ in [GC], (E) Small kink (roll of 20°) towards the major groove at C₅A₆ in [Atract-1], (F) Small kink (roll of 20°) towards the minor groove at C₁₁G₁₂ in [GC].

Harmonic analysis of base pair hinges

Bending stiffness constants for base pair junctions along the sequence are shown as a function of global bending in Fig.4.18. Quasi-normal distribution of the amplitude of the hinge angle ($a_{i,j}$) for every base pair junction was checked by several techniques, i.e. individual $a_{i,j}$ density histograms, quantile-quantile normal probability plots, Chi-Square and Shapiro-Wilk statistical tests (not reported for clarity). A quasi-harmonic analysis of hinge angles (112,114,240,241) was then performed in each window (see method section).

The force constants obtained from this analysis, presented in Fig. 4.18, show several interesting features. Firstly, for all oligomers the force constants decrease with increasing global bending. When out of the relaxed structure the further deformation seems to be easy. This leads to the linear dependence of the free energy on the bending angle we obtain (Fig. 4.13) and which was also observed experimentally. For relaxed structures and for the [GC] oligomer both types of junctions, GC and CG have similar hinge force constants around 0.03

kcal/(mol*deg²). In contrast, for [AT] the two types of junctions are clearly distinguishable, the AT hinge constants can go as high as 0.06 kcal/(mol*deg²) and the TA junction values group around 0.02 kcal/(mol*deg²). In the two Atracts it is easy to see the two components, GC lower and AA junctions higher.

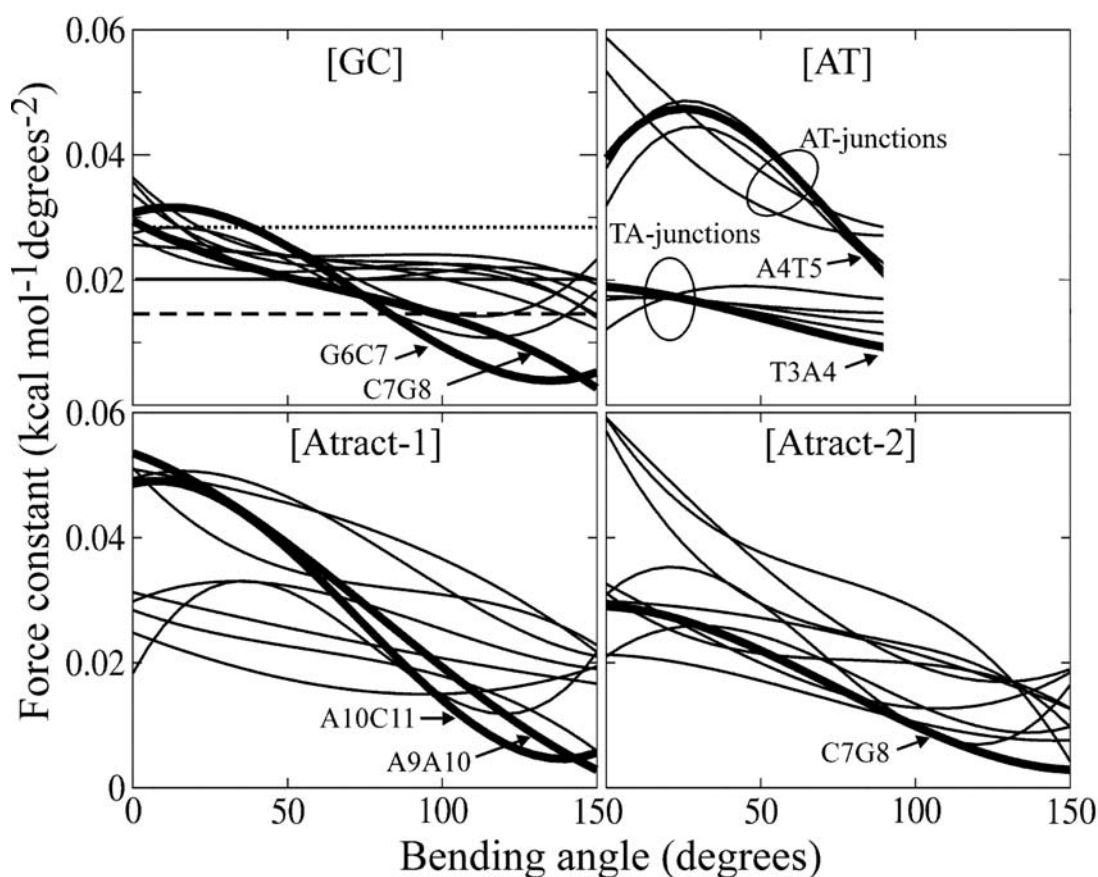


Figure 4.18. Bending force constants k_{ij} of each junction for the four oligomers for each simulation window. k_{ref} is obtained from a harmonic approximation to the junction hinge angle defined as $\sqrt{(\text{roll}^2 + \text{tilt}^2)}$. The results have been smoothed using a cubic polynomial. Junction force constants discussed in the text are highlighted in bold. For CpG several values from other studies are indicated by a straight horizontal line: thermal stability data (252) in dots, EPR (253) in dashes, x-ray crystallography (114) and unrestrained molecular dynamics simulations (241) in bold. For [AT] the force constant curves are not shown beyond 80°.

Secondly, there are few junctions which reduce their force constants dramatically, mainly AT and AA steps. In the case of [GC] these are the two junctions of the kink II G6-C7 and C7-G8. In the same time the junctions C5-G6 and G10-C11 increase their force constants starting from 120°-130° and return to values corresponding to low bending (range of values inferred from experiments (114,252,253) and unrestrained MD simulations (241)). The oligomer [AT]

presents a different picture. TA junctions keep their constants rather flat with a small increase in the end. The lowest of the TA junctions, T3-A4, denatures in the last 4 windows. The next junction A4-T5 shows in contrast severe variation since it drops off dramatically between 30° and 130° as a result of its kinking transition. Other junctions show an S shape i.e. A8-T9 and A12-T13. In [Atract-1] the two descending lines correspond to the kink II junctions A9-A10-C11 and the S shaped to G4-C5. Most interestingly, the central junction A₉A₁₀ shows the highest reduction. Finally, in [Atract-2] the AA junctions loose their stiffness very steeply but the strongest decrease in stiffness corresponds to a junction CG. This relates to the wide distribution of mini-kink motifs discussed previously.

In summary the two junctions of type II kink in [Atract-1], A₉A₁₀ and A₁₀C₁₁, present a roughly 10 fold reduction in their stiffness constant. This is partially attributable to the known initial rigidity of these DNA junctions in a run of consecutive Adenines. Indeed this reduction is only half an order of magnitude for junctions G₆C₇ and C₇G₈ of type II kink in [GC]. This correlate with a lower value of the linear coefficient of free energy obtained for [Atract-1] compared to [GC].

Investigation by Replica Exchange Umbrella Sampling

Umbrella Sampling was also carried out using the Replica-Exchange Umbrella Sampling (REUS, 15,16,17) method in case of the [GC] oligomer, for both Amber force fields parm94 and parmbsc0. REUS was used to address the (non-) reversible nature of DNA base pair kink motifs discussed in this chapter. In particular the type I kink was also observed in current study using conventional umbrella sampling on the [GC] oligomer and the same force field as in (108), parm94. Thus the bending of [GC] is explained by a type I kink with parm94 but by a type II kink with parmbsc0 and it is of great interest to see if the greater variety of accessible conformations in phase space provided in theory by the REUS sampling will change this conclusion. For a description of the REUS method I refer to the section 2.9.2 of the introduction and more particularly to the section “Method” of chapter 3 on the REUS approach, since the protocol is similar to here except for the use of a different umbrella coordinate. As shown in Fig.4.19 the type I kink does not show the property of a reversible (elastic) deformation in the restrained sampling windows of the REUS ansatz. Indeed type I kink does not occur but is substituted by a more homogeneous distribution of bending strain along the DNA sequence. Moreover the free energy curve calculated with WHAM from the REUS simulation (Fig. 4.19) indicates that a transition between two distinct regimes in the

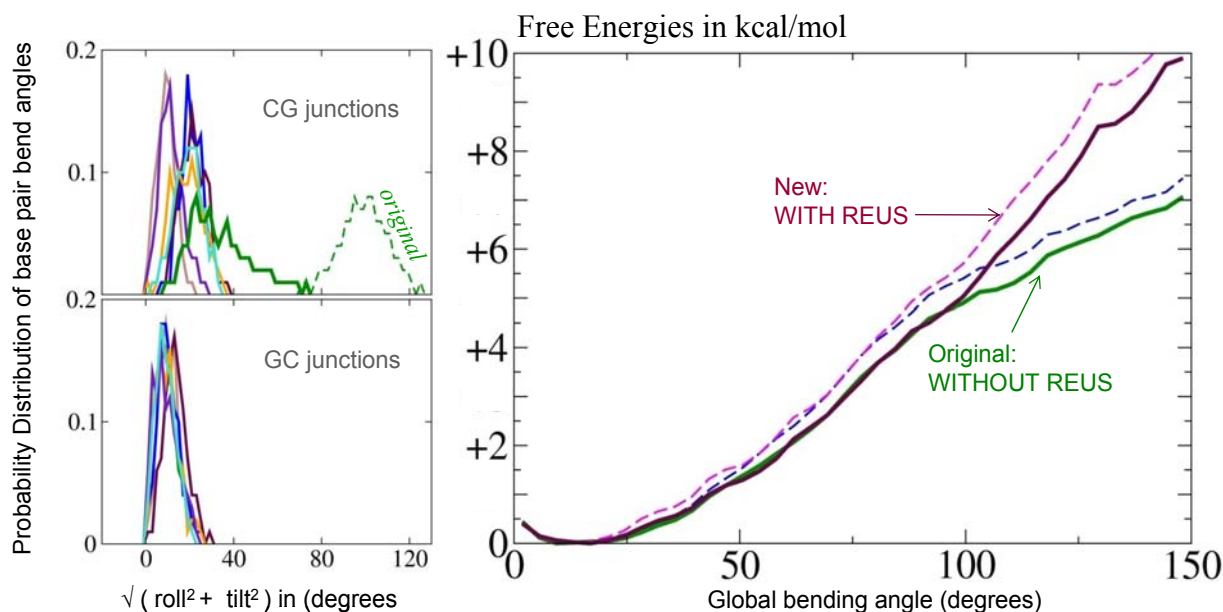


Figure 4.19. Analysis of Replica Exchange Umbrella Sampling on [GC] using the Amber parm94 force field. At the left: Base pair kink density histogram observed for high DNA bending strain (150°). The green curve (full line: REUS; dashed line: original Umbrella Sampling) corresponds to C11G12. At the right: Comparison of the Potential of Mean Force calculated from the conventional Umbrella Sampling where type I kink occurs and from the REUS where more homogeneous distribution of bending strain along the sequence substitutes the occurrence of a severe kink at C11G12. The PMF curves were calculated after 1 ns (dashes) and 2 ns in the Umbrella window at 150° of bending.

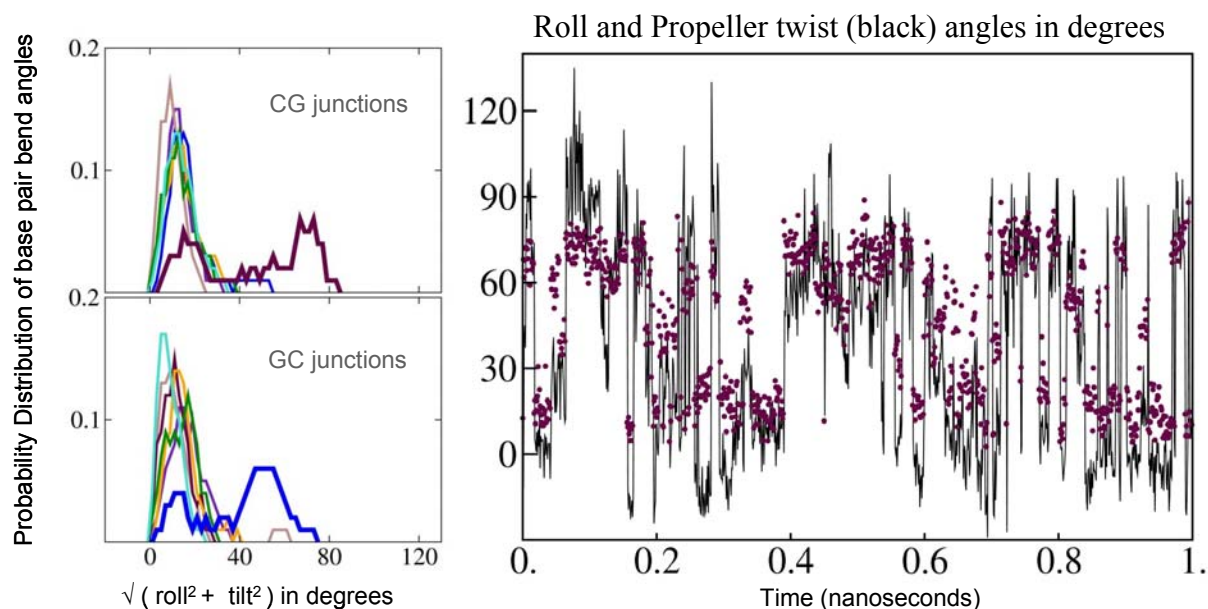


Figure 4.20. Analysis of Replica Exchange Umbrella Sampling on [GC] using the Amber parmbse0 force field. At the left: Base pair kink density histogram observed for high DNA bending strain (150°). The purple curve corresponds to C7G8 and the blue curve to G6C7. At the right: Time series of the

roll angle at C7G8 (purple) and propeller twist (black) at the base pair C7:G24 (window 150°). Note C7:G24 is the base pair shared by the C7G8 and G6C7 base pair steps.

PMF observed in the conventional umbrella sampling does not occur in the REUS simulation. This PMF transition marked the regime where type I kink occurred in the conventional umbrella sampling. Thus the comparative analysis of Fig.4.19 using parm94 force field suggest that type I kink is a plastic deformation due to some phenomena of strain energy localization (254-256) and not a significant DNA propensity at the regime of bending probed in current simulation (up to 150°).

As shown in Fig.4.20 the type II kink does show property of a reversible (elastic) deformation in the restrained sampling windows of the REUS ansatz with parmbsc0. First the base pair bend density histograms appear as bi-stable distributions for the two junctions that form a type II kink in [GC] (C7G8 in purple and G6C7 in blue). This means that these two neighbour junctions are both bent during a significant part of the trajectory as in conventional umbrella sampling and are both unbent during another significant part of the same trajectory (the window at 150° of bending is shown). Second, the graphics on the right hand side of Fig.4.20 looks at the bend angle of the C7G8 junction and at the propeller twist of the base pair intermediate between G6C7 and C7G8 (C7:G24) as a function of time. The former time-series indicates a number of transitions between large bend angles (~60°) and small bend angles (~15°) corresponding to the two sub-maxima on the density plot (Fig.4.20, left) for C7G8. Moreover these transitions occur in concert with propeller transitions (time series in black curve) between usual values and unusual values specific to the type II kink. Thus the simulations using parmbsc0 force field suggest that type II kink is an elastic deformation. Clearly type II kink appears as a significant DNA propensity at the regime of bending probed in current simulation (up to 150°) since it is a reversible motif that returns spontaneously to its most stable conformation under equilibrium fluctuations.

4.5 Discussion

A new restraint to induce DNA curvature, based on the definition of an average screw axis for the terminal segments of DNA oligo-nucleotides has been developed. In principle, DNA bending could also be induced by simple distance or angle restraints involving selected atoms

at both ends and in the middle of a DNA fragment. However, application of such restraints can result in significant distortion of the local structure near the atoms involved in the restraining coordinate and, depending on the choice of atoms, create a directional bias to bending. The present average screw axis definition is based on the local coordinate frames of each nucleotide of the terminal fragments. This definition allows an optimal distribution of the bending restraining force on the whole terminal segments and all DNA conformational parameters can distribute optimally according to the specific base sequence, without disturbing the helix locally. The method allowed us to systematically investigate the bending deformability of various DNA molecules without any restrictions on the conformational relaxation of the DNA upon addition of the bending restraint.

In addition to free relaxation upon induced bending the algorithm was extended to add a restraint on the bending direction and to rapidly create a 2D DNA-bendability map. Note, that only such approach allows a systematic exploration of the sequence dependent bendability of DNA in contrast to 1D approaches or free simulations that may suffer from sampling only one or a limited number of bending directions. Interestingly, pre-bending in various possible directions followed by removal of the restraint on bending direction resulted in axial rotation towards a preferred bending direction indicating a significant bending anisotropy of DNA.

The method is applicable not only to DNA but also to other chain molecules to investigate global bending properties. This includes possible applications to study RNA and the influence of non-helical elements that frequently interrupt base-paired regions in RNA on global and directional bending properties. For example, the ribosome consists of many RNA helices connected by non-helical elements. Understanding the global deformability of these elements is critical for an understanding of the ribosome function. Other application examples are the bending and kinking flexibility of long helical segments in proteins that are often of functional importance for the protein molecule.

Results on A-tract containing sequences and the Hagerman H1 sequence predict preferential bending towards the minor-groove and resulted in helical parameter changes in good agreement with experiment. In addition, the direction and magnitude of bending was in very close agreement with experimental solution structures. For the (TA)₆ sequence the calculations predict no preferred bending towards the minor groove but a greater bendability towards the major groove relative to the other DNA molecules under study. This is compatible with the preference of TA containing segments at the center of the recognition sequences for minor groove binding proteins that need to bend towards the major groove in order to open the minor groove. The energetic preference of bend vs. straight conformations

was of the same order or smaller than the average energy per degree of freedom at room temperature for H1 and A-tract sequences. This small energetic preference indicates that DNA is in a dynamic equilibrium with many different states and a sequence-specific distribution of possible bend angles.

Significant changes in bending magnitude and preferred bending direction (compared to regular DNA) were found for duplex DNA with central mismatches. In particular for a central G:A or central C:C mismatches the calculations predict a significantly enlarged variety of possible structures that showed different directions and magnitudes of global bending compared to B-DNA. The altered bending flexibility may contribute to the recognition properties of modified DNA by repair enzymes.

The approach was extended to study DNA bending during molecular dynamics simulations including surrounding ions and water explicitly. It should be emphasized that this approach goes significantly beyond free MD simulations of DNA where at most bending flexibility close to the equilibrium state can be investigated. Our approach allows for the first time to induce smooth global bending in a DNA fragment during an MD simulation and to directly extract the associated change in free energy. The calculated bending free energy profile for an A₅-tract containing oligo-nucleotide with the standard Amber force field used up to now, parm94, shows an optimal bending angle close to experiment and the corresponding minimum obtained from the adiabatic mapping calculations. However, the calculated free energy penalty for strong DNA bending was calculated significantly smaller for the explicit solvent free energy simulations than for the adiabatic mapping. The helical parameter changes that occurred upon induced global bending showed good agreement between adiabatic mapping and MD as well as with experimental data from a NMR structure of a similar A-tract DNA. The global bending flexibility estimated in terms of a persistence length was in close agreement with experiment for the free energy simulations whereas the bending rigidity obtained from adiabatic mapping was much larger. The simulations with parm94 predict that the mechanism of DNA bending is similar at higher bending angles that go beyond the equilibrium bending angle for the A-tract DNA (bending of 100°). Another interesting finding of the simulations with parm94 is that although there is some coupling of the global bending with other global variables characterizing the DNA, the fluctuations of these global variables do not change significantly under bending stress. The rigidity of the central A-tract (in terms of groove width, twist and rise fluctuations) was found to be very similar for relaxed and significantly bent DNA.

It should be noted that the current method is mainly limited by the possible incomplete sampling of conformational substates that are compatible with a given bend DNA conformation. In case of the EM adiabatic mapping approach only one conformational state (nearest local energy minimum) is considered for each bending angle (and direction). In case of the MD umbrella sampling approach several substates compatible with a bend conformation can in principle be sampled but due to barriers between the substates it may not be possible to achieve an exhaustive exploration. The good qualitative agreement between EM and MD results for the A-tract DNA indicates, however, that the EM adiabatic mapping method may allow for a rapid and systematic exploration of many other DNA sequences (and possibly other types of biomolecules such as RNA) to get at least qualitative insight into the bending deformability.

By controlling an induced bending of short DNA oligomers during Molecular Dynamics simulation with the state-of-the-art Amber force field parmbsc0, we calculated the bending free energy of four different oligomers, including the Atract system discussed above, and analyzed the molecular details of the bending mechanism. The energy change required to form a bend angle of 90° on 5 nm length scale is close to 5.0 kcal/mol for A-tract sequence and 6.0 kcal/mol for alternating [GC] or [AT] sequence which agrees with recent calculations using the previous series of Amber force-field (80). For bending near 150° the free energy cost is ~ 13 kcal/mol for all sequences with the exception of [A-tract-1] where it is reduced to 10.0 kcal/mol. We have found a linear (non-harmonic) energy dependence and extracted linear coefficients for different base sequences which can be used in Sub-Elastic Chain models (4,72).

The structural analysis of oligomers has shown that in the weak regime curvature is distributed smoothly over the whole oligomer. For the strong bending ($> 100^\circ$) the oligomer behaviour depends on its sequence. [GC] and [Atract-1] produce the type II kink, [AT] denaturates and [Atract-2] responds to the imposed stress by a series of small kinks in the GC part of the sequence. We have not observed the type I kink (parmbsc0 results) most probably due to the recent modifications of the force field aimed at reduction of over-stabilisation of the α/γ backbone transition, characteristic for the older series of amber force field parameterisations. The type II kink was observed twice, in [GC] and [Atract-1] oligomers. Type II kink in [GC] is associated with local bend toward major groove and smoother base pair deformation than in the case of [Atract-1] where it induces local bend towards the minor groove. Both are associated with a kink angle of roughly 90° between linear axes fitted to four base pairs on each side of the motif. Using the REUS sampling technique introduced in

chapter 3 it was demonstrated that the type II kink motif is an elastic and thus reversible DNA motif.

We found that bending pathways of the oligomers with A-tract depend on its position which could be used functionally in vivo through helix phasing of DNA A-tract segments. Harmonic bending stiffness moduli (112,114,240,241) are commonly used in the elastic rod theory to model large scale dynamics of DNA molecules (44). In light of recent experimental measurements of DNA flexibility on short length scale (4-6,12) this harmonic treatment has been questioned. It was suggested that sequence dependent DNA coarse-grained parameters may have to include non linear coefficient accounting for rare DNA elastic breakdown (71-73,108,222). The bending stiffness constant in regime of strong bending calculated in this study could be used to refine constitutive relation of elastic rod models for DNA. Our results show that in short oligomers individual junctions behave differently, in function of their sequence context. DNA base flipping, which was proposed as one main origin of the increased flexibility of DNA on short length scale (224,225), were only transient. Finally, there were no significant transitions in backbone torsion angles associated with DNA bending.

Acknowledgements

Part of these simulations were performed using computational resources of the Computational Laboratories for Analysis, Modeling and Visualization (CLAMV) of Jacobs University in Germany and of the Laboratoire de Bioinformatique et RMN structurales of the Institut de Biologie et Chimie des Proteines (IBCP) in France. Fundings were provided by a VolkswagenStiftung PhD grant to my supervisor, Martin Zacharias, and by the Université Franco-Allemande (UFA), co-tutelle agreement between the Université Denis Diderot in Paris (co-supervision by Krystyna Zakrzewska) and the Jacobs University in Bremen. I thank Richard Lavery for helpful discussions and comments on the subject of this chapter.

Chapter V

Flexibility of the phosphodiester backbone: metastable ϵ/ζ conformations in severely bent and/or damaged DNA

- . **Keywords:** Replica Exchange; Nucleic Acid Backbone Dihedrals; (ϵ - ζ) BI/BII substates.
- . **Reference publications:** Curuksu, J., Zacharias, M. (2009) *J. Chem. Phys.* (submitted).
- . **Recommended Introductions:** 1.4.3, 2.7, 2.9, Chapter IV.

5.1 Abstract

Although molecular dynamics (MD) simulations have been applied frequently to study nucleic acids the sampling of possible states separated by barriers is limited due to the time scale limitation of current MD simulations. Replica-exchange (Rex)MD simulations that allow for exchanges between simulations performed at different temperatures (T-RexMD) can achieve improved conformational sampling. However, the T-REMD method is limited to small systems because the computational demand grows rapidly with system size. A Hamiltonian RexMD method that specifically enhances coupled dihedral transitions in nucleic acids or other molecules has been developed. The method employs added biasing potentials as replica parameters that destabilize available dihedral substates. The biasing potentials can be either fixed at the beginning of the simulation or optimized during an equilibration phase. The method was extensively tested and compared to conventional MD simulations and T-RexMD simulations on an adenine dinucleotide system and on a DNA abasic site. The biasing potential (BP)-RexMD method showed improved sampling of conformational substates compared to cMD simulations similar to T-RexMD simulations but at a fraction of the computational costs. It is well suited to study systematically the fine structure and dynamics of nucleic acids and can be easily extended to other types of molecules.

5.2 Introduction

In principle molecular dynamics (MD) simulations allow studying the fine structure and dynamics of nucleic acids and other biomolecules at atomic resolution and under realistic solution conditions including explicit solvent and surrounding ions. However, the efficiency of MD simulations is often limited by poor sampling of possible states due to high energy barriers that separate low-energy regions of a biomolecule (160).

The parallel tempering or replica exchange molecular dynamics (RexMD) method is a widely used to enhance conformational sampling in molecular simulations (13,257-264). In RexMD simulations, several copies (replicas) of a molecule are simulated independently and simultaneously using classical MD or MC methods at different simulation temperatures (or force fields: Hamiltonians). At preset intervals, pairs of replicas (neighbouring pairs) are exchanged with a specified transition probability. In most RexMD simulations the temperature is used as a parameter that varies among the replicas (T-RexMD). The random walk in temperature allows conformations trapped at a low simulation temperature in locally stable states to escape by exchanging with replicas simulated at higher temperature. Efficient exchange between neighbouring replicas requires overlap of the potential energies sampled at neighbouring simulation temperatures. As a consequence, the number of required replicas grows approximately with the square root of the number of particles in the system (to cover a desired temperature range). In turn, a larger number of replicas requires also longer simulation times to allow efficient “diffusion” of replicas in temperature space.

Instead of using the simulation temperature as a replica coordinate, it is also possible to use the force field or Hamiltonian of the system as a replica-coordinate. In H-RexMD simulations (14-17, 197, 199-202, 265-269) it is possible to modify only part of the Hamiltonian along the replica simulations. The advantage is that exchanges between replicas are then independent of the part of the Hamiltonian that does not differ between replicas. Consequently, such approaches may require much fewer replicas for efficient sampling compared with T-RexMD (202).

RexMD methods have been applied frequently to study peptides and proteins (13-15, 17, 197, 199-202, 257-261, 264-269) but much less for studying the structure and dynamics of nucleic acids (e.g. 16, 262, 263). In case of studying double stranded (ds)DNA or RNA a major problem to apply the standard T-RexMD methodology is the possible rapid dissociation and ion redistribution at elevated simulation temperatures. The relatively fast dissociation into

single or partially single stranded chains can lead to the sampling of conformational states that are largely irrelevant for the analysis of sub-states of dsDNA and dsRNA. Conformational transitions in dsDNA or dsRNA often require coupled transitions of backbone dihedral angles to result in conformations compatible with the helical geometry of the molecule.

Here, a new biasing potential (BP)-RexMD method is presented that allows efficient exploration of DNA and RNA fine structure. The RexMD method employs different levels of specific biasing potentials to promote coupled transitions of nucleic acid backbone dihedral angles along the replicas. The sampled conformations can exchange with a reference replica that is controlled by the original force field. The choice of biasing potentials in the replicas can be either fixed at the beginning of the simulation or optimized during an equilibration phase. The method was extensively tested and compared to conventional MD simulations and T-RexMD simulations on an adenine dinucleotide system and on an abasic site located at the center of a DNA duplex. For both systems the new BP-RexMD method showed improved sampling of nucleic acid substates compared to cMD simulations, similar to T-RexMD simulations but at a fraction of the computational costs. Our results on the DNA abasic system show that the BP-RexMD method is necessary to correctly describe the (ϵ / ζ) metastable conformations on short simulation time-scale. Implication on the globally induced flexibility is discussed in term of coupled changes with the local bending angle.

Since the method does not involve simulations at high temperatures the problem of strand dissociation can be avoided. Since the biasing potential focuses on dihedral transitions and does not involve other solute-solvent or solvent-solvent interactions it is expected to require much fewer replicas to achieve efficient sampling of substates in large nucleic acids compared to T-RexMD. A system-specific flooding potential can be designed in a pre-equilibrating phase and subsequently used for Production and thus the method is directly applicable to other biomolecules.

5.3 Methods

Backbone Biasing Potential Replica Exchange (BP-Rex) MD for Nucleic Acids

During replica exchange molecular dynamics (RexMD) simulations exchanges of conformations between independent MD simulations (replicas) are attempted at preset time intervals. Each replica runs at a different temperature (T-RexMD) or under the control of a

different force field (Hamiltonian: H-RexMD). Exchanges between replicas labelled i and j are accepted or rejected based on the usual Metropolis criterion (198),

$$\begin{aligned} \omega(C_i \leftrightarrow C_j) &= \min(1, \exp(-\Delta)), \\ \Delta &= \beta_i(U_i(q_j) - U_i(q_i)) - \beta_j(U_j(q_j) - U_j(q_i)) \end{aligned} \quad \text{eq.5.1}$$

where U is the potential energy of the system and β the inverse thermal energy $1/k_B T$.

The acceptance ratio decreases exponentially with $(\beta_i - \beta_j)$. In practice exchanges are thus attempted only between pairs of replica which must show overlapping potential energy histograms for the acceptance probability to be significant. Instead of the temperature it is also possible to use the Hamiltonian as replica coordinate.

The advantage of H-RexMD is that it is possible to consider the Hamiltonian as a sum of a replica dependent term v and replica independent term $(U - v)$ and to use the same temperature ($\beta_i = \beta_j$) for every replica. One thereby define a “parameter” space of replica exchanges for which the acceptance ratio can be significantly increased compared to T-RexMD since, $\Delta = \beta(v_i(q_j) - v_i(q_i) + v_j(q_i) - v_j(q_j))$, is determined only by the part of the Hamiltonian that depends on the replicas.

The nucleic acid backbone adopts preferred states that are separated by large energy barriers. In order to allow effective transitions among possible substates we have designed a nucleic acid dihedral angle dependent biasing potential that specifically destabilizes a range of dihedral angles,

$$\begin{aligned} v_i(\tau) &= k \times \\ &\begin{aligned} &((\tau - d_2)^2 - (d_1 - d_2)^2)^2 && \text{if } d_1 < \tau < d_2 \\ &((\tau - d_3)^2 - (d_3 - d_4)^2)^2 && \text{if } d_3 < \tau < d_4 \\ &(d_1 - d_2)^4 && \text{if } d_2 < \tau < d_3 \\ &0 && \text{otherwise} \end{aligned} \end{aligned} \quad \text{eq.5.2}$$

where τ is the dihedral angle variable of interest. The nucleic acid backbone dihedral (NABD) biasing potential, v_{NABD} , has the shape of a quasi-Gaussian with flat ceiling but does not use the $\exp$ function (Fig. 5.1). The width and height of the potential can be independently chosen (with an appropriate choice of d_1 and d_2 ; d_2 and d_3 ; d_3 and d_4). The reference value is thus the zero derivative part of the potential (distance between d_2 and d_3 , “ceiling”) where $v_i(\rho_c) =$

$E_{max} = k \cdot (d_1 - d_2)^4$. Different levels of the potential (with different values of k or E_{max}) were used along the replicas to destabilize a range of dihedral angles in the replica MD runs. In nucleic acid molecules some of the six NABD variables (see section 1.4.3) are coupled and adopt preferred subsets of combinations (115,117,121-123,251). The two most important coupled backbone dihedral angles are the ϵ (rotation around C3'-O3' bond in C4'-C3'-O3'-P) and ζ (rotation around O3'-P bond in C3'-O3'-P-O5') angles which show the largest variation in folded nucleic acids (115,117,122,123). In the current study a two-dimensional v_{NABD} was employed to destabilize a given ϵ/ζ combination,

$$v_i(\epsilon, \zeta) = k \times \begin{cases} ((x-r)^2 - (R-r)^2)^2 & \text{if } r < x < R \\ (R-r)^4 & \text{if } x < r \\ 0 & \text{otherwise} \end{cases} \quad \text{eq.5.3}$$

where $x = \sqrt{(\epsilon - \epsilon_r)^2 + (\zeta - \zeta_r)^2}$ is the coordinate in the two dimensional ϵ/ζ phase-space defined by taking the center of the ceiling (ϵ_r, ζ_r) as origin (see Fig. 5.1). r is the radius of the ceiling and R the radius of the biasing potential. In order to systematically cover the whole 2D-plane of possible ϵ/ζ combinations a set of 36 regularly spaced positions of biasing potentials was used (spaced by 60° in each of the two dihedral angle dimensions).

In each dimension six different sets of the following biasing potential parameters (d_1, d_2, d_3, d_4) were possible: ($0^\circ, 30^\circ, 60^\circ, 90^\circ$); ($60^\circ, 90^\circ, 120^\circ, 150^\circ$); ($120^\circ, 150^\circ, 180^\circ, 210^\circ$); ($180^\circ, 210^\circ, 240^\circ, 270^\circ$); ($240^\circ, 270^\circ, 300^\circ, 330^\circ$) and ($300^\circ, 330^\circ, 360^\circ, 390^\circ$). The width of each potential was 90° , the tail region of each potential was 30° (the difference between d_1 and d_2 or d_3 and d_4 , respectively). The overlap between tail regions of two neighbouring biasing potentials resulted in an approximately constant biasing potential at the overlap region almost identical to the level at the plateau region (Fig. 5.1). The width of the biasing potentials was slightly larger than the fluctuation of dihedral angles in a given backbone substate in experimental structures or during conventional MD simulations to guarantee that the biasing potential centred close to a sampled substate is effective in promoting transitions (Fig. 5.2).

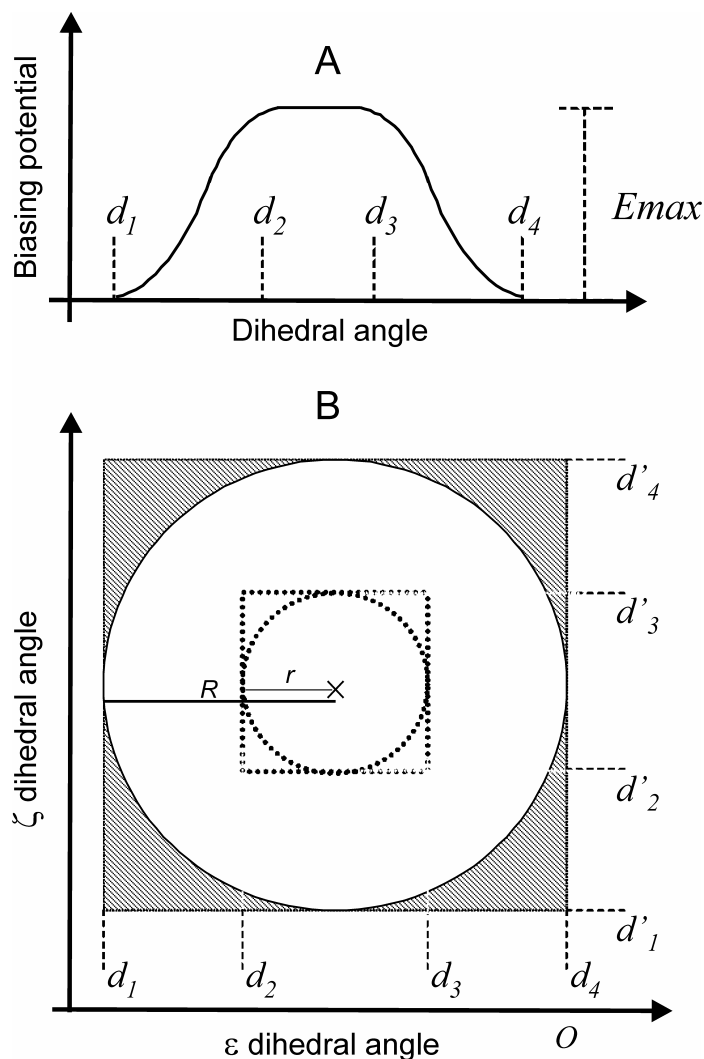


Figure 5.1.

(A) Shape of the dihedral angle biasing potential in one dimension. The potential is constant (E_{max}) between d_2 and d_3 and decreases continuously to zero for dihedral angles in the interval d_2, d_1 as well as d_3, d_4 . Different levels of the biasing potential were added to the force field to control replica runs during BP-RexMD simulations. (B) Shape of the biasing potential in 2 dimensions (ϵ/ζ dihedral angles). The potential is constant for ϵ/ζ dihedral combinations within radius r from the reference (marked by a cross) and decreases smoothly to zero within radius R .

Test simulations on nucleic acid systems indicated that an $E_{max} = 10.0$ kcal/mol was necessary to allow for efficient nucleic acid backbone transitions and it was chosen as the highest biasing level. During replica exchange simulations with five replicas E_{max} levels of 0 (reference replica without biasing), 1, 3, 6 and 10 kcal/mol were used which resulted in an exchange acceptance rate of $\sim 40\%$ for all tested systems. The sampling of conformations within each of the 36 intervals (metavariables) could be monitored by collecting all ϵ/ζ pairs within each interval. The variance of the sampled states within an interval was used to decide on positioning a biasing potential at the corresponding ϵ/ζ location (the mathematical details are given in the appendix II).

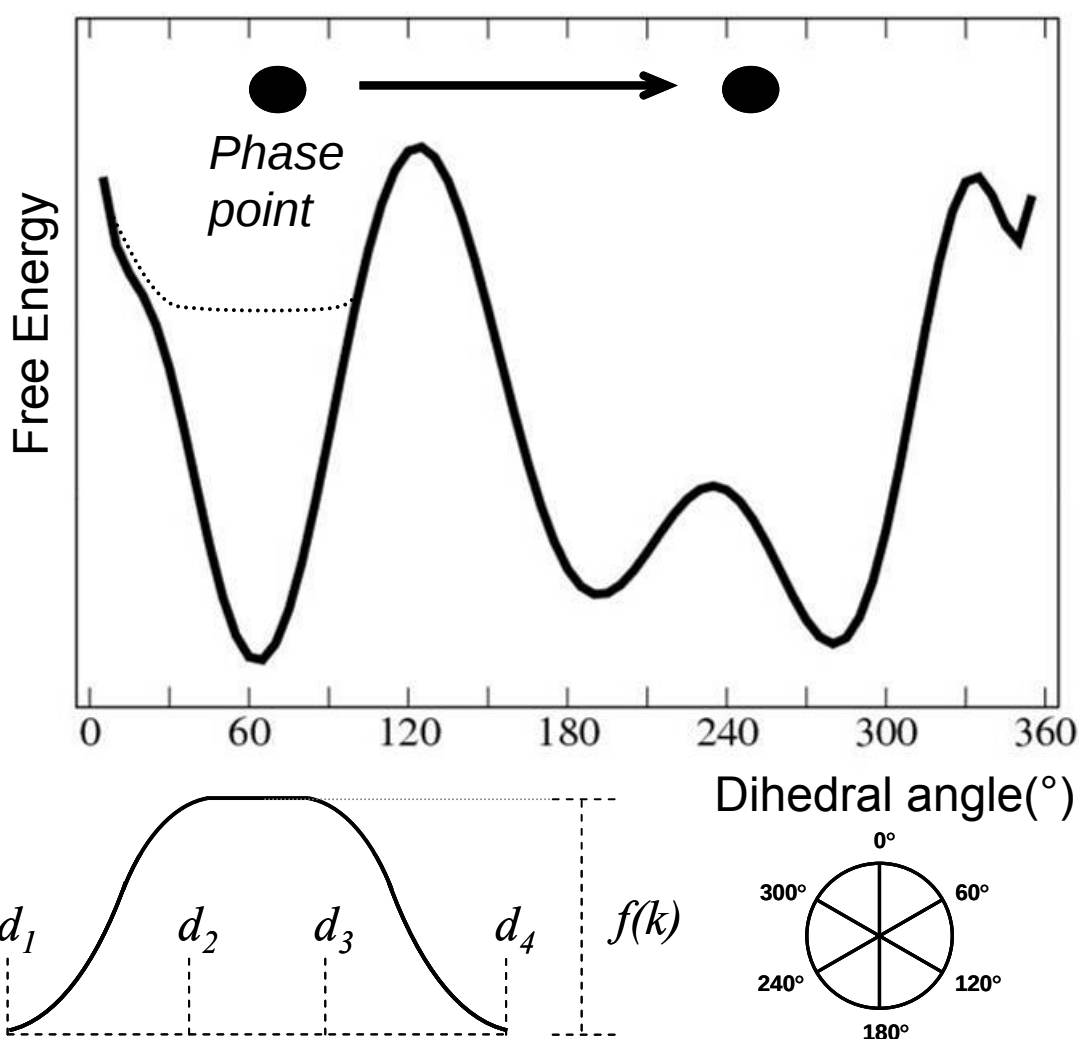


Figure 5.2. Schematic (in one dimension) of the technique used to enhance conformational transition of backbone dihedrals in the biased replicas of BP-RexMD. The 1D model energy surface (bold line) presents three arbitrary wells in *+Gauche*, *Trans* and *-Gauche* respectively. The energetic function is a lifting potential with flat ceiling defined by eq.5.2 / 5.3 and introduced as an additional term in the Hamiltonian function of the scaled replicas. Its position on a given dihedral rotamer can be either fixed in advance in case of ansatz S-BP-RexMD, allocated dynamically on a selected subset of rotamers in case of ansatz D-BP-RexMD, or cloned on a pre-selected subset of rotamers in case of ansatz E-BP-RexMD. In the two latter sampling schemes a communication between the unscaled replica and the scaled replicas is necessary to define this subset of rotamers (using the convergence criterion a_c^i defined in appendix II). In that case the scaled dihedral angle is subdivided into six coarse-grained degrees of freedom (metavariables defined in the text) illustrated in a circular map on the figure. Their convergence, a_c^i , is regularly checked in the unscaled replica. After an equilibrating phase, a specific set of rotamers is selected as “allowable metavariables” for this dihedral angle.

Molecular dynamics simulations

The RexMD method was applied to a desoxy-adenine dinucleotide model system d(ApA) (referred to as AA) and to a more complex DNA molecule that contained a central abasic site (a central sugar phosphate unit missing a nucleobase) with the sequence d(5'ApApGpSpApCpC3'/5'GpGpTpApCpTpT3'). The start structure for the AA system was a canonical B-form DNA dinucleotide step whereas the abasic site DNA (ABA system) corresponded to an experimental structure (pdb-entry: 2HSS, 270).

All MD simulations were performed using a modified version of the *Sander* module of the Amber 8.0 package (203) employing the latest parmbsc0 force field (19). DNA molecules were neutralized by K^+ counter ions and solvated with TIP3P water molecules (191) in a truncated octahedral box (with a 10 Å minimum distance between DNA and box boundary) using the *leap* module. All simulations were carried-out using a 1 fs time step applying periodic boundary conditions (constant temperature) and the particle-mesh Ewald approach (189) with a 9 Å direct space sum cutoff. In order to approximately stay in a helical arrangement similar to double stranded DNA the dApA system was restrained to adopt a B-DNA like structure using four canonical distance restraints between N9, C6, C5 and N3 atoms of each base, respectively, at all simulation stages (i.e. no penalty for atom pair distances within 6-8 Å and a force constant of $0.5 \text{ kcal.mol}^{-1}.\text{Å}^{-2}$ for other distances). Starting AA and ABA solvated systems were energy minimized (1000 steepest descent steps). Subsequently, the solvent and counter ions were allowed to equilibrate for 1 ns at constant volume while adjusting the temperature to 300 K and harmonically restraining the solute to the minimized structure. The solute was then progressively relaxed over a total period of 1.2 ns at constant pressure and temperature, including a final stage of 0.2 ns with no positional restraints. The final structures were used as start structures for all conventional and replica exchange MD simulations. MD trajectories and dihedral angles were analyzed using the *ptraj* module of Amber 8.0. The biasing potential (eq. 5.3) was implemented in the *torng.f* routine of Sander in Amber 8.0. The BP-RexMD and T-RexMD simulations were controlled using a Perl script. During T-RexMD of the AA system the temperature range for the replicas spanned 300-535 K (exponential scaling) which resulted in an acceptance rate for exchanges between neighbouring replicas of 15-40%. In case of the abasic site system a T-RexMD with 16 replicas was performed (T-range: 300K-402K) with an acceptance rate of 15-30% for replica exchanges. See the Annexe II for detailed information on the simulation parameters.

5.4 Results

Comparing sampling efficiency on a dinucleotide system

An adenine (AA) dinucleotide system which contained one pair of dihedral angles ϵ (C4'-C3'-O3'-P) and ζ (C3'-O3'-P-O5') was used for evaluating the performance of BP-RexMD sampling schemes. The starting conformations was in all cases the most common B_I substate (as in standard B-DNA) corresponding to (ϵ/ζ) settled in the (trans/-gauche) regime. Three types of BP-RexMD simulations (each with 5 replicas and for 4 ns) were performed: In one simulation (termed S-BP-RexMD) a single biasing potential centred at the (ϵ/ζ) window that corresponded to the B_I state was employed and various levels of this potential were added in the force fields for the replicas (one reference replica was always controlled by the original force field without added biasing potential).

In the second type of BP-RexMD the placement of added biasing potential was switched to one of the 36 regularly spaced biasing potentials closest to the ϵ/ζ dihedral angle pair that was sampled in the reference replica. This switch was performed after the fluctuation of dihedral angles around the current state was reasonably converged. The convergence of the fluctuation of a given dihedral angle was monitored according to the procedure described in the Methods section. The effect of the “dynamical” allocation of a biasing potential in the replica runs results in destabilization of the dihedral angles sampled in the reference replica and typically resulted in the exploration of states not yet sampled in the reference replica. Here the replicas runs plays the role of some “retailers” ensembles to which the reference replica (Boltzmann sampling) has periodically access. In spirit the smoothing procedure of the energy surface is similar to the meta-dynamics approaches (271,272). In this approach the conformational region sampled during a simulation is continuously destabilized by the addition of small Gaussian (271-274) or Gaussian-like (275) functions centred close to the sampled structure along a relevant degree of freedom of the system. Aim is to completely flatten a free energy surface along some coordinate after many iterations (additions of Gaussians). The canonical distribution can then be recovered by appropriate reweighting of the final trajectory (272-274). In the present approach the exact flattening of the free energy surface (which can be very time-consuming) is not required (or even desired). In contrast to meta-dynamics relatively strong biasing potentials are added at various levels only along the replica runs to enhance transitions (sampling is primarily monitored in the reference replica).

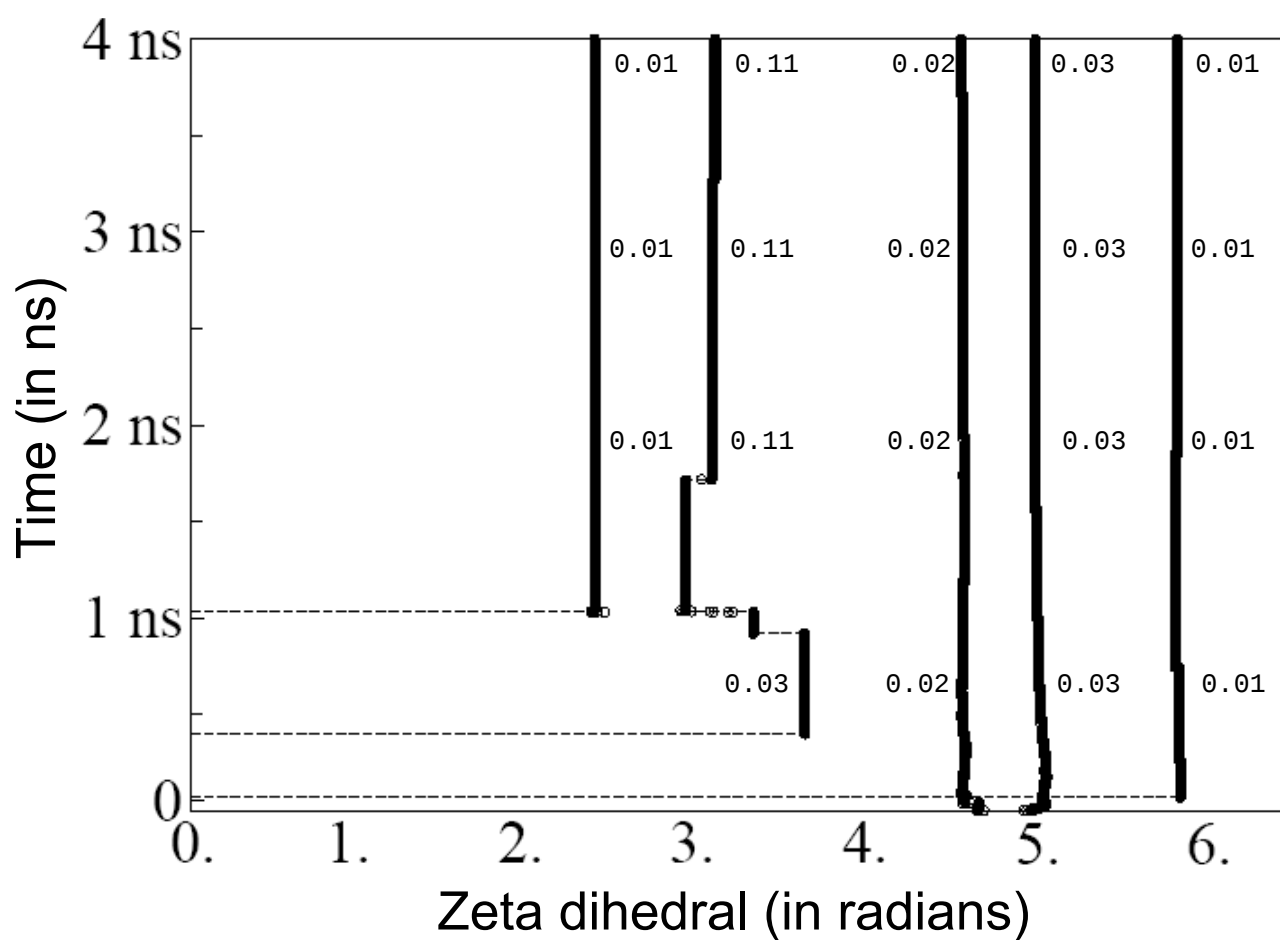
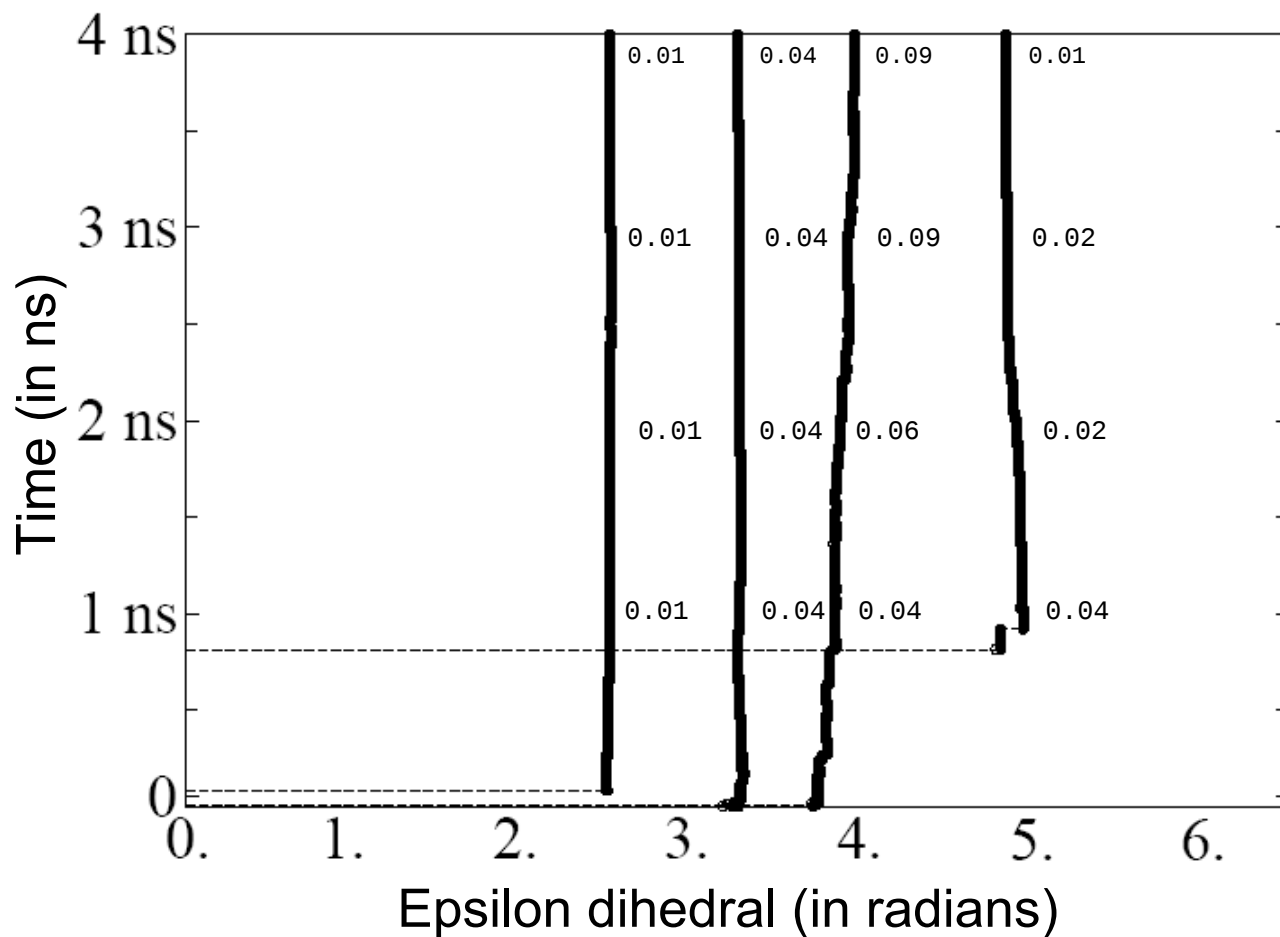


Figure 5.3. Convergence and location of sampled dihedral angle centers (averages over conformations that belong to one substate) during the D-BP-RexMD simulations of the adenine dinucleotide (for the reference replica that runs with the original force field). After exploration of a particular set of metavariables for a dihedral angle variable, the metavariables are sampled alternatively and their respective sub-averages equilibrate around some value with constant square deviation σ^2 (σ^2 is shown on the figure at regular time-interval for each sampled dihedral center).

It is important to note that this type of RexMD (D-BP-RexMD) is not expected to produce a canonical sampling in the reference replica because the potential added to the replica runs is not constant (non-equilibrium conditions in the replicas). The intention was to use it only to rapidly explore putative semi-stable substates of the system. As illustrated in Fig. 5.3 already after about 1 ns simulation time the D-BP-RexMD does not produce any new substates.

In the third BP-RexMD variant (E-BP-RexMD) the information obtained from the D-BP-RexMD was used to setup a combination (sum) of biasing potentials that represents all allocated biasing potentials from the D-BP-RexMD. Purpose of this potential is to destabilize all the dominant states explored during the D-BP-RexMD to various degrees along the replicas. Again one replica runs without added biasing potential. Since the latter type of H-RexMD contains different but constant levels of added biasing potentials controlling the replicas it is expected to produce canonical sampling in each replica because detailed balance conditions are fulfilled.

In addition to the BP-RexMD simulations and for comparison, one 20 ns conventional MD (cMD) simulation and a T-RexMD with 26 replicas (20 ns) on the same system and starting from the same start structure were also performed. Comparison of the ϵ/ζ dihedral angle plots (Fig. 5.4) indicates that during cMD simulation only two substates close to the BI starting state were sampled (Fig. 5.4a) despite the fact that the simulation was five time longer than the BP-RexMD simulations. In contrast, the T-RexMD simulation (in the lowest temperature replica) sampled both states around the B_I configuration but also additional ϵ/ζ dihedral combinations. The distribution of ϵ/ζ dihedral states is actually very similar for the first and last 10 ns of the T-RexMD simulation (compare Fig. 5.4b and 5.4c) indicating good convergence of the sampling in the T-RexMD. It should be noted, that T-RexMD simulations using fewer replicas (6 and 16) were also performed on the AA system. However, these simulations produced ϵ/ζ sampling not as well converged as the T-RexMD with 26 replicas (not shown). Presumably, the application of distance restraints to keep the system close to a

B-form like helical geometry resulted in increased barriers for dihedral transitions that could only be efficiently overcome in the T-RexMD with 26 replicas.

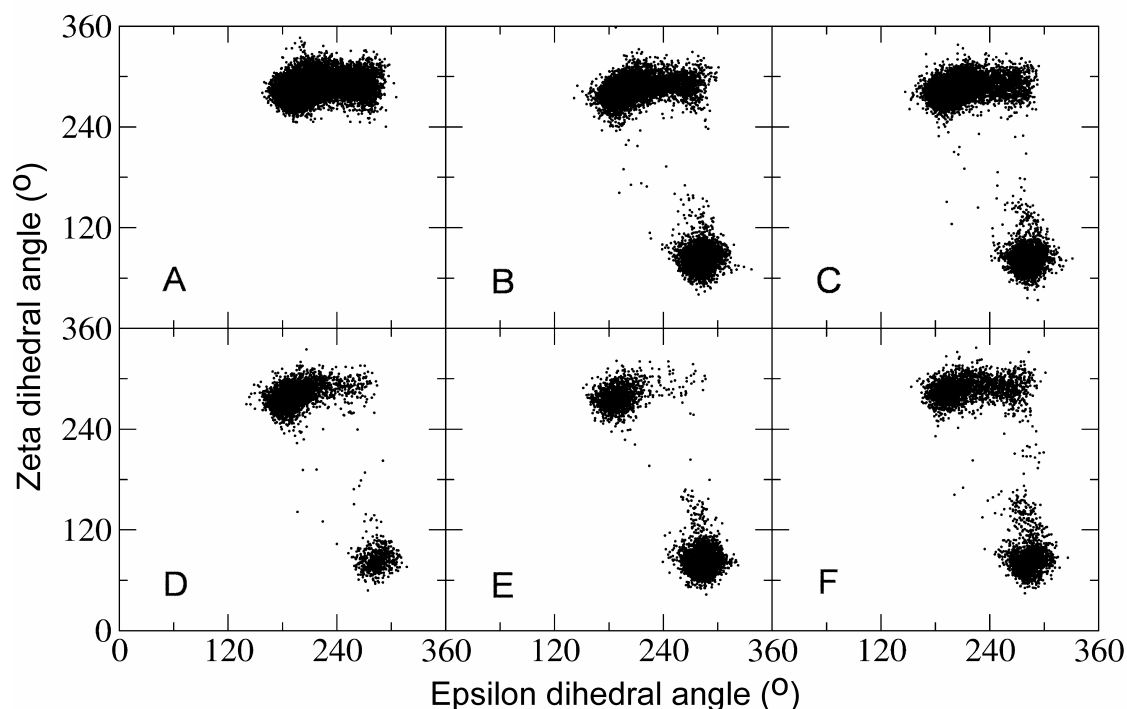


Figure 5.4. Comparison of backbone dihedral angle sampling of the adenine dinucleotide system with different molecular dynamics sampling techniques: (A) conventional MD simulation during 20 ns at 300K; (B) First 10ns sampling of the replica at 300K of T-RexMD using 26 replicas; (B) Final 10ns sampling of the replica at 300K of T-RexMD using 26 replicas; (D) Sampling of reference replica of the 4 ns S-BP-RexMD simulation (using different levels of a single biasing potential centered near the B_I state to control the replica runs); (E) Sampling of reference replica of the 4 ns D-BP-RexMD simulation (using dynamical relocations of the biasing potential in the replicas within 36 possible locations); (F) Sampling of reference replica of the 4 ns E-BP-RexMD simulation (using different levels of all biasing potentials identified during the D-BP-RexMD). A,B,C and D,E,F respectively contain the same number of sampled ϵ/ζ pairs.

All three variants of the BP-RexMD protocol resulted in increased sampling of the ϵ/ζ dihedral angle space within 4 ns compared to the five times longer cMD simulation (Fig. 5.4 d-f). However, both the S-BP-RexMD (single biasing of the B_I state) as well as the D-PB-RexMD (dynamical adaptation of biasing potentials) show quantitative differences of ϵ/ζ sampling with respect to the T-RexMD simulation. For example, D-BP-RexMD significantly undersamples one B_I substate (-gauche/-gauche) compared to the more dominant (trans/-gauche) combination (Table 5.1). The effect is more severe in case of the D-BP-RexMD

presumably because in this type of replica exchange, the reference-replica exchanges with simulations that are not at thermodynamic equilibrium due to the dynamic changes of the added biasing potential. In contrast, in case of the E-BP-RexMD the sampling of conformational substates (within 4 ns) was in almost quantitative agreement with the results of the extended T-RexMD simulations (compare Fig. 5.4f and Fig 5.4 b,c and Table 5.1) at a fraction of the computational cost already for this small system.

Simulation approach	trans/-gauche	-gauche/ -gauche	-gauche/trans	-gauche /+gauche
cMD (20 ns)	0.75 (195 / 280)	0.25 (250 / 290)	0	0
T-RexMD, 0-10ns	0.49 (195 / 282)	0.17 (247 / 292)	0	0.34 (281 / 82)
T-RexMD, 10-20 ns	0.56 (196 / 283)	0.16 (247 / 292)	0	0.28 (281 / 83)
S-BP-RexMD (5 ns)	0.65 (186 / 276)	0.19 (224 / 291)	0	0.15 (284 / 84)
D-BP-RexMD (5 ns)	0.29 (189 / 281)	0.06 (223 / 289)	0	0.65 (283 / 90)
E-BP-RexMD (5 ns)	0.49 (195 / 285)	0.12 (253 / 291)	0	0.39 (282 / 84)

Table 5.I. Sampling of ϵ/ζ dihedral angle states of the adenine dinucleotide system

In order to estimate the efficiency for exploration of different conformational states the variance of the ϵ dihedral angle up to a simulation time τ relative to the variance after 4 ns was investigated (Fig. 5.5). Mathematical details to calculate $S(\tau)$ are in appendix II. This ratio approached values close to 1 already after less than 0.5 ns in case of the T-RexMD and the E-BP-RexMD whereas the S-BP-RexMD (with a single center biasing potential) required longer times to reach the final variance of the ϵ dihedral angle

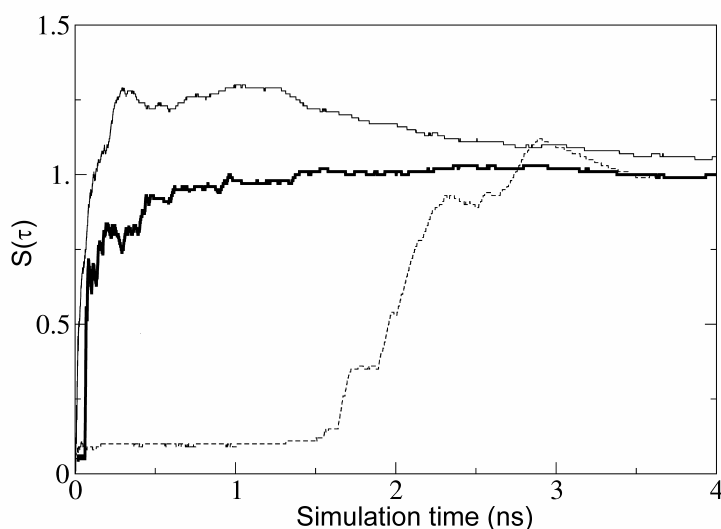


Figure 5.5. Variance of the backbone dihedral angle ϵ up to a time τ relative to the variance after 4 ns simulation time for the adenine dinucleotide system. The bold line corresponds to conformers sampled in the reference replica of the E-BP-RexMD (thin line: reference replica of the T-RexMD and dashed line: reference replica of the S-BP-RexMD simulation).

Application to a DNA system with central abasic site

One major goal of MD applications to nucleic acids is to better understand the fine structure of DNA and RNA and how it modulates recognition properties for binding of proteins and other biomolecules. Major computational efforts employing conventional MD techniques are underway to systematically explore the sequence dependence of the backbone dihedral angle states compatible with a DNA helical geometry (123). Chemical or radiation damage of DNA can result in the loss of a base (abasic site) in one strand leaving an unpaired base in the opposite strand (129). Conventional MD simulations on such abasic sites indicate significantly altered dynamics compared to regular DNA (130). In addition, systematic adiabatic mapping calculations found a variety of metastable substates and different global geometry that are accessible for DNA molecule with an abasic site (131).

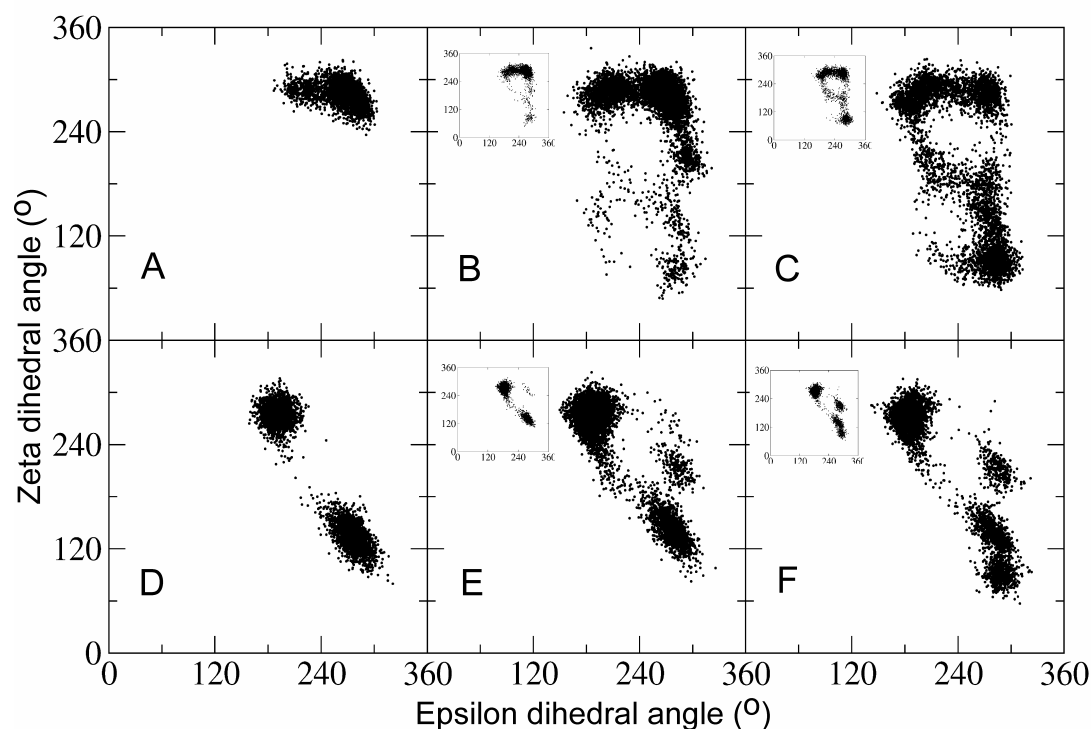


Figure 5.6. Comparison of ϵ/ζ backbone dihedral angle sampling at the abasic site (upper panels) and the opposing adenine (lower panels) of the double stranded DNA with a central abasic site. (A, D) cMD simulation during 5 ns at 300K; (B, E) Sampling in the reference replica of the T-RexMD simulation during 5ns per replica; (C, F) Sampling in the reference replica of the BP-RexMD simulation during 5 ns per replica. The insets in the panels B, C, E, and F indicate ϵ/ζ states sampled during the first 2.5 ns of the corresponding simulations. Each dot in the plots corresponds to a ϵ/ζ pair of a conformation sampled every 0.75 ps.

As an application to further test the BP-RexMD approach, the method was applied to an experimental structure of an abasic site embedded in double strand DNA (seven central base-pairs of a DNA abasic site deposited in the protein data bank: pdb-entry: 2HSS, 270). The system was investigated using a cMD simulation (5ns, 300K), a T-RexMD (16 replicas, see Annexe II) and using the E-BP-RexMD (5 ns, 300K) after employing a D-BP-RexMD for 5 ns equilibration time. The biasing potential (36 possible potentials) was only applied to the abasic site itself and the adenine nucleotide on the opposite strand. All other nucleotides were only controlled by the original force field (in each replica). Dihedral angle sampling was focussed on the sampling at the abasic site and at the opposing adenine (Fig. 5.6). The cMD simulation indicates sampling of the (ϵ / ζ) B_I region at the abasic site (with a dominant sampling of the (-gauche / -gauche) regime (90%) compared to the standard (trans / -gauche). Interestingly, this is opposite to the behaviour of the AA system where the standard (trans / -gauche) was more frequently sampled. For the opposing adenine nucleotide both B_I and B_{II} states were sampled during cMD. Both the T-RexMD as well as the BP-RexMD showed much more extensive sampling of the possible (ϵ / ζ) pairs compared to cMD on the same time scale (in the reference replica, Fig. 5.6 b,e and 5.6 c,f).

abasic site				
simulation method	trans/-gauche	-gauche/-gauche	-gauche/trans	-gauche /+gauche
cMD (5 ns)	0.12 (227 / 290)	0.88 (273 / 283)	0	0
T-RexMD (5ns)	0.22 (206 / 284)	0.63 (272 / 285)	0.06 (266 / 186)	0.08 (268 / 97)
BP-RexMD (5ns)	0.30 (198 / 279)	0.22 (267 / 284)	0.17 (253 / 177)	0.31 (277 / 97)
adenine nucleotide opposite to abasic site				
cMD (5 ns)	0.53 (191 / 275)	0	0.47 (276 / 134)	0
T-RexMD (5ns)	0.74 (185 / 272)	0.07 (267 / 210)	0.19 (277 / 142)	0
BP-RexMD (5ns)	0.63 (184 / 269)	0.07 (281 / 211)	0.17 (277 / 142)	0.13 (290 / 93)

Table 5.II. Sampling of ϵ/ζ dihedral angle states of the DNA abasic site

Both replica simulations predict a more even distribution of the B_I substates (ϵ / ζ) = (-gauche / -gauche) vs. (trans / -gauche) and also sampled several additional states (e.g. in B_{II} region) both for the abasic site as well as for the opposing adenine. The quantitative difference in sampling of four major (ϵ / ζ) dihedral states was further analysed (Table 5.2) indicating that cMD and T-RexMD undersample (or do not sample) the (-gauche/trans) and (-

gauche/+gauche) regime for the abasic site and the (-gauche/-gauche) and the (-gauche/+gauche) regime in case of the opposing adenine. Interestingly, in case of the BP-RexMD simulation the sampled states within the first 2.5 ns of the simulations (insets of Fig. 5.6 c,f) look already quite similar to the sampling in the full time range (5 ns). In contrast, in case of the T-RexMD parts of the (ϵ/ζ) dihedral angle plots show additional sampled regions not sampled within the first 2.5 ns. This indicates better convergence of the BP-RexMD with only 5 replicas compared to 16 replicas in the T-RexMD case. In addition, there are some additional (ϵ/ζ) combinations in the BP-RexMD that are not yet sampled during the T-RexMD (compare Fig. 5.6e and Fig. 5.6f).

Correlation between BI / BII transitions and DNA bending

It was shown in Chapter 4 (Fig. 4.5) that slightly different bending energy curves could be obtained by adiabatic mapping for the intrinsically bent Hagermann sequence due to strong pre-bending that has induced transition toward irreversible conformational DNA backbone substates within JUMNA. Recall that in JUMNA ϵ and ζ dihedrals are the only independent variables in the DNA backbone (section 2.3). As could be expected, the above backbone substates with influence on DNA bending were systematically due to ϵ and ζ . Now it is interesting to recall also that in Chapter 4 we concluded that no significant BI / BII backbone transitions were associated with strong base pair bends (in contrast to (118-120)). Two possibilities arise then. Either the coupling observed in (118-120) between BI / BII and unbent / bent base pair junctions (bending roughly in the range 20° to 50°) has implications only in a very few specific cases (such as the HIV-1-kappa-b DNA binding site (119-120)), or the unusual stereochemical arrangement of bases inside strong base pair bends precludes an exhaustive exploration of the conformational landscape associated to backbone dihedrals on 3 ns time scale of the current MD Umbrella Sampling. Some insights onto this question can be found by carrying-out the REUS sampling since one property of this technique is to accelerate the occurrence of rare events in MD. To gain further insight onto the question Fig.5.7 illustrates DNA bending induced by the abasic site when using the method of the current chapter. An abasic site is an example of unusual DNA structure as is the type II kink motif. But here the problematic is tackled in its inverse direction since the parameter space of the current Replica Exchange strategy does not promote specifically DNA bending (as in Chapter

4) but DNA backbone transitions instead. Fig.5.7 represents the free energy profile of the (ϵ, ζ) phase space for the abasic residue.

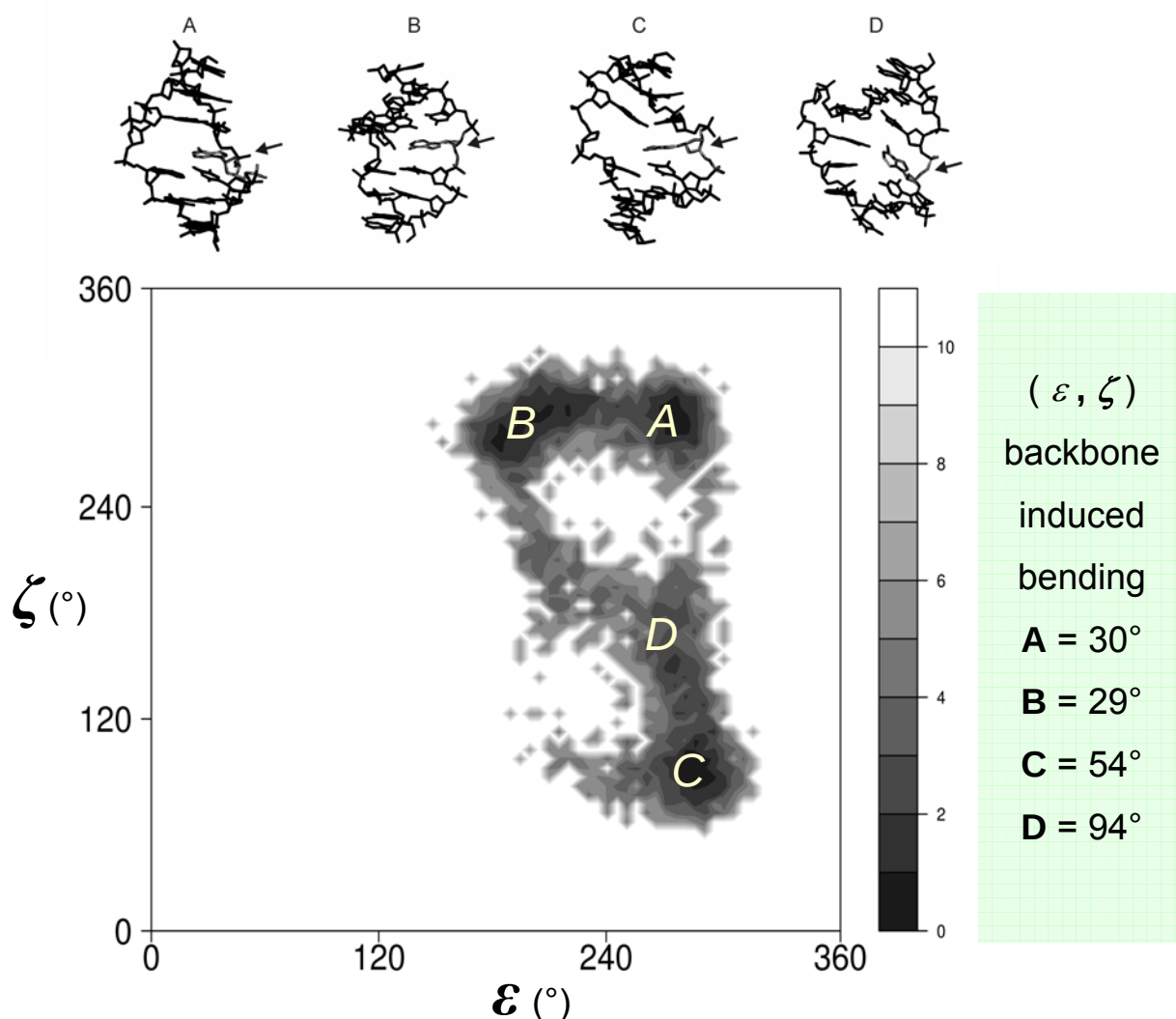


Figure 5.7. Two dimensional free energy taken as $-\log P(\epsilon, \zeta)$ for epsilon/zeta backbone conformations inferred from the E-BP-RexMD sampling after 5 ns in the abasic residue of the ABA oligomer. The free energy is indicated by a grey scale contour map in unit of $k_B T$. The position of four cluster-centroids obtained by a 2D least square fit to the (ϵ, ζ) phase space is indicated by a letter (A-D). The induced global bend angle of representative conformation with (ϵ, ζ) values corresponding to each centroid is indicated in a green box on the right hand side. Stick representation of these conformations is shown on top of the graphic with unpaired nucleotide in light grey indicated by an arrow.

The bold numbers in degrees indicate the bend angle calculated between two linear axis generated by Curves on each side of the abasic residue for each representative structure of (ϵ, ζ) clusters. Each of the latters is a conformation for which (ϵ, ζ) value approximately equals the cluster-centroid A, B, C or D in Fig. 5.7 obtained by least square fitting (k-means

clustering) on this 2D plot. As is evident from Fig.5.7 the (ϵ, ζ) BI/BII backbone mode has obvious consequences on the bend angle induced in the local environment. It is thus possible that the absence of BII states in base pair kink motifs characterised in Chapter 4 simply illustrates the main method-limitation, i.e. an incomplete sampling of conformational substates that are compatible with severely bent DNA conformation.

5.5 Discussion

A new Hamiltonian replica exchange method has been designed that uses a biasing potential as replica coordinate that specifically promotes dihedral transitions in the nucleic acid backbone. Application of the method to a DNA dinucleotide and a DNA with a central abasic site showed significantly improved sampling of the ϵ and ζ dihedral angle substates compared to cMD simulations and similar (or even better sampling in case of the abasic site) compared to T-RexMD at a smaller computational demand. An advantage of the BP-RexMD methodology compared to T-RexMD is the fact that the energy differences between replicas (used to accept or reject exchanges) are only affected by the force field term which changes upon going from one replica to another replica run. Hence, the exchange probability is only affected by the backbone dihedral angle terms and not affected by solvent-solvent and solute-solvent (and many other solute-solute) contributions. The number of required replicas should only grow with the number of dihedral angles involved in the application of a biasing potential. An additional advantage is the possibility to focus the biasing of dihedral transitions to parts of a nucleic acid (e.g. a damaged nucleotide or a central sequence in case of a comparison between DNA with different central sequence but otherwise identical flanking regions) keeping the “unbiased” (original) backbone dihedral angle potential for the rest of the nucleic acid in all replicas. This opens the possibility to specifically enhance the sampling of only parts of a DNA or RNA under realistic simulation conditions and using only few replicas (e.g. folding simulations of structural motifs in nucleic acids such as internal loops or hairpin loops). Compared to the temperature as a replica coordinate it is also expected that this specifically improves the sampling of sub-states without disrupting other parts of the structure (e.g. base pairing, strand dissociation). The enhanced barrier crossing will lead to an improved sampling of sub-states relevant for the nucleic acid structure at the desired temperature.

The BP-RexMD in its current setup requires mapping out the location of putative stable substates for a given set of dihedral angles during an equilibration phase. It is possible to extend the method by checking even during the production phase if new states have been sampled and to update the biasing potentials for these regimes until no further state is sampled. The step could also be completely avoided if one employs expert knowledge on the relevant nucleic acid substates from the many experimental nucleic acid structures and to preset an appropriate biasing potential based on the experimental distribution of backbone dihedral angles.

The study reported in this chapter allows characterizing the fine structure of dsDNA systematically under realistic conditions and may allow also the systematic evaluation of the performance and precision of current force fields. It suggests that the new method is necessary to correctly describe the (ϵ / ζ) metastable conformations on short simulation time-scale on unusual Nucleic Acids such as the DNA abasic system. As long as the set up of a flooding network in the dihedral phase space (selection of a specific set of some “metavariables”) is rapid, the automatic E-BP-RexMD approach will be an efficient tool for exploring energetic landscapes associated to dihedral angle phase space of Nucleic Acids and of Biomolecules in general. The equilibration phase of E-BP-RexMD will be considered “rapid” if it last only a fraction of the sampling time chosen for Production. In case of AA and ABA such selection of metavariables was complete after approximately 1 nanosecond. Generalization of the sampling method to polypeptide backbone dihedral Phi/Psi is currently under test.

Acknowledgements

This study was performed using computational resources of the CLAMV laboratory (Computational Laboratories for Analysis, Modeling and Visualization) at Jacobs University, Germany and was supported by a grant from the VolkswagenStiftung.

Chapter VI

Variation on the method: replica-sampling with a server-client relation

. **Keywords:** Replica Exchange; Molecular Dynamics; Adaptive Sampling; Dihedral angles.

. **Reference publications:** Curuksu, J. (2009) *in preparation*.

. **Recommended Introductions:** 2.7, 2.9, chapter V.

6.1 Abstract

Several advanced sampling techniques are available to improve the multiple minima problem in molecular dynamics simulation of biological systems. Yet a trade off between efficiency and accuracy has not been amenable to one general methodology. I present a flexible replica exchange algorithm which explores the phase space of a given order parameter by self-regulating a scaled ladder of sampling copies (Server replicas). The reference sampling of interest (Client replica) is under the control of an Hamiltonian function without biasing potential and provides a canonical limit in the chosen thermodynamic Ensemble (NVT). The automatic scaling is based on a Tabu search which consists to orient the searching in phase space by learning (definition of either forbidden or prohibited locations in phase space). It is defined here by the integration of a short term memory and a long term memory functions which expand the Hamiltonian of some Server replicas. The Client “sampling” provides requests which trigger adaptive solutions to be searched for by the Servers, as the simulation proceeds. The approach is applied to nucleic acid backbone degrees of freedom and allows us to sample a wide range of both canonical and non canonical nucleic acid backbone sub-states into a Boltzmann sampling on a dinucleotide DNA model system. The sampling efficiency is indicated by a comparative analysis with Temperature replica exchange and BP-RexMD methods of chapter 5. Possible application of the method to nucleic acids, proteins and other biological molecules is discussed.

6.2 Introduction

Molecular dynamics simulation is a standard tool to account for conformational variability and dynamics underlying the function of biomolecules but is limited by poor sampling due to kinetic trapping events in low-energy regions of the complex energy landscape. Broken ergodicity (section 2.7) at room temperature can be reduced by many techniques amongst which Temperature replica exchange (13) and Hamiltonian replica exchange (14) methods are standard examples. The BPRExMD technique (Chapter 5) is an Hamiltonian replica exchange method which ensures that low-energy regions are canonically sampled in the unbiased replica but that frequent transitions in the subspace of backbone dihedrals occur due to conformation exchanges with the scaled replicas. In contrast to the standard Temperature replica exchange, the scaled replicas of Hamiltonian replica exchange algorithms do not provide a canonical limit but mainly enable alternative solution in the phase space to “take over” in the unbiased replica to avoid kinetic trapping in the latter. Thereby if the ladder of scaled replicas adapts to the current state of this unbiased replica as well as to its cumulative history, and without violating the criterion of detailed balance governing inter-replica exchanges, the method would take a full advantage of its formalism. An attempt for such an algorithm is derived in this chapter.

6.3 Background

Early adaptive sampling techniques consisting in recognizing conformations sampled before were based on iterative updates of an umbrella potential with the Weighted Histogram Analysis Method (194) in the course of the simulation. An Adaptive Umbrella Sampling algorithm using the potential energy (276) as the umbrella parameter (formally equivalent to Multicanonical Sampling, 277-279) was further developed to obtain uniform sampling of the Potential Energy Surface (PES) and faster transition in the dihedral angle distribution function. A long convergence time over low energy minima due to the *global* nature of the umbrella potential and to the iterative procedure is a natural drawback of such self-regulating schemes for complex system. For such system residual free energy barriers after convergence may also impede fast transition between important conformational states and in that case selected parameters in the potential energy function are a useful solution to a specific

conformational subspace which is to be explored (276). Adaptive biasing force schemes go into this direction (230,231).

Learning processes during MD sampling consisting to deform the PES or Free Energy Surface (FES) were also pioneered by Local Elevation method (LE, 271) where Gaussian potentials are cumulated to penalize conformations previously visited and for which height is updated proportionally to the number of time a parameter set (range of dihedral angle values in the original LE method) has been sampled before. Metadynamics (272,273,282) raises energy wells based on the effective FES as the simulation proceeds, thereby permitting to recover the (imaged) FES once the simulation has converged - see also the self-healing Umbrella Sampling scheme (283). Self-modification of the energy surface techniques can also consist to introduce a bias potential to the true potential such that potential surfaces near the minima are raised (Hyperdynamics, 284) and those near barriers or saddle point are left unaffected (or lower as in 285,286). Hyperdynamics also evolves as the simulation proceeds without any advanced knowledge of the hypersurface shape, but requires either calculating the Hessian matrix or minimising its first derivatives numerical approximation at each step (284). This can be avoided by relying on a threshold boost energy value and predefined height of biasing potentials as in accelerated MD (287), but in that case these two parameters have to be fine tuned using short trial runs beforehand and thus cannot guarantee optimal sampling for the whole available configurational space. Multicopy search based techniques are a common way to improve sampling and a self-regulating regime of conformational search on the PES (optimal update of the PES deformation operator) was recently achieved (288) by propagating replicas of an identical system which compete against each other. To do so these authors made the replicas share a history dependent variable (the PES deformation operator) for which sum over all replicas is subject to holonomic constraint.

On the other hand enhanced sampling method based conformation exchanges between some parallel replicas (13-17, 197, 199-202, 257-261, 262-269 and Chapter 5) are attractive compared to other generalized ensemble approaches because no pre-production runs are required for the determination of the replica weight factors and canonical distribution is directly obtained in the unmodified replica. The method yet becomes prohibitively time consuming if replica conditions differ in their actual temperature (as in 257-264) or non-ergodic sampling may not be avoided if only specific parameters of the potential energy function differ between the replicas as in BPRexMD (202) and (14-17, 197, 199-201, 265-269). For these reasons, the canonical replica sampling (physiological temperature or unmodified force-field) may better be seen as a Client in a scheme where the remaining

replica ladder will respond effectively as a function of the current state and history recorded in the production (unbiased) replica.

The *Client* acts then as a target searching copy for which trajectory furnishes direct requests to other replicas. These *Server* copies may integrate optimal biasing potentials of diverse strength so as to propose adapted alternative solutions to the Client replica as the simultaneous simulations proceed. The loss of efficiency on the whole available conformation space compared to competing potential energy function scaling replicas (288) is replaced by faster search on a phase-space of interest, the NABD (Nucleic Acid Backbone Dihedral) subspace. In addition such *Server-Client* scheme will be conditioned by the criterion of sampling a canonical distribution in the Client replica. In other words the Client replica produces Boltzmann weighted ensembles of conformation. Details of the self-updating NABD biased search in the Server replicas and its relation to the Client production-replica that I have developed will be presented in this chapter. Nucleic acid backbone elastic properties are the focus of the current work and, as a follow-up of chapter 5, results will concern the DNA ϵ - ζ backbone crankshaft mode. The goal of the development is to provide a method fully self-regulatory thus the script of the Server-Client algorithm, given in Annexe IV, can be applied directly to any dihedral variable or couple of dihedrals.

6.4 Method

BPRexMD ansatz in NABD

BPRexMD is an Hamiltonian replica exchange ansatz where the scaling potential or replica coordinate v_{NABD} takes the form of eq.5.3 in Chapter 5:

$$v_i(\rho^{\epsilon}_c, \rho^{\zeta}_c) = k \times \begin{cases} ((x-r)^2 - (R-r)^2)^2 & \text{if } r < x < R \\ (R-r)^4 & \text{if } x < r \\ 0 & \text{otherwise} \end{cases}$$

where $x = \sqrt{(\rho^{\epsilon}_c - O)^2 + (\rho^{\zeta}_c - O)^2}$ corresponds to the current value $(\rho^{\epsilon}_c, \rho^{\zeta}_c)$ of a chosen ϵ/ζ combination (see Chapter 5 for details). We use increasing k value across the replica ladder

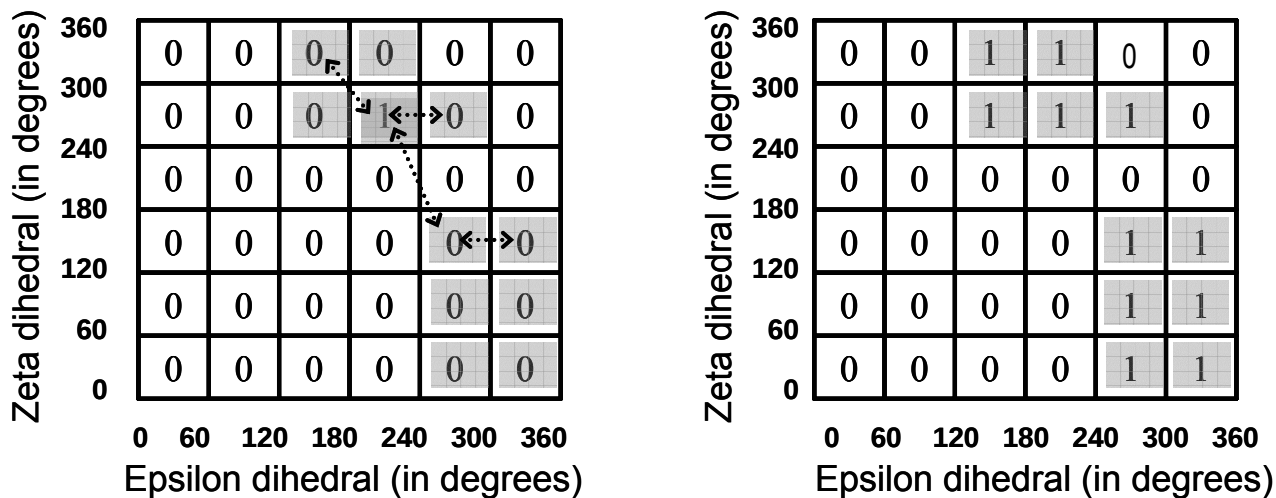
and thus v_{NABD} fills potential wells of a given subset of nucleic acid backbone dihedrals in (n-1) replicas. These n-1 replicas play the role of some “retailers” ensembles to which a canonical sampling (unbiased replica) has periodically access. From now on these n-1 replicas will be called the Servers and the unbiased replica the Client (Server Client Replica Exchange Method, SCREM). I also refer to the six equi-spaced intervals defined for each dihedral variable in Chapter 5 as meta-variables.

Server Client SC-REM ansatz: the short term memory function

The BPRexMD ansatz of Chapter 5 (E-BP-RexMD) defines a pre-equilibrating phase during which a number of positions where to introduce a biasing potential is determined in the two dimensional ε / ζ phase space. When BPRexMD equilibrates (i.e. D-BP-RexMD) the test of a convergence criterion (a_c^i , Annexe II) decides automatically if the biasing potential is switched to the position of the metavariable which is currently visited and the value of this criterion is stored in memory. Thus the many positions of metavariable, initially unknown, are determined on the fly. This self-regulated flooding of current phase point (D-BP-RexMD) will be referred to as the short term memory function.

As shown in Fig.6.1 the *pre-equilibrating* phase with one biasing potential dynamically allocated (D-BP-RexMD) performs more efficiently than the *production* phase because the latter requires evaluation of several biasing potentials at each time step, that is, the ones selected during the pre-equilibrating phase. Hence I discuss the approximations made in the efficient self-equilibrating D-BP-RexMD and accordingly build-up an ansatz that takes full advantage of using just one biasing potential of torsion-metavariable. A quasi-equilibrium situation of the self-equilibrating D-BP-RexMD can be resumed as follows. In the limit of long-time sampling i.e. every metavariable’s average is converged, the metavariables are sampled alternatively and each determines a Hamiltonian function with a specific position of the biasing potential invariant with time. The sampling device defined by the n-1 Server replicas then becomes a sequence of repeated sampling over these stationary metavariables. From the theory of non-equilibrium fluctuation such a repeated sampling of substates converges toward equilibrium in the long time limit, that is after many switches between these substates (289-291). In this limit the alternative biased Hamiltonians define alternative exchange criterions which are exact (time independent). Since the true probability distributions (the exact Hamiltonians) are used in the equations for the exchange criterions, these exchange criterions in principle satisfy detailed balance (292).

Sampling phases of E-BP-RexMD:



Equilibration - - - - → **Production**

Figure 6.1: Flooding of the ϵ / ζ backbone dihedral sub-space in the equilibration (D-BP-RexMD) and the production phase of the BPRexMD method introduced in Chapter 5.

Server Client SC-REM ansatz: the long term memory function

In Hamiltonian replica exchange algorithms the scaled replicas are often of no interest for the property of interest (see Introduction). Second the automatic mapping of some dihedral metavariables proposed in D-BP-RexMD is flawed in a sense that can be stated as follows. First the adaptive flooding of some metavariables with just one potential dynamically allocated does not guaranty a self-regulated sampling over *every* accessible conformations even in the long time limit. In other words an accelerated transition between only a few minima may result even if other important regions in the phase space exist farther away in the actual sequence of transitions between dihedral substates. Furthermore (second flaw) the complex shape of the potential energy wells characterising the visited metavariables may prevent excursion to some sub-minima of these wells in the Servers, because the discretisation of dihedral angles (metavariables) used to allocate the biasing potential v_{NABD} does not distinguish between these sub-minima.

These two impediments can be avoided if the Server replicas avoid proposing the very dihedral values in which the Client replica has thermalized in the past. This can be done using

the approach of Metadynamics (271-273) which were recently shown to not violate detailed balance when combined to the replica exchange formalism (274,293,294). Thus a long term memory function is introduced to fill the energy wells associated to every visited conformation in the PES of the Server replicas, as the simulation proceeds, using Gaussians with width sufficiently small to not mask any details (distinct sub-minima). For peptide and nucleotide backbone dihedrals a maximum of 16 states per dihedral is sufficient to distinguish any conformational change (see 271). In the long time limit the long term memory function tends to induce an approximately flat energy surface as in Metadynamics (271-273).

The current Gaussian functions take on the form:

$$g_l(\tau, t) = w_G \times \sum_{t' < t} \exp - \frac{(\tau_l(t') - \tau_l)^2}{2\sigma_G^2} \quad \text{eq.6.1}$$

where τ is the dihedral variable of interest, t is current time, l is the index of the discrete reference value τ_l corresponding to the current value of the dihedral τ , w_G is an *energy* term per re-sampling of a conformation (chosen to be several order of magnitude lower than typical DNA backbone dihedral barrier; w_G determines the force constant of each unit Gaussian function). σ_G is the width of this memory penalty function and defines a complete set of discrete torsion values for g_{NABD} (width = $360^\circ / 16$). The complete Server-Client ansatz is illustrated in Fig.6.2.

In summary, the above function g_{NABD} exponentially discourages the re-sampling of dihedral discretized values which are visited and does not require pre-production trials neither *on the fly* iteration. Similarly the parameters for the “coarse-grained potential” v_{NABD} of Chapter 5 (E_{max} and d_l-d_4) do not involve tedious determination: these parameters are chosen only to assure very fast transition between (-gauche), (+gauche) and (trans) substates = the short term memory of the SCREM ansatz. Statistical accuracy is obtained through self-regulating rules of Server vs. Client interactions, the detailed balance equation:

$$\omega(C_i \leftrightarrow C_j) = \min \left\{ 1, \exp \beta \left[\begin{aligned} & (v_i(q_i) - v_i(q_j)) + (v_j(q_j) - v_j(q_i)) \\ & + (g_i(q_i) - g_i(q_j)) + (g_j(q_j) - g_j(q_i)) \end{aligned} \right] \right\} \quad \text{eq.6.2}$$

The inclusion of time dependent Gaussian functions g_i is possible because Gaussian units have an infinitesimal height and thus preserve an approximately canonical distribution in the

Client (unbiased) replica (274). Fig.6.2 schematizes the Server-Client relation. In the Servers the systematic biasing of NABD metavariables is adjusted to sampling points of the Client replica which are stationary in phase space (D-BP-RexMD), but as the Gaussian potentials accumulate in the Servers, conformations made available for replica exchange are increasingly unlikely to have been previously visited by the Client. Eventually the exploration of an entire metavariable in the Servers (*trans* in Fig.6.2) will be discouraged by long term memory functions. This

happens if the phase point in the Client does not by itself migrate to unexplored

metavariables. Thus banning this metavariable to be sampled by the Servers and requesting new options in the state-to-state transition sequence to be involved in subsequent replica exchange attempts.

Fig.6.3 indicates the complete flow chart of the SCREM method.

The time series

recorded for ϵ and ζ in the Client and one Server replicas of a test simulation (Fig.6.4) show the basic idea of the SCREM sampling strategy. One of the Server replicas (in grey) samples dihedral values which are not sampled currently in the Client replica (in black). The Client replica selects only low energy conformations (close to -80° , 180° and $+60^\circ$) amongst propositions given by the Server replica with time. For example when the Server replica wanders in the zone 90° - 140° (see the ζ time series between 0.2/0.3 ns) the Client replica only accepts conformations for torsions $> 130^\circ$ (*trans* rotamer).

Long-term memory function:

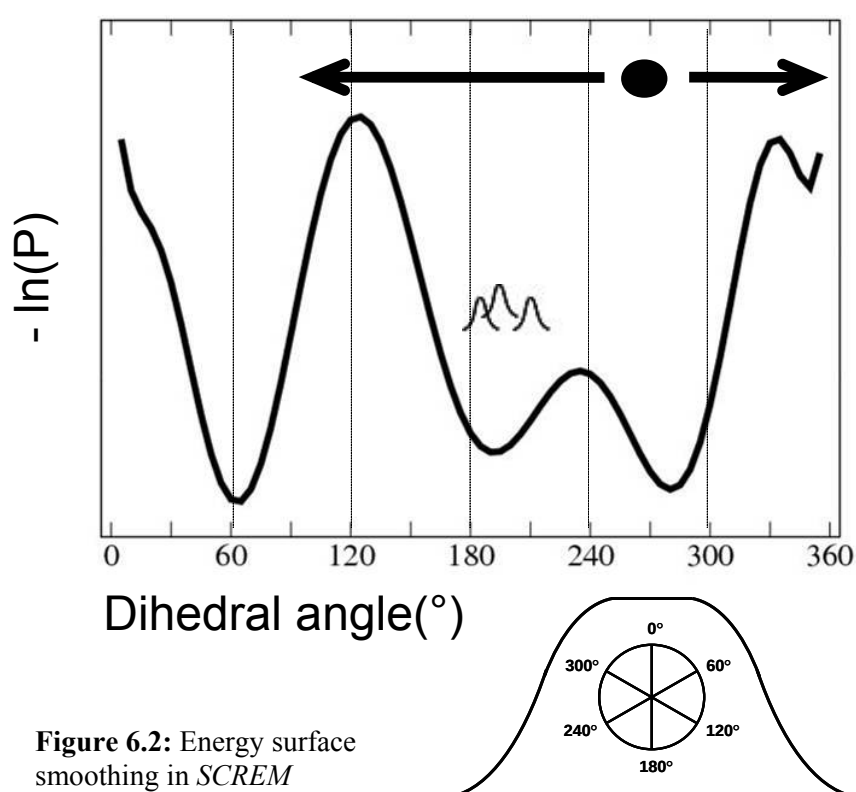
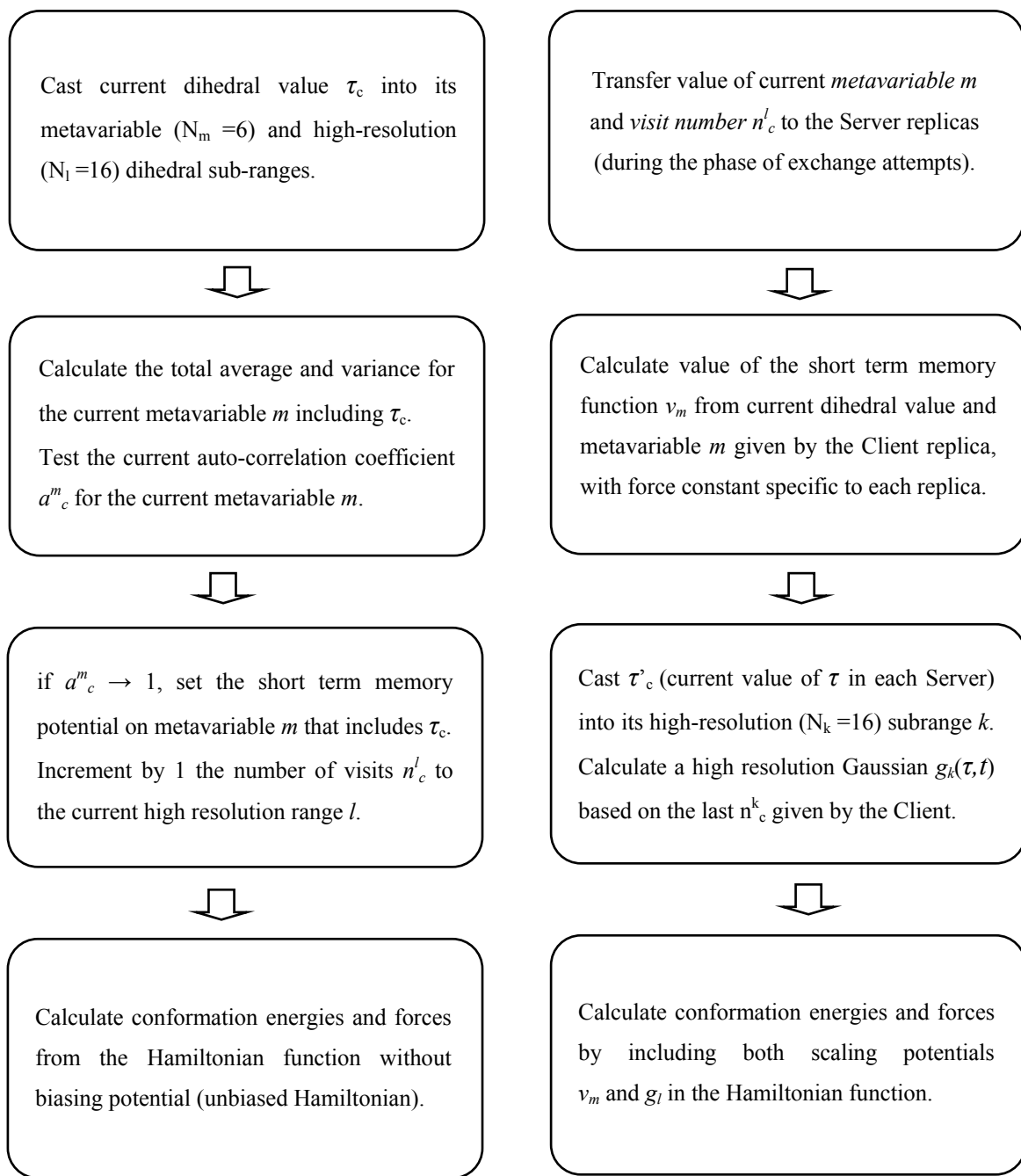


Figure 6.2: Energy surface smoothing in SCREM



Client replica

Send scaling-updates

Provide canonical (unscaled) sampling

Server replicas

No control on scaling updates

Scaling of dihedral-rotamers: 6 metavariables

Scaling by high resolution Gaussian functions

Figure 6.3. Flow chart of the SCREM (Server Client replica exchange Method) sampling scheme.

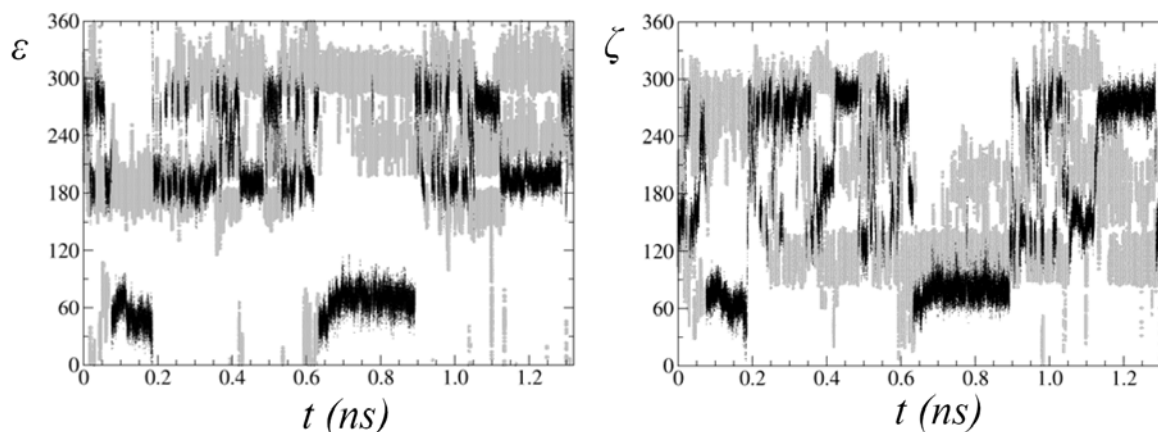


Figure 6.4. A practical example of the Server vs. Client relation in the SCREM method. The black line corresponds to the Boltzmann sampling in the Client replica and the grey line to the sampling in one Server.

Simulation Details

Parameters for v_{NABD} and for the Gaussians do not need numerical treatment as explained above. Strength (E_{max}) of v_{NABD} are taken as in Chapter 5. Simulation parameters for every experiment are the same as in Chapter 5 or otherwise mentioned (see Table 6.I in the Result section). Local updates of conformation in each replica were also realised by 1fs molecular dynamics moves in explicit solvent at constant NVT. Distribution of the exchange frequencies between replica pairs are reported in Annexe II. Results obtained with the SCREM method are compared with the ones from standard molecular dynamics, Temperature replica exchange and BPRexMD obtained in Chapter 5 on the Adenine dinucleotide model system d(AA). The new parmbse0 force field (19) and identical starting conformations for each replica were used.

6.5 Results

Sampling performance and parameter tuning on di-Adenine

The SCREM approach cumulates infinitesimal Gaussian functions in addition to the more coarse grained flooding potential allocated on stationary metavariables, for smoothing the energy surface associated to backbone dihedral angles in some Server replicas. The

introduction of Gaussian functions depends on phase points sampled in the past in the Client replica. However such time-dependent Gaussian functions were shown to not violate detailed balance in replica exchange as long as the height of the Gaussian unit is small (274, 293, 294) and the penalty is also included in the detailed balance equation (292), as done here. These Gaussians provide strategic forgetting (prohibited locations in phase space) with time (295) and eventually will correct the non equilibrium sampling over short time in the “cheap” D-BP-RexMD scheme introduced in chapter 5 (Fig.6.5, Table 6.I first line).

This is still an on-going project and clearly finding a trade off in term of optimal interplay between short and long term memory functions is not trivial. The situation can be summarized as follows. Using a Gaussian unit’s height sufficiently low (10^{-5}) to not offset the height of the coarse-grained potential v_{NABD} (short term memory function) after a relatively long sampling, their effect does not sufficiently correct the sampling of the D-BP-RexMD scheme (Table 6.I). They nevertheless significantly improve the sampling of phase space (compare line 1 and line 2 of Table 6.I with the reference TRExMD simulation).

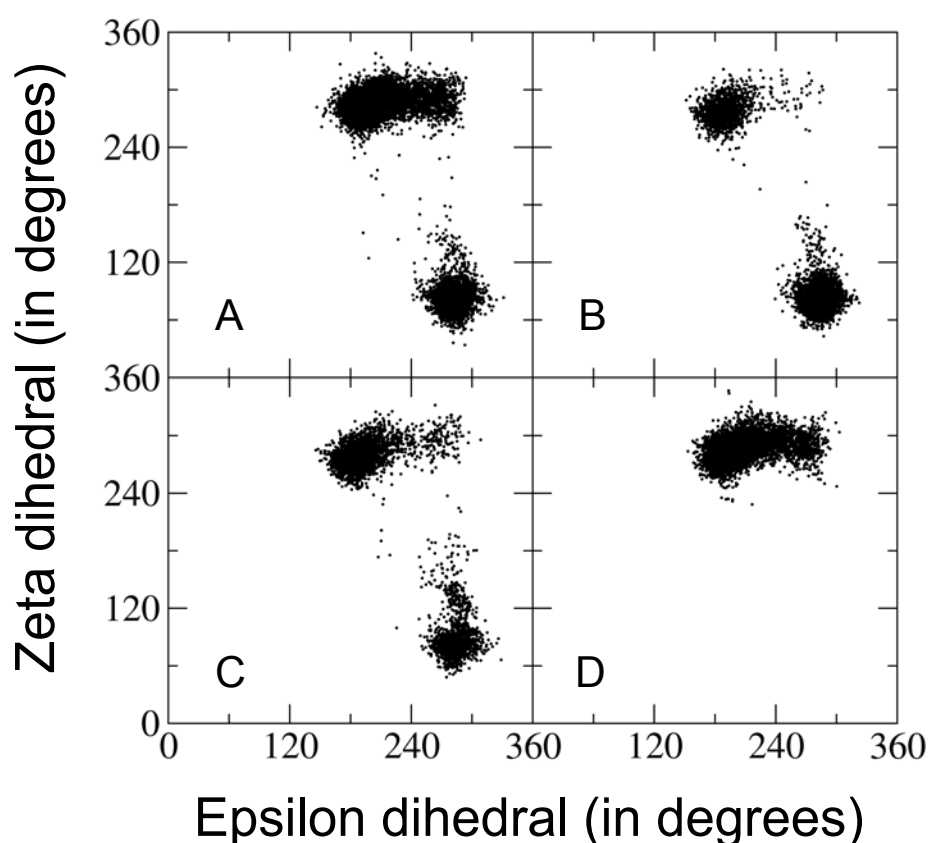


Figure 6.5. Comparison of ϵ / ζ backbone dihedral angle sampling of the adenine dinucleotide AA system with (A) TRExMD, (B) D-BP-RexMD and (C, D) SCREM for unit Gaussian height of 10^{-5} (C) and 10^{-3} (D) respectively. See Table 6.I and text for simulation details.

As second result of Fig. 6.5 / Table 6.I, we see that increasing the height of the Gaussian unit is not able to improve the situation (Fig.6.5 D, Table 6.I third line). Indeed the Gaussians g_{NABD} now quickly offset the height of the short term memory function v_{NABD} and lead to an artificial landscape common in every scaled replica (because g_i are identical in every scaled replica) which dominates the exchange equation. This means that the exchange probability for a pair of Server replicas is systematically $P = 1$ (not shown) and in contrast the exchange probability with the Client replica, where $g_i = 0$ kcal/mol, vanishes: $P = 0.08$. We see that a first current outlook for parameters tuning would be to try different (increasing) Gaussian unit's heights across the ladder of Server replicas.

SC-REM PARAMETERS	Trans / -Gauche	-Gauche/-Gauche	-Gauche/+Gauche
Only v_i (short term memory) 5 replicas	0.29	0.06	0.65
$\omega_G = 10^{-5}$ 5 replicas	0.38	0.27	0.35
$\omega_G = 10^{-3}$ 5 replicas	0.74	0.26	0.
Only g_i (long term memory) $\omega_G = 5.10^{-4}$, 2 replicas	0.66	0.13	0.21

REFERENCE SAMPLING	Trans / -Gauche	-Gauche/-Gauche	-Gauche/+Gauche
TREM (Chapter 5) 26 replicas, 20 ns per replica	0.53	0.16	0.31

Table 6.I. Results of a cluster analysis performed on epsilon-zeta backbone dihedral angles densities of the adenine dinucleotide AA system, using D-BP-RexMD sampling (Chapter 5), SCREM sampling with unit Gaussian height of 10^{-5} and 10^{-3} respectively, and SCREM sampling with only the Gaussians and just two replicas. The reference TRexMD sampling of Chapter 5 is shown again for direct

comparison. Rotamer density matrices were computed using a k-mean cluster analysis with pre-selected starting centroids for the least-square fitting optimization.

Note in the last situation it is almost equivalent to replace the ladder of Server replicas by a single Server since the short term memory v_{NABD} , which in this set up is the only difference between the Server replicas, is not “in balance with” the effect of the Gaussian functions (g_{NABD} offsets v_{NABD}).

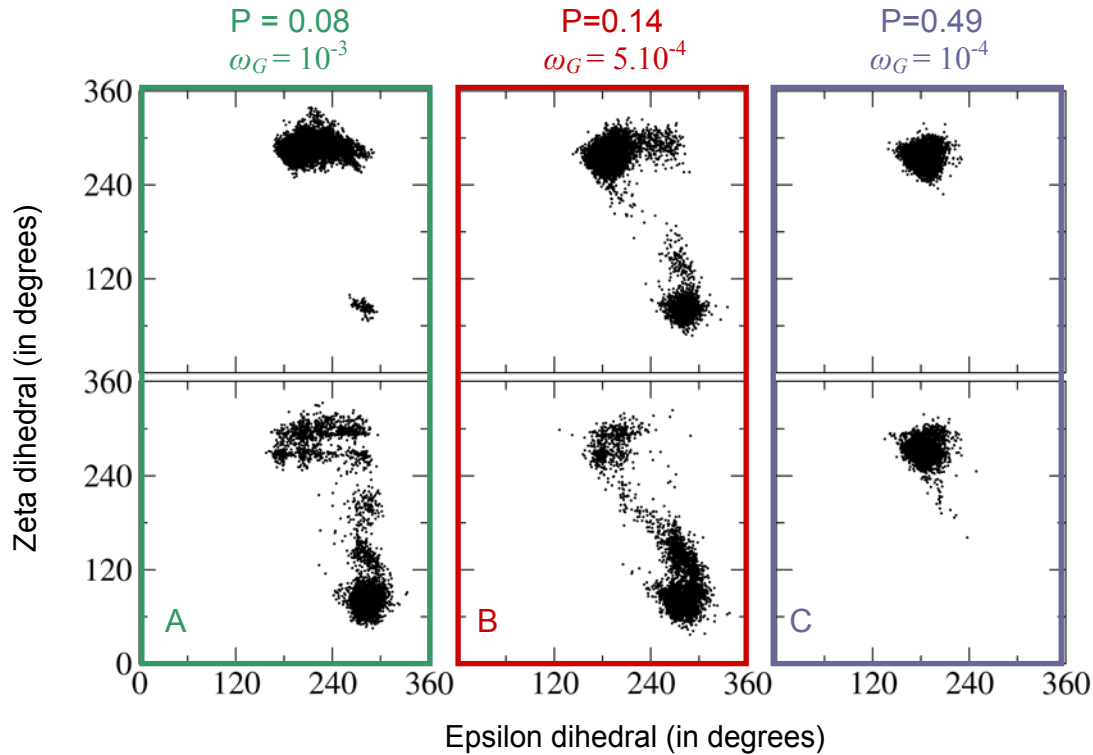


Figure 6.6. Sampling of ϵ/ζ backbone dihedral angles of the DNA dinucleotide system “AA” in the client (up) and the server (down) replicas of a 2-replicas SCREM simulation. The height of unit Gaussians ω_G introduced in the server is indicated on top of each result as well as the total exchange rate P between the Client replica and the Server replica after 5ns.

Using only one Server thus and one Client the ansatz becomes a self-regulating 2-replicas approach and is thus highly efficient. Results of the 2-replicas-SCREM scheme with Gaussian unit’s height of 10^{-3} , 5.10^{-4} and 10^{-4} are shown in Fig.6.6, both for the Server and the Client. Self-complementarity between the two emerges (see A and B in Fig.6.6): the Server explores many different conformational areas in phase space and the Client only selects the areas which are compatible with its unbiased Hamiltonian energy function. Moreover the results of Fig.6.6 have pinned the origin of the challenge of SCREM methods. For a relatively low

amplitude of the long term memory based function g_{NABD} (blue pair of plots) the Server does not make a lot of transition after 5 ns. For a relatively high amplitude of each unit function g_{NABD} (green pair of plots) the Server has made many transitions after 5 ns but the exchange acceptance rate between the biased Server and the unbiased Client vanishes, thus the Client does not make a lot of transitions. For intermediary amplitude of each unit function g_{NABD} (red pair of plots) the sampling is optimal after 5 ns and agrees with TRexMD sampling (Fig.6.5, Table 6.I).

Method	# replicas	# days per replica	# effective days
TRexMD	26	6.0	156
TRexMD	5	5.0	25
S-BP-RexMD	5	2.0	10
D-BP-RexMD	5	3.0	15
E-BP-RexMD	5	5.0	25
SCREM $v_{NABD} + g_{NABD}$	5	3.5	17.5
SCREM only g_{NABD}	2	2.5	5

Table 6.II. Comparison of the number of days required to carry out 5 ns of sampling per replica on the dinucleotide system AA. It is important to keep in mind that in RexMD a latence occurs at each replica exchange attempt (every 0.5 ps) because every replica-simulation needs to be completed before a new cycle starts. This means that the more replicas, the largest is the increase in total simulation time.

These results indicate that the challenge of SCREM is in finding an optimal trade-off: finding the amplitude of the Gaussian unit which, on one hand, can promote a frequent transition between substates in the Server and, on the other hand, a frequent exchange between the Server and the Client replicas. This choice is non trivial. An interplay with the short term memory function for which goal is to promote transition between large volumes in phase space (metavariables) is an attractive route for fine tuning. Table 6.II summarizes the relative

efficiency of the different methods used in Chapter 5 and 6 in term of effective computing days on a standard cluster

6.6 Discussion

SC-REM constitutes an extension of BP-RexMD (Chapter 5) in the subspace on nucleic acid backbone dihedrals by including self-regulation of competing replicas. The current approach is amenable to structure refinement of short nucleic acid fragments containing non-canonical backbone substates (in progress) as well as any other biomolecules of interest. Here I have tested the efficiency and the accuracy of the SCREM extension by comparing results obtained with a standard MD simulation, the TRexMD method and also the BPRexMD method using the same amount of statistics on a d(AA) test model which constitutes the simplest DNA system containing a complete set of backbone dihedral parameters.

Parameter tuning in SCREM is a non trivial trade-off, difficult to generalize at the moment. Possible outlooks include trying with different (increasing) height of unit Gaussians across the ladder of Server replicas, replacing the exponential expression by a linear kernel-like expression such as the one used in Chapter 5 and in (275), and applications to larger molecular systems to see if one set of parameters enjoys the trait of transferable properties for generalizing the method. Satisfactory results were obtained on the dinucleotide system. However first trials on larger DNA systems, an abasic site, have yet not lead to improved efficiency compared to BPRexMD and thus are not shown. Further tests in this direction are in progress. This discussion closes with a take home message for the development made in the current chapter. The SCREM method rests on the Tabu search (295,296) which is a general set-up for searching strategies in complex environment. SCREM relies on two principal assumptions:

1. The short term memory: *Dihedral angle variables can be discretized into intervals (coarse grained from 0° to 360°) and their alternative sampling can be seen as a non-equilibrium sampling between substates, which converges toward equilibrium (289-291) after many switches between these substates. An equilibrium situation is assumed when autocorrelation functions in each substate are stationary.*

2. The long term memory: *The inclusion of time dependent Gaussian functions preserves a canonical distribution in the unbiased replica because each Gaussian unit has an infinitesimal height (274, 293, 294) and the equation for exchange between replicas is derived from true probability distributions of conformations including every biasing potential (292).*

Acknowledgement

I acknowledge Martin Zacharias, Krystyna Zakrzewska, Richard Lavery, Martin Spitchy and Carlos Simmerling for useful discussions and comments on the subject of this chapter. Calculations carried-out with the computational resources of CLAMV at Jacobs University.

General Conclusion

The PhD thesis provides insights into the conformational flexibility of nucleic acids by molecular dynamics computer simulations and into the two main challenges of such approach, the force field problem and the sampling problem. Molecular dynamics is sometimes referred to as a computer experiment in that both scientific questions which are issued and interpretations of the results may be tackled in the same spirit as a physical experiment. Pragmatic responses given by molecular dynamics simulations to scientific questions are the ones which pass the grade of using different force fields and applying enhanced sampling techniques. However method development should not substitute motivated analysis and conclusion thereof from a molecular dynamics simulation. Hence important insights into the conformational flexibility of nucleic acids are resumed.

Our results give a better idea on the free energy cost for the curvature and the bending deformability of diverse nucleic acid motifs. Umbrella sampling simulations on the RNA kink-turn kt38 using the opening (or closing) of the key A-minor interaction as a reaction coordinate indicated an elbow motion which involves free energy changes of ~ 1.5 kcal/mol in favour of the closed form. The simulations indicated a global change in bending and twist of the C stem with respect to the NC stem correlated with the opening of the kt38 structure and controlled by the conformation of the second A-minor motif. The latter shifted from a type I to a type 0 in the meta-stable semi-open conformation. The fully open form observed also experimentally requires higher free energy changes (~ 4 kcal/mol) and is characterized by a complete disruption of A-minor motifs. The relatively small calculated free energy changes for k-turn opening/closing support the idea that k-turn motifs can act as flexible elements to mediate global changes of helical stem elements in large RNA containing molecules such as the ribosome.

A new restraint to induce DNA curvature, based on the definition of an average screw axis for the terminal segments of DNA oligo-nucleotides has been developed. This definition allows us to systematically investigate the bending deformability of various DNA molecules without any restrictions on the conformational relaxation of the DNA upon addition of the bending restraint. Results on A-tract containing sequences predict preferential bending towards the minor-groove and resulted in helical parameter changes in good agreement with experiment.

Results on central G:A or central C:C mismatches predict a significantly enlarged variety of possible structures that showed different directions and magnitudes of global bending compared to B-DNA. The altered bending flexibility may contribute to the recognition properties of modified DNA by repair enzymes. The approach was extended to study DNA bending during molecular dynamics simulations including surrounding ions and water explicitly. We found an optimal bending angle close to experiment for the different DNA sequences studied and a coupling of the global bending with other global variables characterizing the DNA. The fluctuations of these global variables did not change significantly under bending stress up to 100° of the current bending definition ($\sim 80 / 90^\circ$ in the standard definition of Curves).

By using the most recent Amber force field, parmbsc0, the energy change required to form a bend angle of 90° on 5 nm length scale is close to 5.0 kcal/mol for A-tract sequences and 6.0 kcal/mol for alternating [GC] or [AT] sequences which agrees with recent calculations of DNA bending driven by minor/major groove opening (80). For bending near 150° the free energy cost is on average 13 kcal/mol for all sequences with the exception of [A-tract-1] where it is reduced to 10.0 kcal/mol. We have found a linear (non-harmonic) energy dependence and extracted linear coefficients for different base sequences which can be used in Sub-Elastic Chain models (4). For the strong bending ($> 100^\circ$) the oligomer behavior depends on its sequence. [GC] and [Atract-1] produce the type II kink, [AT] denaturates and [Atract-2] responds to the imposed stress by a series of small kinks in the GC part of the sequence. The type II kink was observed twice, in [GC] and [Atract-1] oligomers. Type II kink in [GC] is associated with local bend toward major groove and smoother base pair deformation than in the case of [Atract-1] where it induces local bend towards the minor groove. Both are associated with a kink angle of roughly 90° between linear axes fitted to four base pairs on each side of the motif. Using the replica exchange umbrella sampling technique it was demonstrated that the type II kink motif is an elastic and thus reversible DNA motif. In contrast the type I kink was observed only with the parm94 force field and appeared as a plastic deformation due to localization of the strain energy along sequence, i.e. type I was not observed with REUS.

Our results show that in short oligomers individual junctions behave differently, in function of their sequence context, without necessity of invoking rare events such as severe base pair kink motifs. However these occur systematically at some particular di- or trinucleotide in DNA, in regime of strong bending. Harmonic bending stiffness modula are commonly used in the

elastic rod theory to model large scale dynamics of DNA molecules and one feature which emerged for every DNA kink (i.e. roll $> 20^\circ$) in these simulation, unexpectedly, is a lower bending stiffness constant (4/5 down to 1/10).

There was no significant transition in backbone torsion angles associated with DNA bending in chapter IV. Then a new Hamiltonian replica exchange method was designed in chapter V which uses a biasing potential as replica coordinate that specifically promotes dihedral transitions in the nucleic acid backbone (BP-RexMD). The conclusion there was that the application of the method to unusual DNA motifs, a central abasic site, is necessary to correctly describe the (ϵ / ζ) metastable conformations on short simulation time-scale (5 to 20 ns). In the umbrella sampling approach of chapter IV thus, several substates compatible with a bend conformation, due to barriers between the backbone conformational substates, may not have been exhaustively explored. The study reported in chapter V allows characterizing the fine structure of dsDNA systematically under realistic conditions and may allow also the systematic evaluation of the performance and precision of current force fields. Another possibility would be to use one of the other force fields available for molecular dynamics simulation of nucleic acids and see if the type II kink motif stands still within alternative force field parameters. If yes DNA kinks, according to the molecular dynamics simulation tool, would occur spontaneously in vivo, confirming a 20 years old theory developed by F. Crick and A. Klug.

It has been demonstrated that coupling of umbrella sampling simulations with a Hamiltonian replica exchange methodology significantly improves the convergence of calculated free energies and resulted in a closer match of forward and backward simulations compared to “standard” umbrella sampling. Finally, the Hamiltonian replica exchange formalism was revisited on the ground that its key advantage is to provide a Boltzman ensemble of conformation in the unbiased replica. The biased replicas were seen as searching copies aimed at avoiding kinetic trapping in the unbiased replica. The latter was seen as a client in a scheme where the remaining replica ladder would respond effectively as a function of the current state and history recorded in this client-replica. This lead to the development of an automatic scaling grounded on a Tabu search and defined by the integration of a short term memory and a long term memory functions which expand the Hamiltonian of some server-replicas. Despite effort on this front the ideal parameterization for general applications also on unusually complex systems such as an abasic site or a type II kink motif in DNA is still a prospect. Based on two-replicas test simulations on simple system however, the adaptive Hamiltonian

replica exchange (SCREM) can in principle be tuned into an efficient tool for exploring energetic landscapes associated to dihedral angle phase space of nucleic acids and of biomolecules in general. An improved sampling has been demonstrated throughout every chapter in case of the standard Hamiltonian replica exchange method using a biasing (BP-RexMD) or an umbrella (REUS) potential.

References

1. Bustamante C. (2008) In singulo biochemistry: when less is more. *Annu. Rev. Biochem.* **77**: 45-50.
2. Moffitt J.R., Chemla Y.R., Smith S.B., Bustamante C. (2008) Recent advances in optical tweezers. *Annu. Rev. Biochem.* **77**: 205-28.
3. Mulholland A.J. (2008) Introduction. Biomolecular simulation. *J. R. Soc. Interface* **3**: S169-72.
4. Wiggins P.A., Van Der Heijden T., Moreno-Herrero F., Spakowitz A., Phillips R., Widom J., Ceekers C. and Nelson P.C. (2006) High flexibility of DNA on short length scales probed by atomic force microscopy. *Nature Nanotechnology* **1**: 137-41.
5. Shroff H., Reinhard B.M., Siu M., Agarwal H., Spakowitz A. and Liphardt J. (2005) Biocompatible force sensor with optical readout and dimensions of 6 nm³. *Nano. Lett.* **5**: 1509-14.
6. Yuan C., Chen H., Lou X.W. and Archer L.A. (2008) DNA bending stiffness on small length scales. *Phys. Rev. Lett.* **100**: 018102.
7. Van der Kamp M.W., Shaw K.E., Woods C.J., Mulholland A.J. (2008) Biomolecular simulation and modelling: status, progress and prospects. *J R Soc Interface.* **3**: S173-90.
8. Born M. and Oppenheimer R. (1927) Zur Quantentheorie der Molekeln. *Ann. Phys.* **84**: 457-84.
9. (a) Chipot C., Shell M., Pohorille A. (2007) Introduction. pp. 1-33. And (b) Shell M. Panagiotopoulos A., Pohorille A. (2007) Methods Based on Probability Distributions and Histograms. pp. 77-119. In Chipot C., Pohorille A. (Eds.) *Free Energy Calculations*. Series: Springer Series in Chemical Physics, **86**.
10. Simonson T., Archontis G., Karplus M. (2002) Free energy simulations come of age: protein-ligand recognition. *Acc. Chem. Res.* **35**: 430-7.
11. Torrie G.M., Valleau J.P. (1974) Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.* **28**: 578-81.
12. Cloutier T.E. and Widom J. (2004) Spontaneous sharp bending of double-stranded DNA. *Molecular Cell.* **14**: 355-62.
13. Hansmann U.H. (1997) Parallel Tempering Algorithm for Conformational Studies of Biological Molecules. *Chem. Phys. Lett.* **281**: 140-50.
14. Fukunishi H., Watanabe O., Takada S. (2002) On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **116**: 9058-67

15. Sugita Y., Kitao A., Okamoto Y. (2000) Multidimensional replica-exchange method for free energy calculations. *J. Chem. Phys.* **113**: 6042-51.
16. Murata K., Sugita Y., Okamoto Y. (2004) Free energy calculations for DNA base stacking by replica-exchange umbrella sampling. *Chem. Phys. Lett.* **385**: 1-7.
17. Woods C. J., King M. A., Essex J. W. (2003) The Development of Replica-Exchange-Based Free-Energy Methods. *J. Phys. Chem.* **107**: 13703-10.
18. Norberg J., Nilsson L. (2004) Comment on: Free energy calculations for DNA base stacking by replica-exchange umbrella sampling by K. Murata and al. *Chem. Phys. Lett.* **393**: 282-3.
19. Perez A., Marchan I., Svozil D., Sponer J., Cheatham T.E. 3rd, Laughton C.A., Orozco M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.*, **92**: 3817-29.
20. Klein W., Thannhauser S.J. (1933) *Z. physiol. Chem.* **218**: 173 ; Klein W., Thannhauser S.J. (1934) *Z. physiol. Chem.* **224**: 252; Klein W., Thannhauser S.J. (1935) *Z. physiol. Chem.* **231**: 96.
21. Blackburn G.M., Gait M.J. (1996) *Nucleic Acids in Chemistry and Biology* 2nd edition. Oxford university press inc., New York, pp. 15-83.
22. Lavery R., Zakrzewska K., Sklenar H. (1995) JUMNA: Junction Minimisation of Nucleic Acids. *Comp. Phys. Comm.* **91**: 135-58.
23. Calladine C.R., Drew H., Luisi B., Travers A. (2004) *Understanding DNA: the Molecule and How it Works*. Academic Press, London, pp. 39-139.
24. Yakushevich L.V. (1998) *Nonlinear Physics of DNA*. Wiley-VC West Sussex, England. pp. 6-9.
25. Kudritskaya Z.G., Danilov V.I. (1976) Quantum mechanical study of bases interaction in various associates in atomic dipole approximation. *J. Theor. Biol.* **59**: 303-18.
26. Sponer J., Jurecka P., Marchan I., Luque F.J., Orozco M., Hobza P. (2006) Nature of base stacking: reference quantum-chemical stacking energies in ten unique B-DNA base-pair steps. *Chemistry* **12**: 2854-65.
27. Tan Z.J., Chen S.J. (2008) Electrostatic free energy landscapes for DNA helix bending. *Biophys. J.* **94**: 3137-49.
28. (a) Range K., Mayaan E., Maher L.J. 3rd, York D.M. (2005) The contribution of phosphate-phosphate repulsions to the free energy of DNA bending. *Nucleic Acids Res.* **33**: 1257-68. & (b) McDonald R.J., Dragan A.I., Kirk W.R., Neff K.L., Privalov P.L., Maher L.J. 3rd (2007) DNA bending by charged peptides: electrophoretic and spectroscopic analyse. *Biochemistry* **46**: 2306-16.

29. Han W., Dlakic M., Zhu Y.J., Lindsay S.M., Harrington R.E. (1997) Strained DNA is kinked by low concentrations of Zn²⁺. *Proc. Natl. Acad. Sci. U.S.A.* **94**: 10565-70.
30. Dickerson R.E., Ho-Leung Ng. (2001) DNA structure from A to B. *Proc. Natl. Acad. Sci. U.S.A.* **98**: 6986-88.
31. Calladine C.R. (1982) Mechanics of sequence-dependent stacking of bases in B-DNA. *J. Mol. Biol.* **161**: 343-52.
32. Arnott S., Chandrasekharan R., Birdsall D.L., Leslie A.G.W., Ratliff R.L. (1980) Left handed helices. *Nature* **283**: 743-45.
33. Lavery R., Sklenar H. (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* **6**: 63-91.
34. Lavery R., Sklenar H. (1989) Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.* **6**: 655-67.
35. Olson W.K., Bansal M., Burley S.K., Dickerson R.E., Gerstein M., Harvey S.C., Heinemann U., Lu X.J., Neidle S., Shakked Z., Sklenar H., Suzuki M., Tung C.S., Westhof E., Wolberger C., Berman H.M. (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **313**: 229-37.
36. Soumpasis D.M., Tung C.S. (1988) A rigorous basepair oriented description of DNA structures. *J. Biomol. Struct. Dyn.* **6**: 397-420.
37. Lu X.J., Babcock M.S., Olson W.K. (1999) Overview of nucleic acid analysis programs. *J. Biomol. Struct. Dyn.* **16**: 833-43.
38. Barbic A., Crothers D.M. (2003) Comparison of analyses of DNA curvature. *J. Biomol. Struct. Dyn.* **21**: 89-97.
39. Dickerson R.E. (1989) Definitions and nomenclature of nucleic acid structure parameters. *J. Biomol. Struct. Dyn.* **6**: 627-34.
40. McLachlan A.D. (1979) Gene duplications in the structural evolution of chymotrypsin. *J. Mol. Biol.* **128**: 49-79.
41. Goldstein H. (1981). *Classical Mechanics*. 2nd ed. Addison Wesley publishing Co., New York, pp: 1-80 for Kinematics and pp. 161 for the Rotation Formula.
42. Bhattacharya D., Bansal M. (1988) A general procedure for generation of curved DNA molecules. *J. Biomol. Struct. Dyn.* **6**: 93-104.
43. Schellman J.A., Harvey S.C. (1995) Static contributions to the persistence length of DNA and dynamic contributions to DNA curvature. *Biophys. Chem.* **55**: 95-114.
44. Manning R.S., Maddocks J.H., Kahn J.D. (1996) A Continuum Rod Model of Sequence-Dependent DNA Structure. *J. Chem. Phys.* **105**: 5626-46

45. Moakher M., Maddocks J.H. (2005) A double-strand elastic rod theory. *Archives for Rational Mechanics and Analysis* **177**: 53-91.
46. Chasles M. (1830) Note sur les propriétés générales du système de deux corps semblables entre eux et placés d'une manière quelconque dans l'espace; et sur le déplacement fini ou infiniment petit d'un corps solide libre. *Bulletin des Sciences Mathematiques, Astronomiques, Physiques et Chimiques* **14**: 321-26. [Notes on the general properties of a system of two identical bodies randomly located in space; and on the finite or infinitesimal motion of a free solid body].
47. Mazur J., Jernigan R.L. (1995) Comparison of rotation models for describing DNA conformations: Application to static and polymorphic forms. *Biophys. J.* **68**: 1472-89.
48. Babcock M.S., Pednault E.P., Olson W.K. (1994) Nucleic acid structure analysis. Mathematics for local Cartesian and helical structure parameters that are truly comparable between structures. *J. Mol. Biol.* **237**: 125-56.
49. Srinivasan R., Geetha V., Seetharaman J., Mohan S. (1993) A unique or essentially unique single parametric characterisation of biopolymeric structures. *J. Biomol. Struct. Dyn.*, **11**: 583-96.
50. Aumann F., Lankas F., Caudron M., Langowski J. (2006) Monte Carlo simulation of chromatin stretching. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* **73**: 041927.
51. Wedemann G., Langowski J. (2002) Computer simulation of the 30-nanometer chromatin fiber. *Biophys. J.* **82**: 2847-59.
52. Luger K., Mäder A.W., Richmond R.K., Sargent D.F., Richmond T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**: 251-60.
53. Gabrielian A., Pongor S. (1996) Correlation of intrinsic DNA curvature with DNA property periodicity. *FEBS Lett.* **393**: 65-8.
54. Zhurkin V.B. (1983) Specific alignment of nucleosomes on DNA correlates with periodic distribution of purine-pyrimidine and pyrimidine-purine dimers. *FEBS Lett.* **158**: 293-7.
55. Ong M.S., Richmond T.J., Davey C.A. (2007) DNA stretching and extreme kinking in the nucleosome core. *J. Mol. Biol.* **368**: 1067-74.
56. Satchwell S.C., Drew H.R., Travers A.A. (1986) Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191**: 659-75.
57. Segal E., Fondufe-Mittendorf Y., Chen L., Thåström A., Field Y., Moore I.K., Wang J.P., Widom J. (2006) A genomic code for nucleosome positioning. *Nature* **442**: 772-8.
58. Kiyama R., Trifonov E.N. (2002) What positions nucleosomes?--A model. *FEBS Lett.* **523**: 7-11.

59. Boffelli D., De Santis P., Palleschi A., Savino M. (1991) The curvature vector in nucleosomal DNAs and theoretical prediction of nucleosome positioning. *Biophys. Chem.* **39**: 127-36.
60. Ruscio J.Z., Onufriev A. (2006) A computational study of nucleosomal DNA flexibility. *Biophys. J.* **91**: 4121-32.
61. Roccatano D., Barthel A. and Zacharias M. (2007) Structural flexibility of the nucleosome core particle at atomic resolution studied by molecular dynamics simulation. *Biopolymers* **85**: 407-21.
62. Vaillant C., Audit B., Arneodo A. (2007) Experiments confirm the influence of genome long-range correlations on nucleosome positioning. *Phys. Rev. Lett.* **99**: 218103.
63. Vaillant C., Audit B., Thermes C., Arnéodo A. (2006) Formation and positioning of nucleosomes: effect of sequence-dependent long-range correlated structural disorder. *Eur. Phys. J. E. Soft Matter* **19**: 263-77.
64. Garcia H.G., Grayson P., Han L., Inamdar M., Kondev J., Nelson P.C., Phillips R., Widom J., Wiggins P.A. (2007) Biological consequences of tightly bent DNA: the other life of a macromolecular celebrity. *Biopolymers* **85**: 115-30.
65. Li G., Widom J. (2004) Nucleosomes facilitate their own invasion. *Nat. Struct. Mol. Biol.* **11**: 763-9.
66. Li G., Levitus M., Bustamante C., Widom J. (2005) Rapid spontaneous accessibility of nucleosomal DNA. *Nat. Struct. Mol. Biol.* **12**: 46-53.
67. Hagerman P.J. (1988) Flexibility of DNA. *Annu. Rev. Biophys. Chem.*, **17**: 265-86.
68. Fitzgerald D.J., Anderson J.N. (1999) DNA distortion as a factor in nucleosome positioning. *J. Mol. Biol.* **293**: 477-91.
69. Sobell H.M., Tsai C.C., Gilbert S.G., Jain S.C., Sakore T.D. (1976) Organization of DNA in chromatin. *Proc. Natl. Acad. Sci. U.S.A.* **73**: 3068-72.
70. Zhurkin V.B., Lysov Y.P., Ivanov V.I. (1979) Anisotropic flexibility of DNA and the nucleosomal structure. *Nucleic Acids Res.* **6**: 1081-96.
71. Wiggins P.A., Phillips R., Nelson P.C. (2005) Exact theory of kinkable elastic polymers. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* **71**: 021909.
72. Wiggins P.A. and Nelson P.C. (2006) Generalized theory of semiflexible polymers. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* **73**: 031906.
73. Yan J. and Marko J.F. (2004) Localized single-stranded bubble mechanism for cyclization of short double helix DNA. *Phys. Rev. Lett.* **93**: 108108.

74. Becker N.B., Wolff L., Everaers R. (2006) Indirect readout: detection of optimized subsequences and calculation of relative binding affinities using different DNA elastic potentials. *Nucleic Acids Res.* **34**: 5638-49.
75. Travers A., Muskhelishvili G. (2005) Bacterial chromatin. *Curr. Opin. Genet. Dev.* **15**: 507-14.
76. Zakrzewska K. (2003) DNA deformation energetics and protein binding. *Biopolymers* **70**: 414-23.
77. Rhodes D., Schwabe J.W., Chapman L., Fairall L. (1996) Towards an understanding of protein-DNA recognition. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **351**: 501-9.
78. Paillard G., Lavery R. (2004) Analyzing protein-DNA recognition mechanisms. *Structure* **12**: 113-22.
79. Privalov P.L., Dragan A.I., Crane-Robinson C., Breslauer K.J., Remeta D.P., Minetti C.A. (2007) What drives proteins into the major or minor grooves of DNA? *J. Mol. Biol.* **365**: 1-9.
80. Zacharias M. (2006) Minor groove deformability of DNA: a molecular dynamics free energy simulation study. *Biophys. J.* **91**: 882-91.
81. Schneider B., Berman H.M. (2008) Basics Of Nucleic Acid Structure: Concepts, Tools, and Archives In Sponer J., Lankas F. (Eds) *Computational studies of RNA & DNA*, Series: Challenges & advances in computational chemistry & physics, Volume 2, Springer pp. 1-44.
82. Brennan R.G., Matthews B.W. (1989) Structural basis of DNA-protein recognition. *Trends Biochem. Sci.* **14**: 286-90.
83. Rohs R., Sklenar H., Shakked Z. (2005) Structural and energetic origins of sequence-specific DNA bending: Monte-Carlo simulations of papillomavirus E2-DNA binding sites. *Structure* **13**: 1499-509.
84. Sreekanth R., Pattabhi V., Rajan S.S. (2008) Characterization of alpha helices interacting with nucleic acids. *Comput. Biol. Chem.* **32**: 378-81.
85. Somers W.S., Phillips S.E. (1992) Crystal structure of the met repressor-operator complex at 2.8 Å resolution reveals DNA recognition by beta-strands. *Nature* **359**: 387-93.
86. Schmiedeskamp M. and Klevit R.E. (1994) Zinc Finger Diversity. *Current Op. Struct. Biol.* **4**: 28-35.
87. Müller C.W., Rey F.A., Sodeoka M., Verdine G.L., Harrison S.C. (1995) Structure of the NF-kappa B p50 homodimer bound to DNA. *Nature* **373**: 311-7.
88. Matthews B.W. (1988) Protein-DNA interaction. No code for recognition. *Nature* **335**: 294-5.

89. Dickerson R.E. and Chiu T.K. (1997) Helix bending as a factor in protein/DNA recognition. *Biopolymers* **44**: 361-403.
90. Vilar J.M. and Saiz L. (2005) DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr. Opin. Genet. Dev.* **15**: 136-44.
91. Sandmann C., Cordes F., Saenger W. (1996) Structure model of a complex between the factor for inversion stimulation (FIS) and DNA: modeling protein-DNA complexes with dyad symmetry and known protein structures. *Proteins* **25**: 486-500.
92. Maurer S., Fritz J., Muskhelishvili G., Travers A. (2006) RNA polymerase and an activator form discrete subcomplexes in a transcription initiation complex. *EMBO J.* **25**: 3784-90.
93. Travers A., Muskhelishvili G. (1998) DNA microloops and microdomains: a general mechanism for transcription activation by torsional transmission. *J. Mol. Biol.* **279**: 1027-43.
94. Koo H.S., Drak J., Rice J.A., Crothers D.M. (1990) Determination of the extent of DNA bending by an adenine-thymine tract. *Biochemistry* **29**: 4227-34.
95. Hagerman P.J. (1985) Sequence dependence of the curvature of DNA: a test of the phasing hypothesis. *Biochemistry* **24**: 7033-37.
96. Koo H.S., Wu H.M., Crothers D.M. (1986) DNA bending at adenine.thymine tracts. *Nature* **320**: 501-06.
97. Hagerman P.J. (1986) Sequence-directed curvature of DNA. *Nature* **321**: 449-50.
98. Stefl R., Wu H., Ravindranathan S., Sklenar V., Feigon J. (2004) DNA A-tract bending in three dimensions: solving the dA4T4 vs. dT4A4 conundrum. *Proc. Natl. Acad. Sci. U.S.A.* **101**: 1177-82.
99. Wu H.M., Crothers D.M. (1984) The locus of sequence-directed and protein-induced DNA bending. *Nature* **308**: 509-13.
100. Arnott S., Chandrasekaran R., Hall I.H., Puigjaner L.C. (1983) Heteronomous DNA. *Nucleic Acids Res.* **11**: 4141-55.
101. Park Y.W., Breslauer K.J. (1991) A spectroscopic and calorimetric study of the melting behaviors of a "bent" and a "normal" DNA duplex: [d(GA4T4C)]₂ versus [d(GT4A4C)]₂. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 1551-55.
102. Selsing E., Wells R.D., Alden C.J., Arnott S. (1979) Bent DNA: visualization of a base-paired and stacked A-B conformational junction. *J. Biol. Chem.* **254**: 5417-22.
103. Trifonov E.N., Sussman J.L. (1980) The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proc. Natl. Acad. Sci. U.S.A.* **77**: 3816-20.

104. Ulanovsky L.E., Trifonov E.N. (1987) Estimation of wedge components in curved DNA. *Nature* **326**: 720-22.
105. Dickerson R.E., Goodsell D., Kopka M.L. (1996) MPD and DNA bending in crystals and in solution. *J. Mol. Biol.* **256**: 108-25.
106. Barbic A., Zimmer D.P., Crothers D.M. (2003) Structural origins of adenine-tract bending. *Proc. Natl. Acad. Sci. U.S.A.* **100**: 2369-73.
107. Bolshoy A., McNamara P., Harrington R.E., Trifonov E.N. (1991) Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 2312-16.
108. Lankas F., Lavery R., Maddocks J.H. (2006) Kinking occurs during molecular dynamics simulations of small DNA minicircles. *Structure* **14**: 1527-34.
109. Crothers D.M. (1998) DNA curvature and deformation in protein-DNA complexes: a step in the right direction. *Proc. Natl. Acad. Sci. U.S.A.* **95**: 15163-5.
110. Landau L.D., Lifshitz E.M. (1986) *Theory of Elasticity*, 3rd ed., Butterworth-Heinemann: Oxford, UK, pp. 1-87.
111. Antmann S.S. (1995). *Nonlinear Problems of Elasticity*. New York: Springer-. Verlag. pp. 479-587.
112. Lankas F. (2004) DNA sequence-dependent deformability--insights from computer simulations. *Biopolymers* **73**: 327-39.
113. Karplus M., Kushick J.N. (1981) Method for estimating the configurational entropy of macromolecules. *Macromolecules* **14**: 325-32.
- 114 Olson W.K., Gorin A.A., Lu X.J., Hock L.M., Zhurkin V.B. (1998) DNA sequence dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U.S.A.* **95**: 11163-8.
115. Djuranovic D. and Hartmann B. (2004) DNA fine structure and dynamics in crystals and in solution: the impact of BI/BII backbone conformations. *Biopolymers* **73**: 356-68.
116. Varnai P., Djuranovic D., Lavery R., Hartmann B. (2002) alpha/gamma Transitions in the B-DNA backbone. *Nucleic Acids Res.* **30**: 5398-406.
117. Packer M.J. and Hunter C.A. (1998) Sequence-dependent DNA structure: the role of the sugar-phosphate backbone. *J. Mol. Biol.* **280**: 407-20.
118. Heddi B., Foloppe N., Bouchemal N., Hantz E., Hartmann B. (2006). "Quantification of DNA BI/BII backbone sub-states in solution. Implications for DNA overall structure and recognition". *J. Am. Chem. Soc.* **128**: 9170-7.

119. Tisne C., Hantz E., Hartmann B., Delepierre M. (1998) Solution structure of a non-palindromic 16 base-pair DNA related to the HIV-1 kB site: evidence for BI-BII equilibrium inducing a global dynamic curvature of the duplex. *J. Mol. Biol.* **279**: 127-42.
120. Tisne C., Hartmann B., Delepierre M. (1999) Nf-kappa b binding mechanism: a nuclear magnetic resonance and modeling study of a GGG→CTC mutation, *Biochemistry* **38**: 3883-94.
121. Varnai P. and Zakrzewska K. (2004) DNA and its counterions: a molecular dynamics study. *Nucleic Acids Res.* **32**: 4269-80.
122. Beveridge D.L., Barreiro G., Byun K.S., Case D.A., Cheatham T.E. 3rd, Dixit S.B., Giudice E., Lankas F., Lavery R., Maddocks J.H. et al. (2004) Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys. J.* **87**: 3799-813.
123. Dixit S.B., Beveridge D.L., Case D.A., Cheatham T.E. 3rd, Giudice E., Lankas F., Lavery R., Maddocks J.H., Osman R., Sklenar H. et al. (2005) Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophys. J.* **89**: 3721-40.
124. Pérez A., Luque F.J., Orozco M. (2007) Dynamics of B-DNA on the. microsecond time scale. *J. Am. Chem. Soc.* **129**: 14739-45.
125. Crick F.H. and Klug A. (1975) Kinky helix. *Nature* **255**: 530-33.
126. Harris S.A., Laughton C.A., Liverpool T.B. (2008) Mapping the phase diagram of the writhe of DNA nanocircles using atomistic molecular dynamics simulations. *Nucleic Acids Res.* **36**: 21-9.
127. Ojha R.P., Dhingra M.M., Sarma M.H., Shibata M., Farrar M., Turner C.J., Sarma R.H. (1999) DNA bending and sequence-dependent backbone conformation NMR and computer experiments. *Eur. J. Biochem.* **265**: 35-53.
128. Elledge S.J. (1996) Cell Cycle Checkpoints: Preventing an Identity Crisis. *Science* **274**: 1664-72.
129. Demple B. and Harrison L. (1994). Repair of oxidative damage to DNA: enzymology and biology. *Annu. Rev. Biochem.* **63**: 915-48.
130. Fujimoto H., Pinak M., Nemoto T., O'Neill P., Kume E., Saito K., Maekawa H. (2005) Molecular dynamics simulation of clustered DNA damage sites containing 8-oxoguanine and abasic site. *J. Comput. Chem.* **26**: 788-98.
131. Ayadi L., Coulombeau C., Lavery R. (1999) Abasic sites in duplex DNA: molecular modeling of sequence-dependent effects on conformation. *Biophys. J.* **77**: 3218-26.

132. Li Y., Zon G., Wilson W.D. (1991) NMR and molecular modeling evidence for a G.A mismatch base pair in a purine-rich DNA duplex. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 26-30.
133. Nissen P., Ippolito J.A., Ban N., Moore P.B., Steitz T.A. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl. Acad. Sci. U.S.A.* **98**: 4899-903.
134. Murray L.J., Arendall W.B. 3rd, Richardson D.C., Richardson J.S. (2003) RNA backbone is rotameric. *Proc. Natl. Acad. Sci. U.S.A.* **100**: 3904-9.
135. Doherty E.A., Batey R.T., Masquida B., Doudna J.A. (2001) A universal mode of helix packing in RNA. *Nat. Struct. Biol.* **8**: 339-43.
136. Razga F., Koca J., Sponer J., Leontis N.B. (2005) Hinge-like motions in RNA kink-turns: the role of the second a-minor motif and nominally unpaired bases. *Biophys. J.* **88**: 3466-85.
137. Leontis N.B., Lescoute A., Westhof E. (2006) The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.* **16**: 279-87.
138. Tamura M., Hendrix D.K., Klosterman P.S., Schimmelman N.R., Brenner S.E., Holbrook S.R. (2004) SCOR: Structural Classification of RNA, version 2.0. *Nucleic Acids Res.* **32**(Database issue): D182-4.
139. Pley H.W., Flaherty K.M., McKay D.B. (1994) Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**: 68-74
140. Jovine L., Djordjevic S., Rhodes D. (2000) The crystal structure of yeast phenylalanine tRNA at 2.0 Å resolution: cleavage by Mg(2+) in 15-year old crystals. *J. Mol. Biol.* **301**: 401-14
141. Ban N., Nissen P., Hansen J., Moore P.B., Steitz T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**: 905-20.
142. Klein D.J., Schmeing T.M., Moore P.B., Steitz T.A. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J.* **20**: 4214-21.
143. Vidovic I., Nottrott S., Hartmuth K., Luhrmann R., Ficner R. (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol. Cell.* **6**: 1331-42.
144. Lescoute A., Leontis N.B., Massire C., Westhof E. (2005) Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res.* **33**: 2395-409.
145. Cojocaru V., Nottrott S., Klement R., Jovi T.M. (2005) The snRNP 15.5 K protein folds its cognate k-turn RNA: a combined theoretical and biochemical study. *RNA* **11**: 197-209.
146. Razga F., Spackova N., Reblova K., Koca J., Leontis N.B., Sponer J. (2004) Ribosomal RNA kink-turn motif- A flexible molecular hinge. *J. Biomol. Struct. Dyn.* **22**: 183-94.

147. Razga F., Zacharias M., Reblova K., Koca J., Sponer J. (2006) RNA Kink-Turns as Molecular Elbows: Hydration, Cation Binding, and Large-Scale Dynamics. *Structure* **14**: 1-11.
148. Tiedge H. (2006) K-Turn Motifs in Spatial RNA Coding. *RNA Biol.* **3**: 133-9.
149. Goody T.A., Melcher S.E., Norman D.G., Lilley D.M.J. (2004) The kink-turn motif in RNA is dimorphic, and metal ion-dependent. *RNA* **10**: 254-64.
150. Liu J. and Lilley D.M. (2006) The role of specific 22-hydroxyl groups in the stabilization of the folded conformation of kink-turn RNA. *RNA* **13**: 200-10.
151. Turner B., Lilley D. M. (2008) The importance of G:A hydrogen bonding in the metal ion- and protein-induced folding of kink turn RNA. *J. Mol. Biol.* **381**: 431-42.
152. Frank J., Spahn C.M.T. (2006) The ribosome and the mechanism of protein synthesis. *Rep. Prog. Phys.* **69**: 1383-417.
153. Gabashvili I.S., Agrawal R.K., Spahn C.M., Grassucci R.A., Svergun D.I., Frank J., Penczek P. (2000) Solution structure of the E. coli 70S ribosome at 11.5 Å resolution. *Cell.* **100**: 537-49.
154. Moore P.B., Steitz T.A. (2003) The Structural Basis of Large Ribosomal Subunit Function. *Annu. Rev. Biochem.* **72**: 813-50.
155. Frank J., Spahn C.M.T. (2006) The ribosome and the mechanism of protein synthesis. *Rep. Prog. Phys.* **69**: 1383-417.
156. Frauenfelder H., Wolynes P.G. (1994) Biomolecules: Where the Physics of Complexity and Simplicity Meet. *Physics today* **47**: 58-64.
157. Feynman R. (1963). *The Feynman lectures on physics*, vol. **1**. Readings / Tapes, MA: Addison-Wesley Publishing Company.
158. Board J.A., Kalé L.V.Jr., Schulten K., Skeel R.D., Schlick T. (1994) Modeling biomolecules: Larger scales, longer durations. *IEEE Computational Science & Engineering* **1**: 19-30.
159. Schlick T. (1996). Pursuing Laplace's Vision on Modern Computers. In Mesirov J.P., Schulten K., Sumners D.W. *Mathematical Applications to Biomolecular Structure and Dynamics*, IMA Volumes in Mathematics and Its Applications. **82**. New York: Springer-Verlag. pp. 218–247
160. Frauenfelder H., Sligar S.G., Wolynes P.G. (1991) The energy landscapes and motions of proteins. *Science* **254**: 1598-603.
161. Watson J. D., Crick F. H. C. (1953) Molecular structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**: 737-8.

162. Cohen A.E., Moerner W.E. (2007) Principal-components analysis of shape fluctuations of single DNA molecules. *Proc. Natl. Acad. Sci. U.S.A.* **104**: 12622-7.
163. Brooks B., Karplus M. (1985) Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* **82**: 4995-9.
164. Balasubramanian C., Ojha R.P., Maiti S. (2007) Hydration pattern of A4T4 and T4A4 DNA: a molecular dynamics study. *Biochem.Biophys.Res.Comm.* **355**: 1081-6.
165. McConnell K.J., Beveridge D.L. (2000) DNA structure: what's in charge? *J. Mol. Biol.* **304**: 803-20.
166. Kuznetsov S.V., Sugimura S., Vivas P., Crothers D.M., Ansari A. (2006) Direct Observation of DNA Bending/Unbending Kinetics in Complex with DNA-Bending Protein IHF, *Proc. Natl. Acad. Sci.* **103**: 18515-20.
167. Horenko I., Dittmer E., Lankas F., Maddocks J.H., Metzner Ph., Schütte Ch. (2008) Macroscopic Dynamics of Complex Metastable Systems: Theory, Algorithms, and Application to B-DNA. *J. Appl. Dyn. Syst.* **7**: 532-60.
168. Ryckaert J.P., Ciccotti G., Berendsen H.J.C. (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**: 327-41.
169. Alder B.J. and Wainwright T.E. (1957) Phase Transition for a Hard Sphere System. *J. Chem. Phys.* **27**: 1208-9.
170. Hawkins G.D., Cramer C.J., Truhlar D.J. (1995) Pairwise solute descreening of solute charges from a dielectric continuum. *Chem. Phys. Lett.* **246**: 122-9.
171. Leach A.R. (1996) *Molecular Modelling. Principles and Applications*. Addison Wesley, 1996. pp. 165-410.
172. Pérez A., Lankas F., Luque F.J., Orozco M. (2008) Towards a molecular dynamics consensus view of B-DNA flexibility. *Nucleic Acids Res.* **36**: 2379-94.
173. Vorobyov I.V., Anisimov V.M., MacKerell A.D. Jr. (2005) Polarizable empirical force field for alkanes based on the classical Drude oscillator model. *J. Phys. Chem. B.* **109**: 18988-99.
174. Walsh T.R., Liang T. (2008) A multipole-based water potential with implicit polarization for biomolecular simulations. *J. Comput. Chem.* ahead of print.
175. Roux B., Simonson T. (1999) Implicit solvent models. *Biophys Chem.* **78**: 1-20.
176. Röhrig U.F., Frank I., Hutter J., Laio A., VandeVondele J., Rothlisberger U. (2003) QM/MM Car-Parrinello molecular dynamics study of the solvent effects on the ground state and on the first excited singlet state of acetone in water. *Chem. Phys. Chem.* **4**: 1177-82.

177. Fletcher R. and Reeves C.M. (1964) Function minimization by conjugate gradients, *Comput. J.* **7**: 149-54.
178. Sklenar H., Wüstner D., Rohs R. (2006) Using internal and collective variables in Monte Carlo simulations of nucleic acid structures: chain breakage/closure algorithm and associated Jacobians. *J. Comput. Chem.* **27**: 309-15.
179. Leimkuhler B.J., Reich S., Skeel R.D. (1996) Integration methods for molecular dynamics. In J. P. Mesirov, K. Schulten and D. W. Sumners. *Mathematical Applications to Biomolecular Structure and Dynamics, IMA Volumes in Mathematics and Its Applications.* **82**. New York: Springer-Verlag. pp. 161–185.
180. Allen M.P., Tildesley D.J. (1987) *Computer Simulation of Liquids*. Clarendon Press, Oxford, pp. 117.
181. Reif F. (1965) *Fundamentals of statistical and thermal physics*, McGraw-Hill. pp. 54.
182. Haile J.M. (1992) *Molecular Dynamics Simulation. Elementary Methods*. Wiley, Chichester. pp. 38-104.
183. Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., DiNola A., Haak J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**: 3684-90.
184. Andersen H. C. (1983) Rattle: a “velocity” version of the Shake algorithm for molecular dynamics calculations. *J. Comp. Phys.* **52**: 24–34.
185. Ryckaert J.P. (1985) Special geometrical constraints in the molecular dynamics of chain molecules. *Mol. Phys.* **55**: 549-56.
186. van Gunsteren W.F., Berendsen H.J.C. (1977) Algorithms. for. macromolecular dynamics and constraint dynamics. *Mol. Phys.* **34**: 1-1327
187. van Gunsteren W.F., Karplus M. (1982) Effect of constraints on. the dynamics of macromolecules. *Macromolecules* **15**: 1528-44.
188. Darden T., Perera L., Li L., Pedersen L. (1999) New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure* **7**: R55-60.
189. Darden T., York D., Pedersen L. (1993) Particle mesh Ewald: an $N \cdot \log(N)$ method for computing Ewald sums. *J. Chem. Phys.* **98**: 10089-92.
190. Ewald P. (1921) Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **369**: 253-87.
191. Jorgensen W.L., Chandrasekhar J., Madura J.D., Impey R.W., Klein M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**: 926-35.

192. Andricioaei I. (2007) Specialized Methods for Improving Ergodic Sampling Using Molecular Dynamics and Monte Carlo Simulations. In Chipot C. and Pohorille A. (Eds.) *Free Energy Calculations Series: Springer Series in Chemical Physics*, **86** pp.277-322.
193. McQuarrie D.A. (2000) *Statistical Mechanics*, 2nd ed. University Science Books. pp. 467-593.
194. Kumar S., Rosenberg J.M., Bouzida D., Swendsen R.H., Kollman P.A. (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**: 1011-21.
195. Roux B. and Souaille M. (2001) Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comp. Phys. Comm.* **135**: 40-57.
196. Roux B. (1995) The calculation of the potential of mean force using computer simulations. *Comput. Phys. Comm.* **91**: 275-82.
197. Mitsutake A., Sugita Y., Okamoto Y. (2001) Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* **60**: 96-112.
198. Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E.J. (1953) Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **21**: 1087.
199. Zacharias M. (2008) Combining elastic network analysis and molecular dynamics simulation by Hamiltonian replica exchange. *J. Chem. Theory Comput.* **4**: 477-87.
200. Kwak W. and Hansmann U.H. (2005) Efficient sampling of protein structures by model hopping. *Phys. Rev. Lett.* **95**: 138102.
201. Chikenji G., Kikuchi M., Iba Y. (1999) Multi-Self-Overlap Ensemble for Protein Folding: Ground State Search and Thermodynamics. *Phys. Rev. Lett.* **83**: 1886-9.
202. Kannan S. and Zacharias M. (2007) Enhanced sampling of peptide and protein conformations using replica exchange simulations with a peptide backbone biasing-potential. *Proteins* **66**: 697-706.
203. Case D.A., Darden T.A., Cheatham T.E.III, Simmerling C.L., Wang J., Duke R.E., Luo R., Maerz K.M., Wang B., Pearlman D.A., Crowley M., Brozell S., Tsui V., Gohlke H., Mongan J., Hornak V., Cui G., Beroza P., Schafmeister C., Caldwell J.W., Ross W.S., Kollman P.A. (2004) *AMBER 8*. University of California, San Francisco.
204. Zacharias M., Hagerman P. J. (1995) Bulge-Induced Bends in RNA: Quantification by Transient Electric Birefringence. *J. Mol. Biol.* **247**: 486-500.
205. Travers A.A. (2004) The structural basis of DNA flexibility. *Phil. Trans. R. Soc. Lond. A.* **362**: 1423-38.
206. Deremble C. and Lavery R. (2005) Macromolecular recognition. *Curr. Opin. Struct. Biol.* **15**: 171-5.

207. von Hippel P.H., Rees W.A., Rippe K., Wilson K.S. (1996) Specificity mechanisms in the control of transcription. *Biophys. Chem.* **59**: 231-46.
208. Hunter C.A. (1993) Sequence-dependent DNA structure. The role of base stacking interactions. *J. Mol. Biol.* **230**: 1025-54.
209. Strahs D., and Schlick T. (2000) A-tract Bending: Insights into Experimental Structures by Computational Models. *J. Mol. Biol.* **301**: 643-63.
210. Sprous D., Young M.A., Beveridge D.L. (1999) Molecular dynamics studies of axis bending in d(G5-(GA4T4C)2-C5) and d(G5-(GT4A4C)2-C5): effects of sequence polarity on DNA curvature. *J. Mol. Biol.* **285**: 1623-32.
211. McConnell K.J., Beveridge D.L. (2001) Molecular dynamics simulations of B'-DNA: sequence effects on A-tract-induced bending and flexibility. *J. Mol. Biol.* **314**: 23-40.
212. Beveridge D.L., Dixit S.B., Barreiro G., Thayer K.M. (2004) Molecular dynamics simulations of DNA curvature and flexibility: helix phasing and premelting. *Biopolymers* **73**: 380-403.
213. Dixit S.B., Pitici F., Beveridge D.L. (2004) Structure and axis curvature in two dA6 x dT6 DNA oligonucleotides: comparison of molecular dynamics simulations with results from crystallography and NMR spectroscopy. *Biopolymers* **75**: 468-79.
214. Mazur A.K. (2006) Evaluation of elastic properties of atomistic DNA models. *Biophys. J.* **91**: 4507-18.
215. Ulyanov N.B., Zhurkin V.B. (1984) Sequence-Dependent Anisotropic Flexibility of B-DNA. A Conformational Study. *J. Biomol. Struct. Dyn.* **2**: 361-85.
216. Sanghani S.R., Zakrzewska K., Harvey S.C., Lavery R. (1996) Molecular modelling of (A4T4NN)*n* and (T4A4NN)*n*: sequence elements responsible for curvature. *Nucleic Acids Res.* **24**: 1632-7.
217. Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz K.M.Jr., Ferguson D.M., Spellmeyer D.C., Fox T., Caldwell J.W., Kollman P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **117**: 5179-97.
218. Pardo L., Campillo M., Bosch D., Pastor N., Weinstein H. (2000) Binding mechanisms of TATA box-binding proteins: DNA kinking is stabilized by specific hydrogen bonds. *Biophys. J.* **78**: 1988-96.
219. Chen C.Y., Ko T.P., Lin T.W., Chou C.C., Chen C.J., Wang A.H. (2005) Probing the DNA kink structure induced by the hyperthermophilic chromosomal protein Sac7d. *Nucleic Acids Res.* **33**: 430-8.
220. Zhang Y., McEwen A.E., Crothers D.M., Levene S.D. (2006) Statistical-mechanical theory of DNA looping. *Biophys. J.* **90**: 1903-12.

221. Stasiak A. and Maddocks J.H. (2000) Best packing in proteins and DNA. *Nature* **406**: 251-3.
222. Travers A. (2005) DNA dynamics: bubble 'n' flip for DNA cyclisation? *Curr. Biol.* **15**: R377-9.
223. Zhurkin V.B., Tolstorukov M.Y., Fei X., Colasanti A.V., Olson W.K. (2005) In Ohyama, T. (ed.), *DNA conformation and transcription*. Eureka.com.
224. Yan J. and Marko J.F. (2004) Localized single-stranded bubble mechanism for cyclization of short double helix DNA. *Phys. Rev. Lett.* **93**: 108108.
225. Yuan C., Rhoades E., Lou X.W., Archer L.A. (2006) Spontaneous sharp bending of DNA: role of melting bubbles. *Nucleic Acids Res.* **34**: 4554-60.
226. Von Kitzing E., Diekmann S. (1987) Molecular mechanics calculations of dA12.dT12 and of the curved molecule d(GCTCGAAAAA)4.d(TTTTTCGAGC)4. *Eur. Biophys. J.* **15**: 13-26.
227. Van Dam H., Louck J.D., Biederharn L.C. (1990) Rotation and angular momentum. In *Encyclopedia of Physics*. 2nd ed. Lerner, R.G. and Trigg, G.L. editors. VCH Publishers, New-York. pp. 1076-81.
228. Cheatham T.E., Cieplak P., Kollman P.A. (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* **16**: 845-62.
229. Zacharias M. (2001) Conformational analysis of DNA-trinucleotide-hairpin-loop structures using a continuum solvent model. *Biophys. J.* **80**: 2350-63.
230. MacDonald D., Herbert K., Zhang X., Pologruto T., Lu P. (2001) Solution structure of an A-tract DNA bend. *J. Mol. Biol.* **306**: 1081-98.
231. Hizver J., Rozenberg H., Frolov F., Rabinovich D., Shakked Z. (2001) DNA bending by an adenine--thymine tract and its role in gene regulation. *Proc. Natl. Acad. Sci. U.S.A.* **98**: 8490-5.
232. Rozenberg H., Rabinovich D., Frolov F., Hegde R.S., Shakked Z. (1998) Structural code for DNA recognition revealed in crystal structures of papillomavirus E2-DNA targets. *Proc. Natl. Acad. Sci. U.S.A.* **95**: 15194-9.
233. Kim S.S., Tam J.K., Wang A.F., Hegde R.S. (2000) The structural basis of DNA target discrimination by papillomavirus E2 proteins. *J. Biol. Chem.* **275**: 31245-54.
234. Hegde R.S., Grossman S.R., Laimins L.A., Sigler P.B. (1992) Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature* **359**: 505-12.

235. de Los Santos C., El-Khateeb M., Rege P., Tian K., Johnson F. (2004) Impact of the C1 configuration of abasic sites on DNA duplex structure. *Biochemistry* **43**: 15349-57.
236. Sanchez A.M., Volk D.E., Gorenstein D.G., Lloyd R.S. (2003) Initiation of repair of A/G mismatches is modulated by sequence context DNA Repair (*Amst*) **2**: 863-78.
237. Boulard Y., Cognet J.A., Fazakerley G.V. (1997) Solution structure as a function of pH of two central mismatches, C . T and C . C, in the 29 to 39 K-ras gene sequence, by nuclear magnetic resonance and molecular dynamics. *J. Mol. Biol.* **268**: 331-47.
238. Isaacs R.J., Rayens W.S., Spielmann H.P. (2002) Structural differences in the NOE-derived structure of G-T mismatched DNA relative to normal DNA are correlated with differences in (13)C relaxation-based internal dynamics. *J. Mol. Biol.* **319**: 191-207.
239. Giudice E., Varnai P., Lavery R. (2003) Base pair opening within B-DNA: free energy pathways for Gc and AT pairs from umbrella sampling simulations. *Nucleic Acids Res.* **31**: 1437-43.
240. Gonzalez O. and Maddocks J.H. (2001) Extracting parameters for base-pair level models of DNA from molecular dynamics simulations. *Theoretical Chemistry Accounts* **106**: 76-82.
241. Lankas F., Sponer J., Langowski J., Cheatham T.E. 3rd. (2003) DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.* **85**: 2872-83.
242. Djuranovic D., Hartmann B. (2005) Molecular Dynamics studies on free and bound targets of the bovine papillomavirus type I e2 protein: the protein binding effect on DNA and the recognition mechanism. *Biophys. J.* **89**: 2542-51.
243. Byun K.S., Beveridge D.L. (2004) Molecular dynamics simulations of papilloma virus E2 DNA sequences: dynamical models for oligonucleotide structures in solution. *Biopolymers* **73**: 369-79.
244. Shatzky-Schwartz M., Arbuckle N.D., Eisenstein M., Rabinovich D., Bareket-Samish A., Haran T.E., Luisi B.F., Shakked Z. (1997) X-ray and solution studies of DNA oligomers and implications for the structural basis of A-tract-dependent curvature. *J. Mol. Biol.* **267**: 595-623.
245. Mack D.R., Chiu T.K., Dickerson R.E. (2001) Intrinsic Bending and Deformability at the T-A Step of CCTTTAAAGG: A Comparative Analysis of TA and AT Steps within A-tracts. *J. Mol. Biol.* **312**: 10371049.
246. Barsky D., Foloppe N., Ahmadi S., Wilson III D.M., MacKerell Jr. A.D. (2000) New insights into the structure of abasic DNA from molecular dynamics simulations. *Nucleic Acids Res.* **28**: 2613-26.
247. Diekmann S., Mazzealli J.M., McLaughlin L.W., von Kitzing E., Travers A.A. (1992) DNA curvature does not require bifurcated hydrogen bonds or pyrimidine methyl groups. *J. Mol. Biol.* **225**: 729-38.

248. Gorin A.A., Zhurkin V.B., Olson W.K. (1995) B-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.* **247**: 34-48.
249. Packer M.J., Dauncey M.P., Hunter C.A. (2000) Sequence-dependent DNA structure: tetranucleotide conformational maps. *J. Mol. Biol.* **295**: 85-103.
250. Sarai A., Siebers J., Selvaraj S., Gromiha M.M., Kono H. (2005) Integration of bioinformatics and computational biology to understand protein-DNA recognition mechanism. *J. Bioinform. Comput. Biol.* **3**: 169-183.
251. Bertrand H., Ha-Duong T., Femandjian S., Hartmann B. (1998) Flexibility of the B-DNA backbone: effects of local and neighbouring sequences on pyrimidine-purine steps. *Nucleic Acids Res.* **26**: 1261-7.
252. Scipioni A., Anselmi C., Zuccheri G., Samori B., De Santis P. (2002) Sequence-dependent DNA curvature and flexibility from scanning force microscopy images. *Biophys. J.* **83**: 2408-18.
253. Okonogi T.M., Alley S.C., Reese A.W., Hopkins P.B., Robinson B.H. (2002) Sequence-dependent dynamics of duplex DNA: the applicability of a dinucleotide model. *Biophys. J.* **83**: 3446-59.
254. Dichmann D.J., Maddocks J.H., Pego R.L. (1996) Hamiltonian dynamics of an elastica and the stability of solitary waves. *Arch. Rational Mech. Anal.* **135**: 357-96.
255. Komarova N.L., Soffer A. (2005) Nonlinear waves in double-stranded DNA. *Bull. Math. Biol.* **67**: 701-18.
256. van Erp T.S., Cuesta-López S., Peyrard M. (2006) Bubbles and denaturation in DNA. *Eur. Phys. J. E. Soft Matter* **20**: 421-34.
257. Sugita Y. and Okamoto Y. (1999) Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**: 141-51.
258. Zhou R. (2007) Replica exchange molecular dynamics method for protein folding simulation. *Methods Mol. Biol.* **350**: 205-23.
259. Rhee Y.M. and Pande V.S. (2003) Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys. J.* **84**: 775-86.
260. Eleftheriou M., Rayshubski A., Pitera J.W., Fitch B.G., Zhou R. and Germain R.S. (2006) Parallel implementation of the replica exchange molecular dynamics algorithm on Blue Gene/L. *D.O.I.: 10.1109/IPDPS.2006.1639538 (on line)*.
261. Nadler W. and Hansmann U.H. (2007) Dynamics and optimal number of replicas in parallel tempering simulations. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* **76**: 065701.
262. Kelso C. and Simmerling C. (2006) Enhanced Sampling Methods for Simulation of Nucleic Acids. In Sponer, J. and Lankas, F. (editors), *Computational Studies of RNA and DNA*, Springer publishers, Netherlands, pp. 147-168.

263. Kannan S. and Zacharias M. (2007) Folding of a DNA hairpin loop structure in explicit solvent using replica-exchange molecular dynamics simulations *Biophys. J.* **93**: 3218-28.
264. Periolo X. and Mark A.E. (2007) Convergence and sampling efficiency in replica exchange simulations of peptide folding in explicit solvent. *J. Chem. Phys.* **126**: 014903.
265. Liu P., Kim B., Friesner R.A., Berne B.J. (2005) Replica exchange with solute tempering: a method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 13749-54.
266. Huang X., Hagen M., Kim B., Friesner R.A., Zhou R., Berne B.J. (2007) Replica Exchange with Solute Tempering: Efficiency on Large Scale Systems. *J. Phys. Chem.* **111**: 5405-10.
267. Affentranger R., Tavernelli I., Di Iorio E.E. (2006) A Novel Hamiltonian Replica Exchange MD Protocol to Enhance Protein Conformational Space Sampling. *J. Chem. Theory Comput.* **2**: 217-28.
268. Faraldo-Gómez J.D. and Roux B. (2006) Characterization of Conformational Equilibria Through Hamiltonian and Temperature Replica-Exchange Simulations: Assessing Entropic and Environmental Effects. *J. Comput. Chem.* **00** D.O.I.: 10.1002/jcc.20652 (on line).
269. Mu Y., Yang Y., Xu W. (2007) Hybrid Hamiltonian replica exchange molecular dynamics simulation method employing the Poisson-Boltzmann model. *J. Chem. Phys.* **127**: 084119.
270. Chen J., Dupradeau F.Y., Case D.A., Turner C.J., Stubbe J. (2007) Nuclear magnetic resonance structural studies and molecular modeling of duplex DNA containing normal and 4'-oxidized abasic sites. *Biochemistry* **46**: 3096-107.
271. Huber T., Torda A.E., vanGunsteren W.F. (1994) Local Elevation: A method for improving the searching properties of molecular dynamics simulation. *J-CAMD* **8**: 695-708.
272. Laio A. and Parrinello M. (2002) Escaping free-energy minima. *PNAS.* **99**: 12562-6.
273. Raiteri P., Laio A., Gervasio F.L., Micheletti C., Parrinello M. (2006) Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *J. Phys. Chem.* **110**: 3533-9.
274. Piana S. and Laio A. (2007) A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem.* **111**: 4553-9.
275. Babin V., Roland C., Sagui C. (2008) Adaptively biased molecular dynamics for free energy calculations. *J. Chem. Phys.* **128**: 134101.

276. Bartels C. and Karplus M. (1997) Probability Distribution for Complex Systems: Adaptive Umbrella Sampling of the Potential Energy. *J. Comput. Chem.* **18**: 865-80.
277. Berg B.A., Neuhaus T. (1991) Multicanonical algorithms for first order phase transitions. *Phys. Lett. B.* **B267**: 249-53.
278. Hansmann U.H.E., Okamoto Y. (1994) Comparative study of multicanonical and simulated annealing algorithms in the protein folding problem. *Physica A.* **212**: 415-37.
279. Wang F., Landau D.P. (2001) Efficient, Multiple-Range RandomWalk Algorithm to Calculate the Density of State. *Phys. Rev. Lett.* **86**: 2050-3.
280. Darve E., Pohorille A. (2001) Calculating free energies using average force. *J. Chem. Phys.* **115**: 9169-83.
281. Chipot C., Henin J. (2005) Exploring the free-energy landscape of a short peptide using an average force. *J. Chem. Phys.* **123**: 244906.
282. Bussi G., Laio A., Parrinello M. (2006) Equilibrium Free Energies from Nonequilibrium Metadynamics. *Phys. Rev. Lett.* **96**: 090601.
283. Marsili S., Barducci A., Chelli R., Procacci P., Schettino V. (2006) Self-healing Umbrella Sampling: A Non-equilibrium Approach for Quantitative Free Energy Calculations. *J. Phys. Chem.* **110**: 14011-3.
284. Voter A.F. (1997) Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events. *Phys. Rev. Lett.* **78**: 3908.
285. Andricioaei I., Straub J.E. (1996) Generalized simulated annealing algorithms using Tsallis statistics: application to conformational optimization of a tetrapeptide of special interest. *Phys. Rev. E.* **53**: R3055-R3058.
286. Gao Y.Q., Zang L. (2006) On the enhanced sampling over energy barriers in molecular dynamics simulations. *J. Chem. Phys.* **125**: 114103.
287. Hamelberg D., Mongan J., McCammon J.A. (2004) Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **120**: 11919-29.
288. Bitetti-Putzer R., Dinner A.R., Yang W., Karplus M. (2006) Conformational sampling via a self-regulating effective energy surface. *J. Chem. Phys.* **124**: 174901.
289. Jarzynski C. (1997) Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **78**: 2690.
290. Crooks G.E. (1999) Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E.* **60**: 2721-6.
291. Maragakis P., Spichty M., Karplus M. (2006) Optimal estimates of free energies from multi-state nonequilibrium work data. *Phys. Rev. Lett.* **96**: 100602.

292. Roitberg A.E., Okur A., Simmerling C. (2007) Coupling of Replica exchange Simulations to a Non-Boltzmann Structure Reservoir. *J. Phys. Chem.* **111**: 2415-8.
293. Bussi G., Gervasio F.L., Laio A., Parrinello M. (2006) Free-Energy Landscape for β Hairpin Folding from Combined Parallel Tempering and Metadynamics. *J. Am. Chem. Soc.* **128**: 13435-41.
294. Camilloni C., Provasi D., Tian G., Broglia R.A. (2007) Exploring the protein G helix free-energy surface by solute tempering metadynamics. *Proteins* **71**: 1647-54.
295. Glover F. (1989) Tabu Search - Part I. *ORSA Journal on Computing* **1**: 190-206.
296. Glover F. (1990) Tabu Search - Part II. *ORSA Journal on Computing* **2**: 4-32.

List of publications

Magnitude and direction of DNA bending induced by screw-axis orientation: Influence of Sequence, Mismatches and Abasic Sites

Curuksu, J., Zakrzewska, K., Zacharias, M.

Nucleic Acids Research (2008) **36**: 2268-83.

The elbow flexibility of the kt38 RNA kink turn motif from free energy molecular dynamics simulations

Curuksu, J., Sponer, J., Zacharias, M.

Submitted to *RNA* on the 15 of December 2008

Local and global effects of strong DNA bending induced during molecular dynamics simulations

Curuksu, J., Zacharias, M., Lavery, R., Zakrzewska, K.

Submitted to *Nucleic Acid Research* on the 22 of December 2008

Enhanced Conformational Sampling of Nucleic Acids by Hamiltonian Replica Exchange
Molecular Dynamics

Curuksu, J. and Zacharias, M.

The Journal of Chemical Physics (in press)

ANNEXE I

Complete analysis of DNA bending parmbsc0 trajectories (chapter IV)

Figure 1.

Angular direction of DNA global bend angle with respect to dyad axis of the central base pair. Average is given for each window of the umbrella sampling and x-axis is the reaction coordinate. Bending direction is calculated as an angle between two projection planes. First plane is defined by the dyad axis and helical axis of central base pair. Second plane is defined by the two ends of best fit curvilinear axis produced by Curves for the whole oligomer (two last helical axis excluding terminal base pairs). An angle equal to zero corresponds to global bending in direction of minor groove located at the middle path of DNA oligomer.

Figure 2.

Induced persistence length P for the four oligomers studied as a function of restrained bending. P is calculated as $P = 2 * \text{contour length (in base pair steps)} / \langle \sigma \rangle$ where $\langle \sigma \rangle$ is the local bend angle variance calculated with Curves for each umbrella window. For unbiased comparison between DNA oligomers studied, 0° is always taken as bend angle reference value. Upper-left: [GC], upper-right: [AT], lower-left: [Atract-1] and lower-right: [Atract-2]. Data points were averaged over group of three consecutive sampling windows for clarity.

Figure 3.

Schematic illustration of base-pair kink motifs observed in current study. Opaque bond-sticks represent linear axis calculated with Curves for four base pairs on each side of the motif (except one shorter line in case of motif **3** since it is located two junctions far from terminal base pair). **1**: *Type II kink* directed toward local minor groove at AAC associated to DNA bending of 88.6° , **2**: *Type II kink* directed toward local major groove at GCG associated to

DNA bending of 91.3° , **3: *Breather*** (flipped base) directed toward local major groove at AT associated to DNA bending of 24.3° , **4: *Kink*** ‘ $+45^\circ$ ’ directed toward local major groove at CG associated to DNA bending of 66.3° , **5: *Kink*** ‘ $+20^\circ$ ’ directed toward local major groove at CA associated to DNA bending of 30.4° and **6: *Kink*** ‘ -20° ’ directed toward local minor groove at CG associated to DNA bending of 38.2° . Snapshots were captured along stable sampling trajectory of motifs discussed in the text.

Figure 4.

Time-series, averages and mean-square deviations of helical roll (first column) and tilt (second column) parameters for trinucleotide $A_9A_{10}C_{11}$ / $G_{20}T_{21}T_{22}$ of [Atract-1] oligomer. The third column corresponds to propeller twist of central $A_{10}:T_{21}$ base-pair. The x-axis corresponds to the bending coordinate used in umbrella sampling.

Figure 5.

Base pair kink density histogram observed for high DNA bending strain with Amber parmbsc0 force-field. Base pair bend angle values were cumulated over four adjacent umbrella windows (135° - 150°) in forward simulation. First set of plots corresponds to roll fluctuation for every YpR (first row plots) and RpY (second row plots) base pair junctions in oligomers [GC] and [AT]. Second set of plots gives roll fluctuation for dinucleotides inside CAAAAAC (first row plots, see detailed legend on the right-hand side) and inside the two sequences flanking the A-tract block (second row plots), in case of oligomers [Atract-1] and [Atract-2]. Each color represents one independent DNA junction indicated through a color code on the right hand side of the figure.

Figure 6.

Breakdown of the reaction coordinate used to induce DNA bending in case of the [AT] oligomer. Shown in Y-axis is the average distance in each umbrella window between the two strands in the first half of the oligomer (dashed line, right hand side legend), taken as the distance between group of atoms on the bases that are used to define the screw vector. Also shown (full line, left hand side legend) is the opening angle averaged over the two first base pairs of the [AT] oligomer, in each umbrella window. The bending angle is in X-axis. The breakdown is at ~ 80 degrees for both order parameters.

Table I. Simulation parameters of umbrella sampling Molecular Dynamics simulation on d(CGCGCGCGCGCGCGC), d(CATATATATATATATC), d(CGCGCAAAAACGCGC) and d(CGCGCGCGCAAAAAC) DNA oligomers referred respectively as [GC], [AT], [Atract-1] and [Atract-2] sequences.

Table II. Summary of base pair kink motif occurrences during umbrella sampling simulation of four DNA oligomers. For each sequence, kink motif positions as well as the bending regime corresponding to the onset of each specific unstacking transition is reported based on observed density of kinked base pairs (see full histograms in supplementary material). Distribution centers are clustered into three types, either “-20°”, “+20°” or “+45°”. Base pair position is indicated through an index for first nucleotide step. Bending angle in bracket refer to upper boundary of bending range where corresponding motif first emerges.

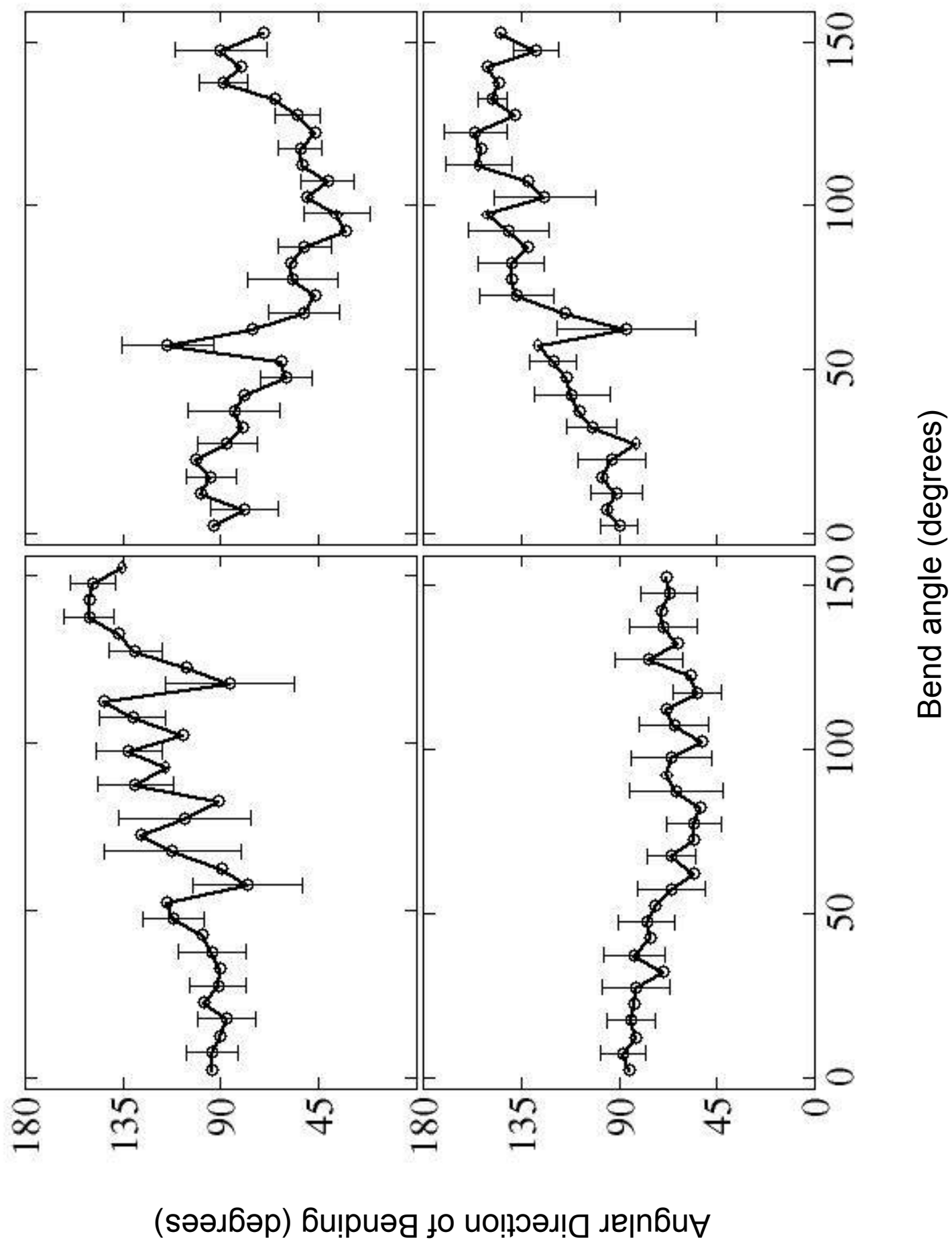


Figure 1

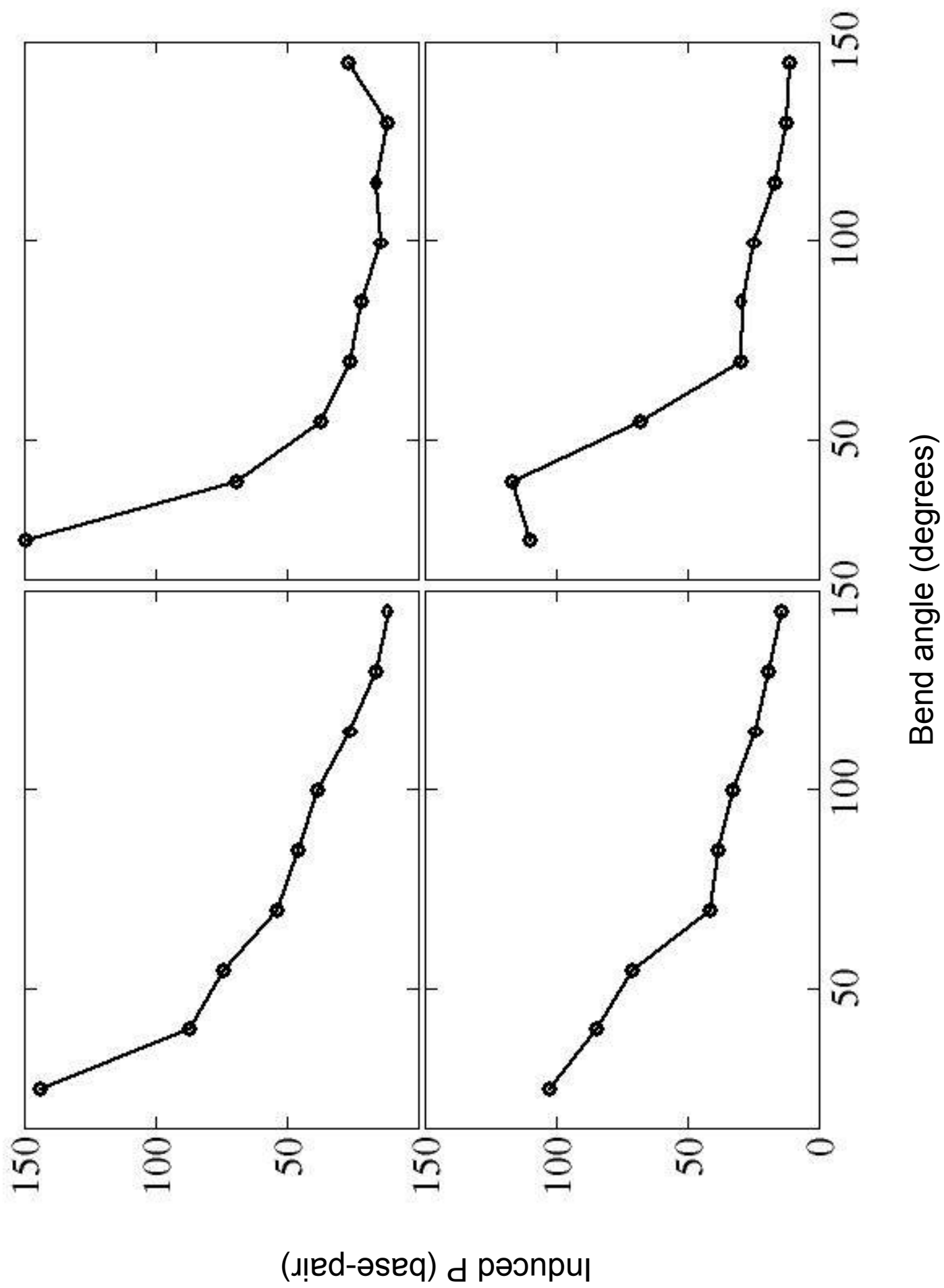
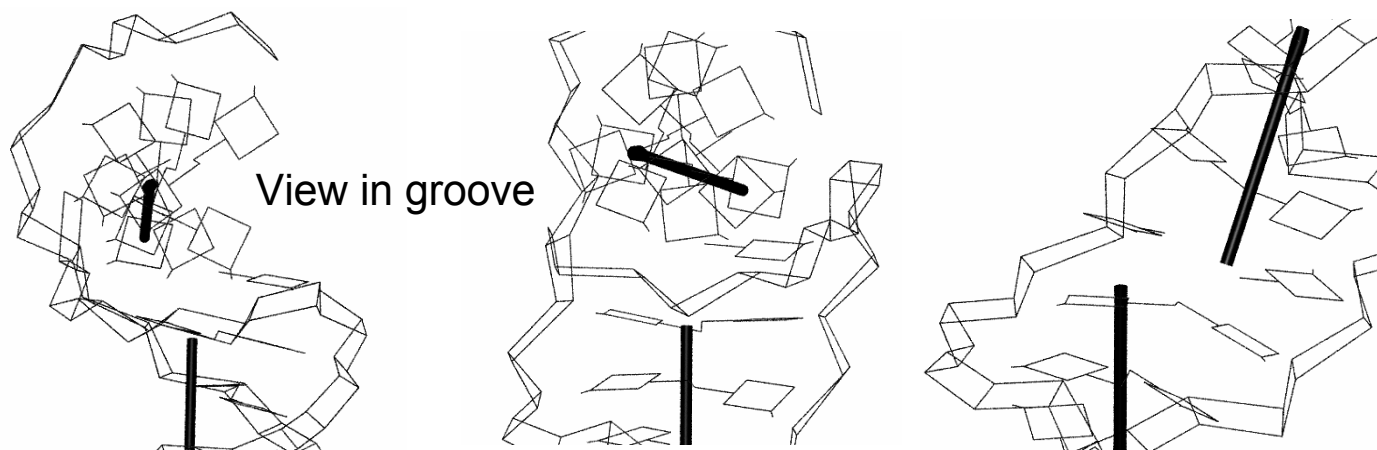
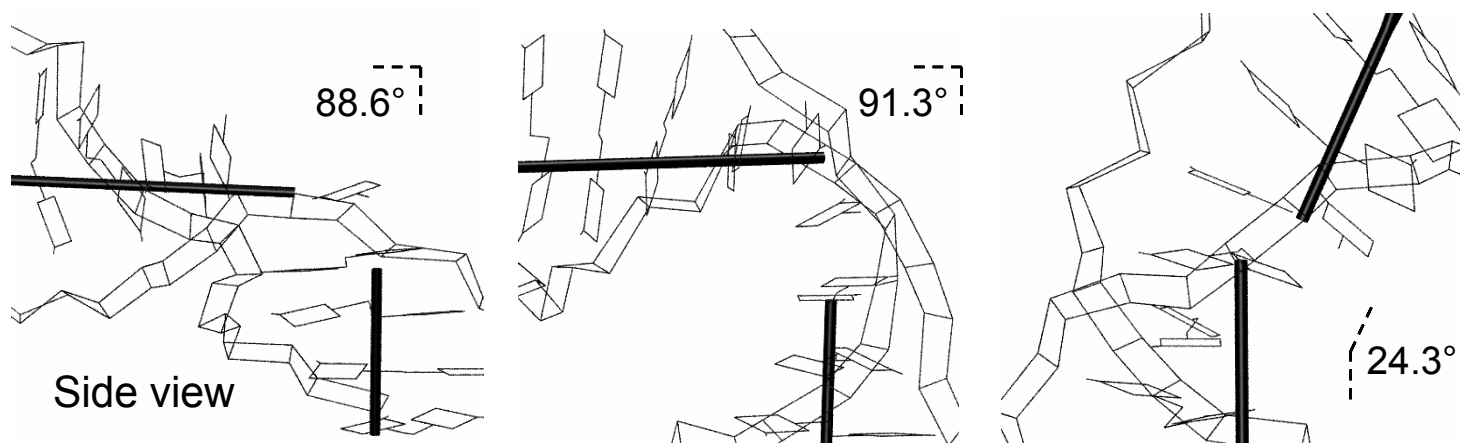


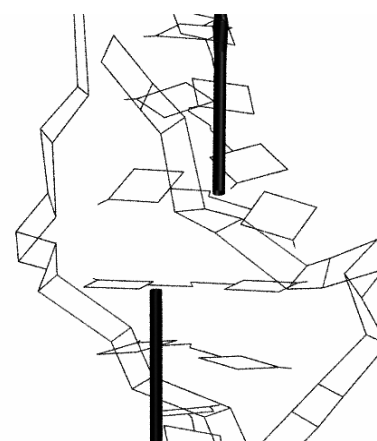
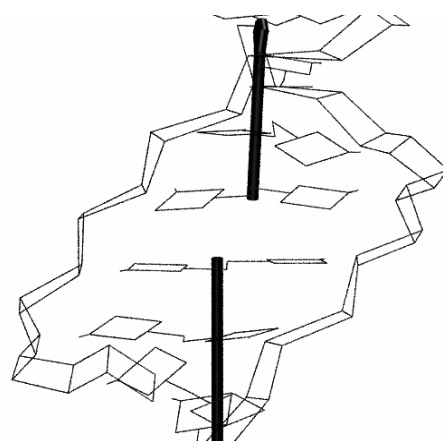
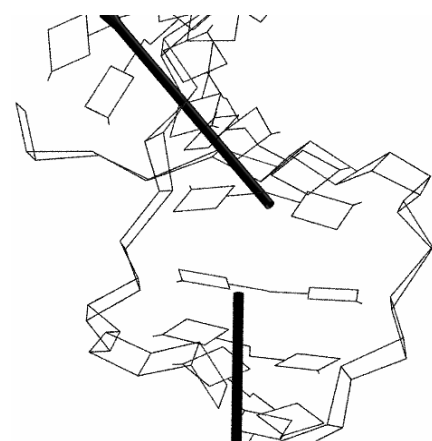
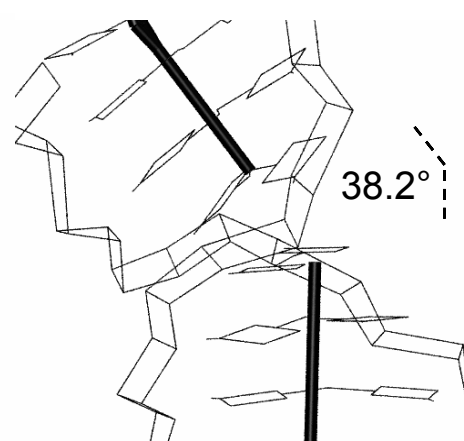
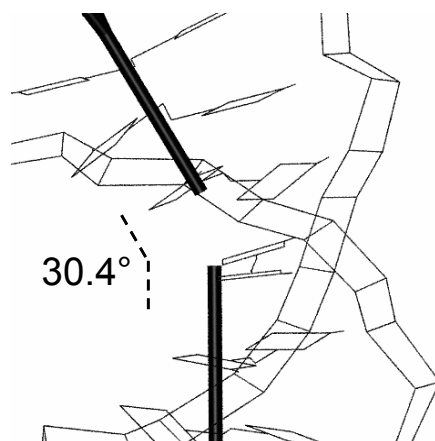
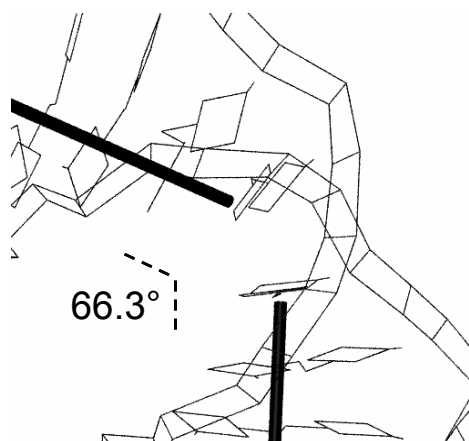
Figure 2



1

2

3



4

5

6

Figure 3

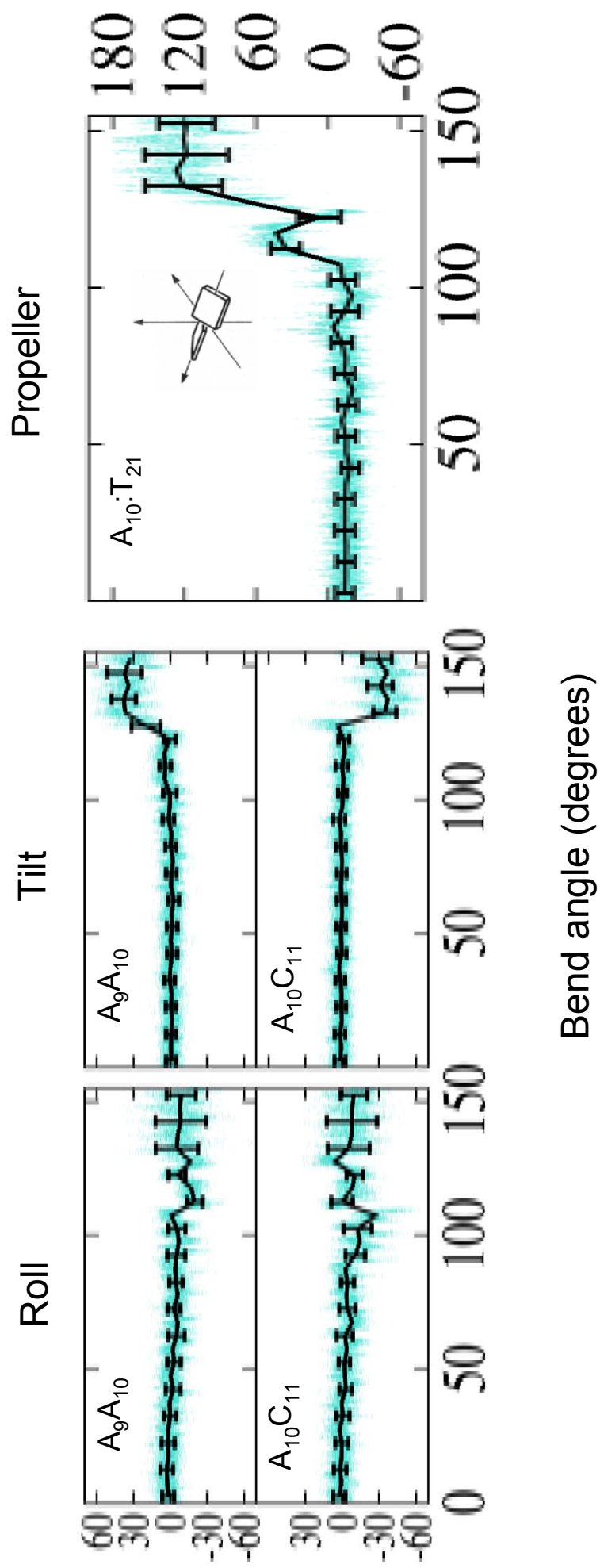


Figure 4.

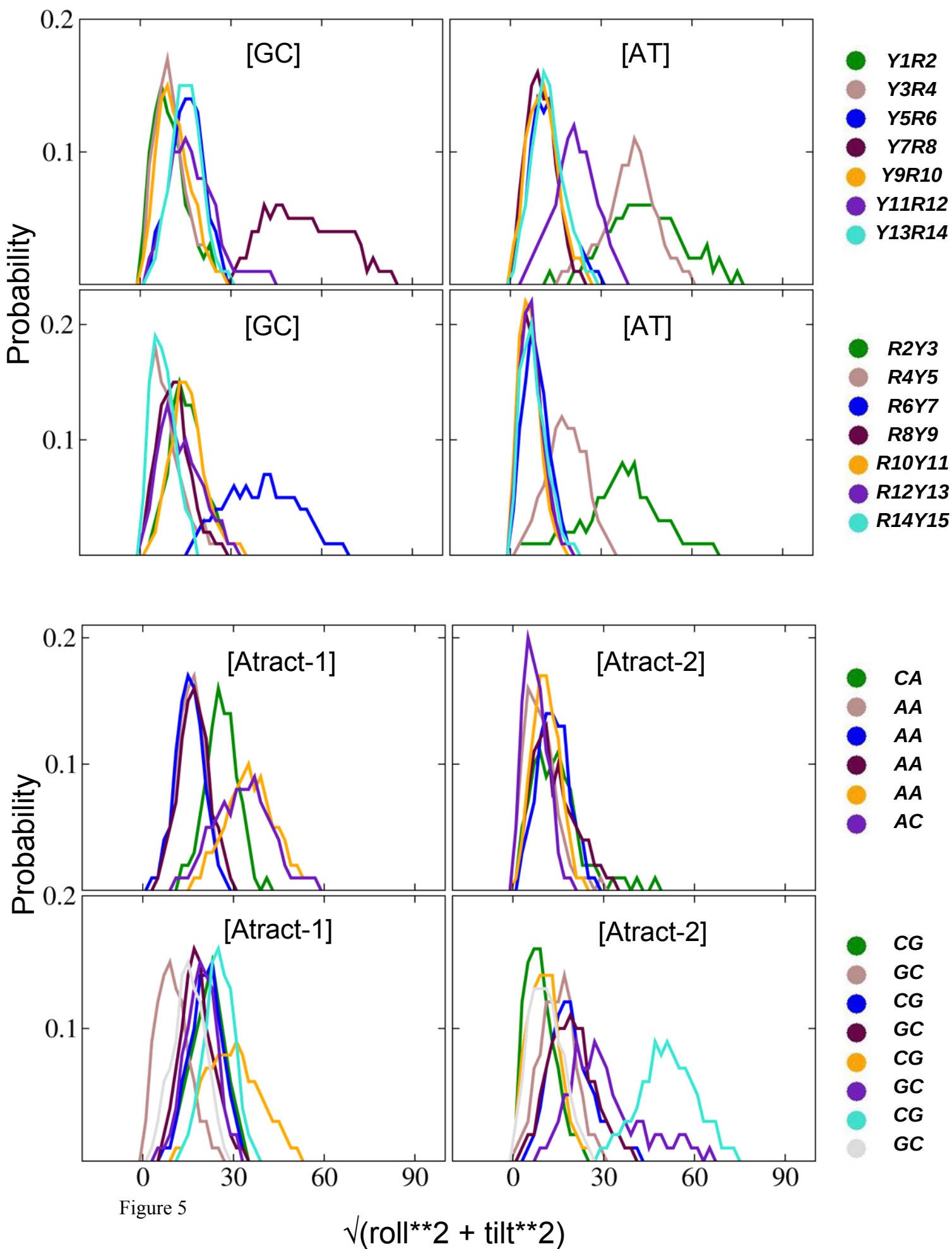


Figure 5

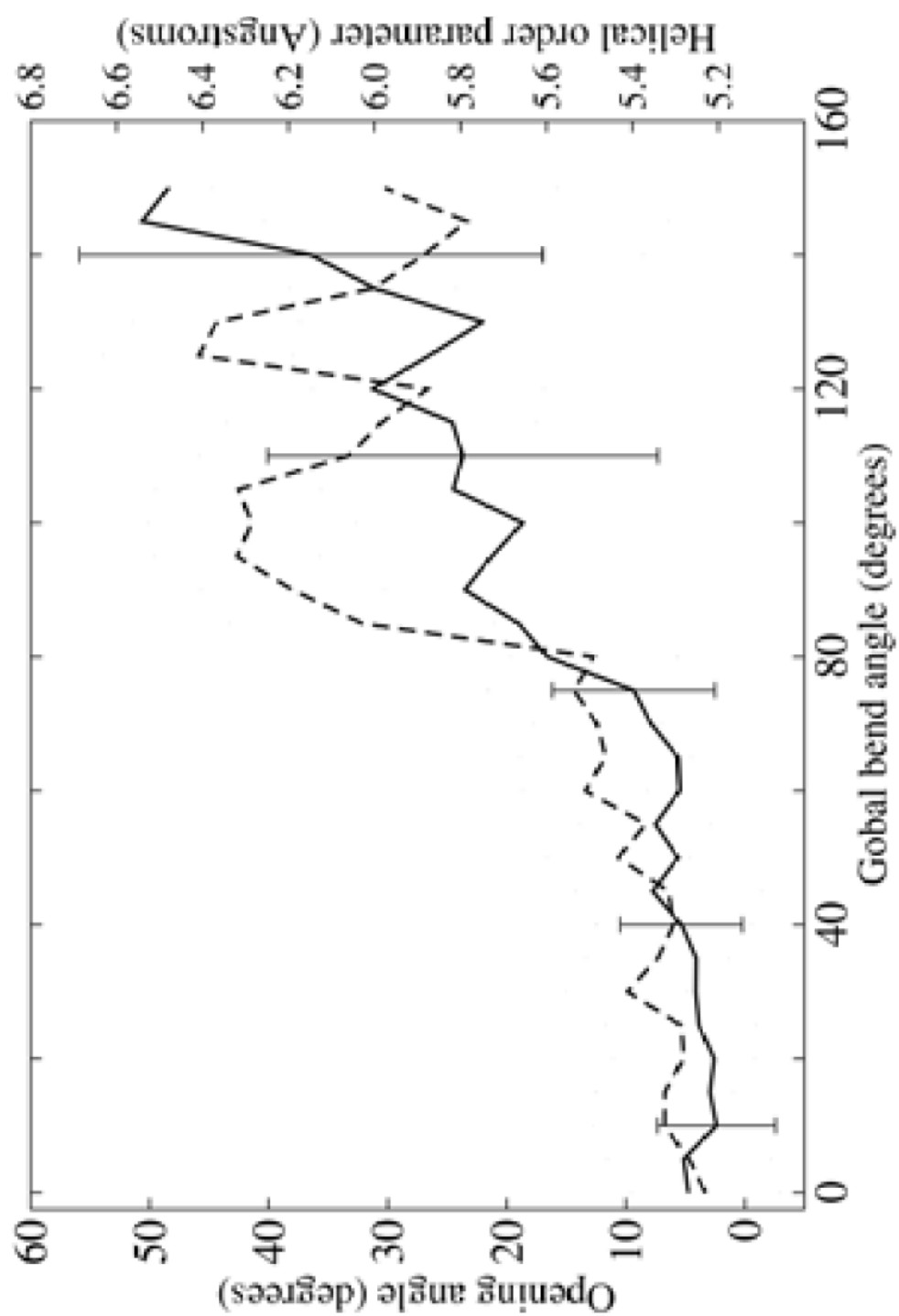


Figure 6.

DNA Sequence	Simulation Length	Bending regime	Temperature	DNA sequence	
[GC]	31 * 3ns (forward) 31 * 1ns (backward)	0-150°	300 K	d (CGCGCGCGCGCGCGC/ GCGCGCGCGCGCGCG)	
[AT]	31 * 3ns (forward) 31 * 1ns (backward)	0-150°	300 K	d (CATATATATATATAC/ GTATATATATATATG)	
[Atract-1]	31 * 3ns (forward) 31 * 1ns (backward)	0-150°	300 K	d (CGCGCAAAAACGCGC/ GCGCGTTTTTGCGCG)	
[Atract-2]	31 * 3ns (forward) 31 * 1ns (backward)	0-150°	300 K	d (CGCGCGCGCAAAAAC/ GCGCGCGCGTTTTTG)	

Table I.

	Base-pair junction with kink motif (Umbrella Sampling windows)			Total	
	-20°	$+20^\circ$	$+45^\circ$	$ 20^\circ $	$+45^\circ$
<i>[GC]</i>	$C_{11}G (150^\circ)$ $G_{10}C (120^\circ)$	$C_7G (70^\circ)$	$C_7G (140^\circ)$ $G_6C (110^\circ)$	3	2
<i>[AT]</i>	$T_1A (150^\circ)$	$T_{11}A (70^\circ)$ $T_{13}A (110^\circ)$ $A_4T (150^\circ)$	$T_3A (150^\circ)$ $A_2T (125^\circ)$	4	2
<i>[Atract-1]</i>	$A_{10}C (110^\circ)$	$C_5A (50^\circ)$ $A_{6-8}A (150^\circ)$ $C_{13}G (110^\circ)$ $C_{3,11}G (150^\circ)$ $G_{4,12}C (150^\circ)$	-	10	0
<i>[Atract-2]</i>	$A_{12}A (130^\circ)$ $G_4C (90^\circ)$ $C_5G (90^\circ)$	$C_{10}A (70^\circ)$ $G_8C (90^\circ)$	$C_7G (130^\circ)$	5	1

Table II.

MORE...

Figure A.

Time-series, averages and mean square deviations of helical roll (first column), tilt (second column), twist (third column) and propeller twist (fourth column) parameters (in y-axis) for each of the ten central base pair step (base pair 3 to 12 for propeller twist). Forward/Backward simulation corresponding to oligomers [GC], A1/A2, [AT], B1/B2, [Atract-1], C1/C2 and [Atract-2], D1/D2. Each row corresponds to one DNA junction. In case of [Atract-2], rotational parameters are reported for junctions 5 to 14 to include the whole A-tract block in the analyse. The x-axis corresponds to bending angle coordinate used in umbrella sampling.

Figure B.

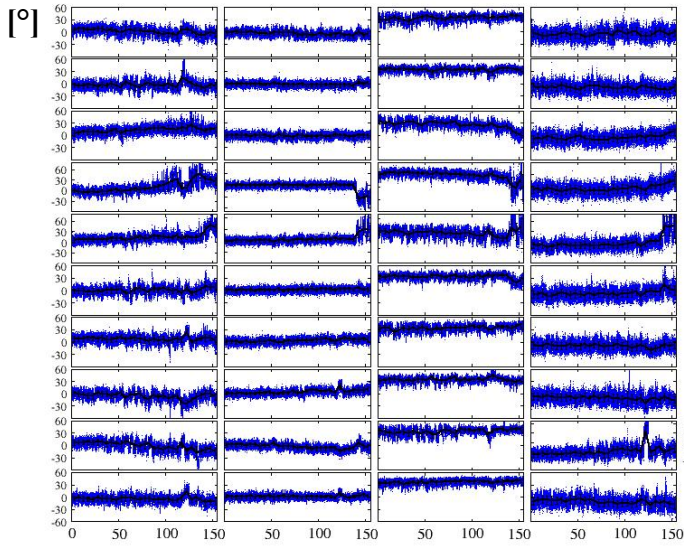
Base pair flip out dynamics during umbrella sampling simulation of DNA bending: Left column shows time-series of every base pair hydrogen bond donor-acceptor distances (three for C:G and two for A:T) and the right column time-series (blue), averages and mean square deviations (black) of helical base pair opening parameter, for central base pairs 3 to 12. Forward/Backward simulation corresponding to oligomers [GC], A1/A2, [AT], B1/B2, [Atract-1], C1/C2 and [Atract-2], D1/D2. Each row corresponds to one DNA junction. Hydrogen bond distances are represented as three or two superposed curves respectively in blue, green and black. The x-axis corresponds to bending coordinate used in umbrella sampling.

Figure C.

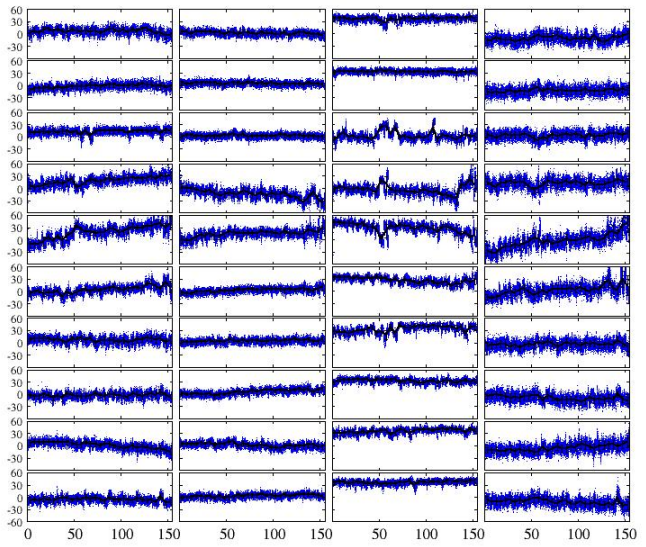
Base pair kink density histogram vs. global bending strain, averaged over consecutive groups of four adjacent umbrella windows (from left to right: 15°-30°, 35°-50°, ..., 135°-150°). Forward/Backward simulation corresponding to oligomers [GC], A1/A2, [AT], B1/B2, [Atract-1], C1/C2 and [Atract-2], D1/D2. First row corresponds to every YpR step and second row every RpY step concerning oligomers [GC] and [AT]. For [Atract-1] and [Atract-2] oligomers, first row corresponds to the A-tract block base pair steps (ApA and base pair junctions) while second row contains every flanking B-DNA sequence base pair steps. Each color represents different base pair junction for which position is indicated in the color-code legend at the bottom of the figure.

Figure D.

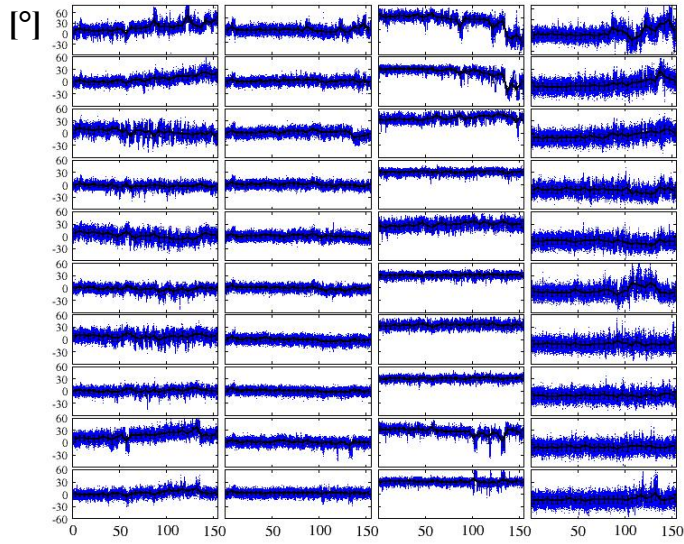
Phase space of backbone dihedral angles epsilon ϵ , zeta ζ and gamma γ (from left to right) vs. base pair bend angle calculated with Curves. A: First row of graphics contains phase point for every CG step and second row every GC step in [GC] oligomer. B: First row of graphics contains phase point for every TA step and second row every AT step in [AT] oligomer. C: First row of graphics contains phase point for every AA step and second row CA and AC junction steps in [Atract-1] oligomer. D: First row of graphics contains phase point for every AA step and second row the CA junction step (not AC since it is a terminal base-pair) in [Atract-2] oligomer.



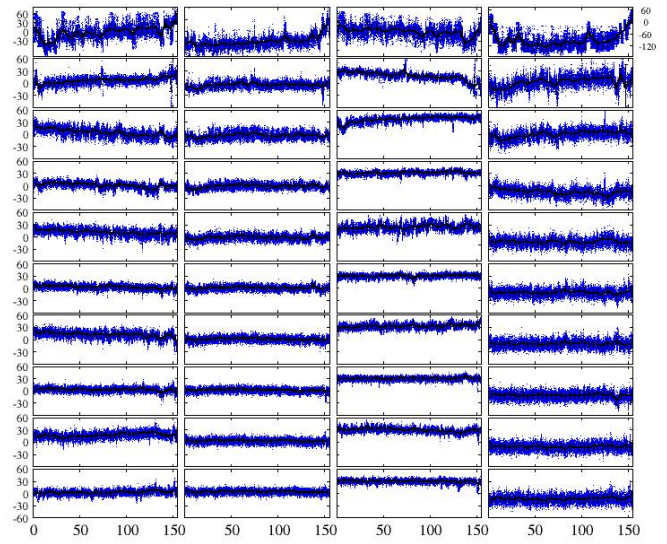
A1



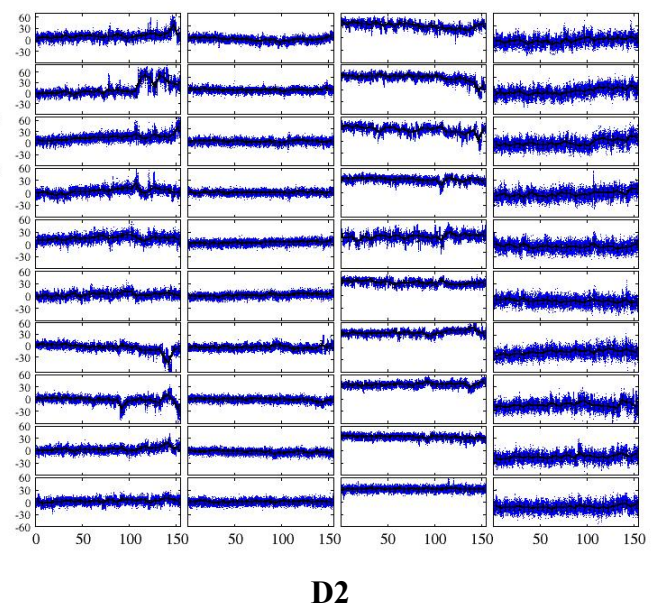
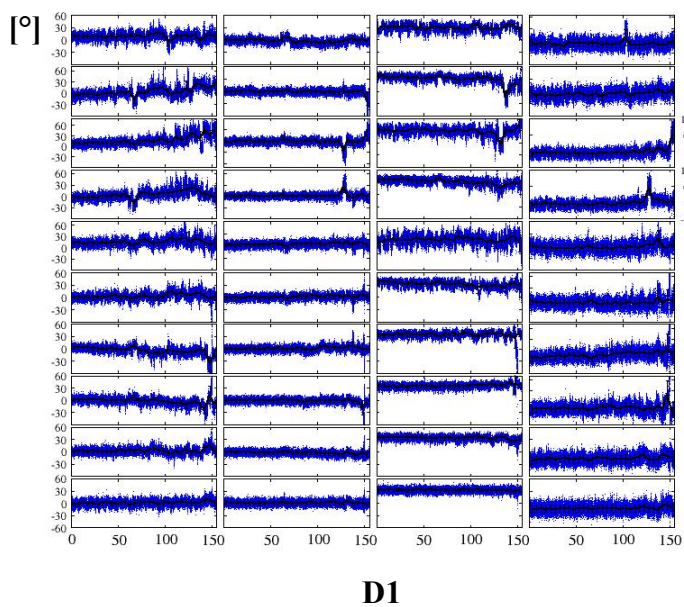
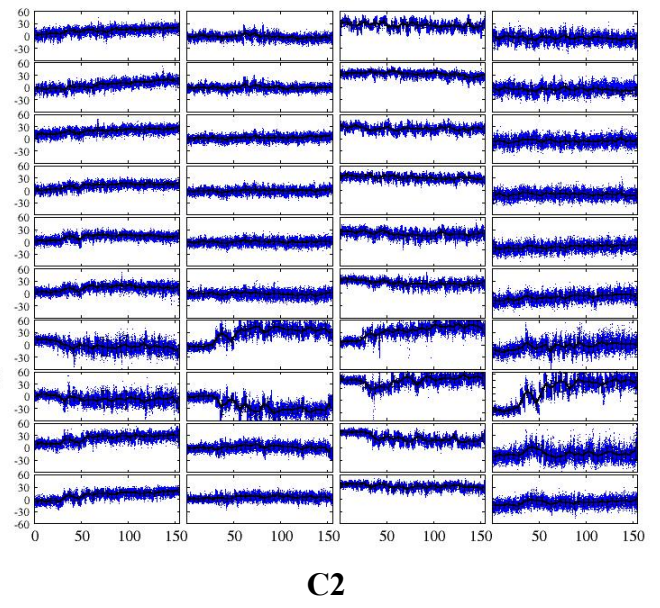
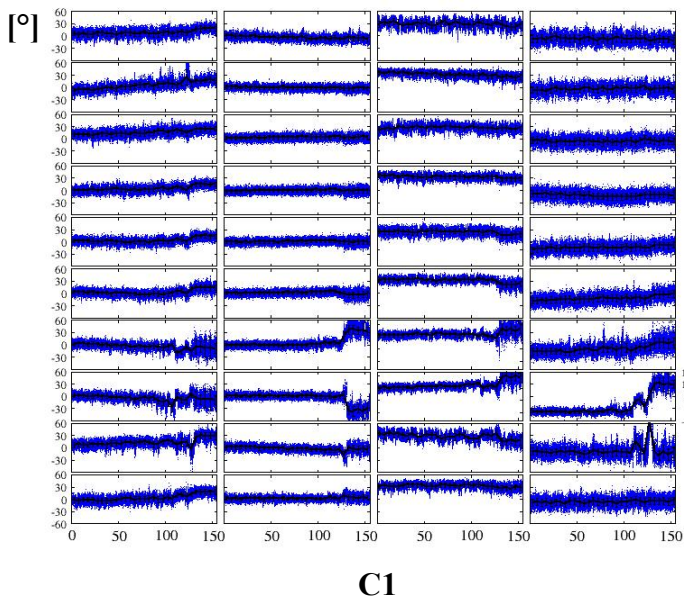
A2



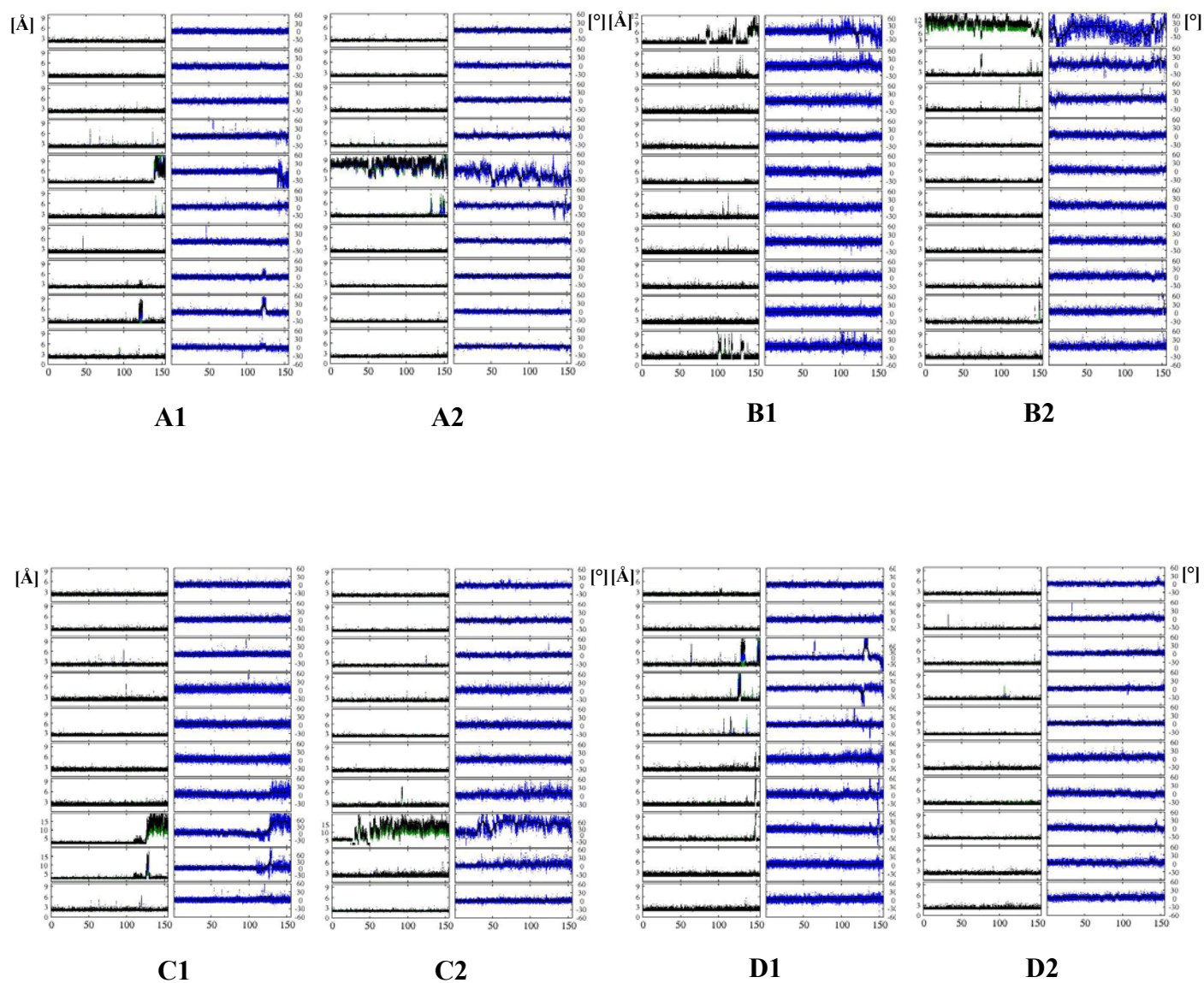
B1



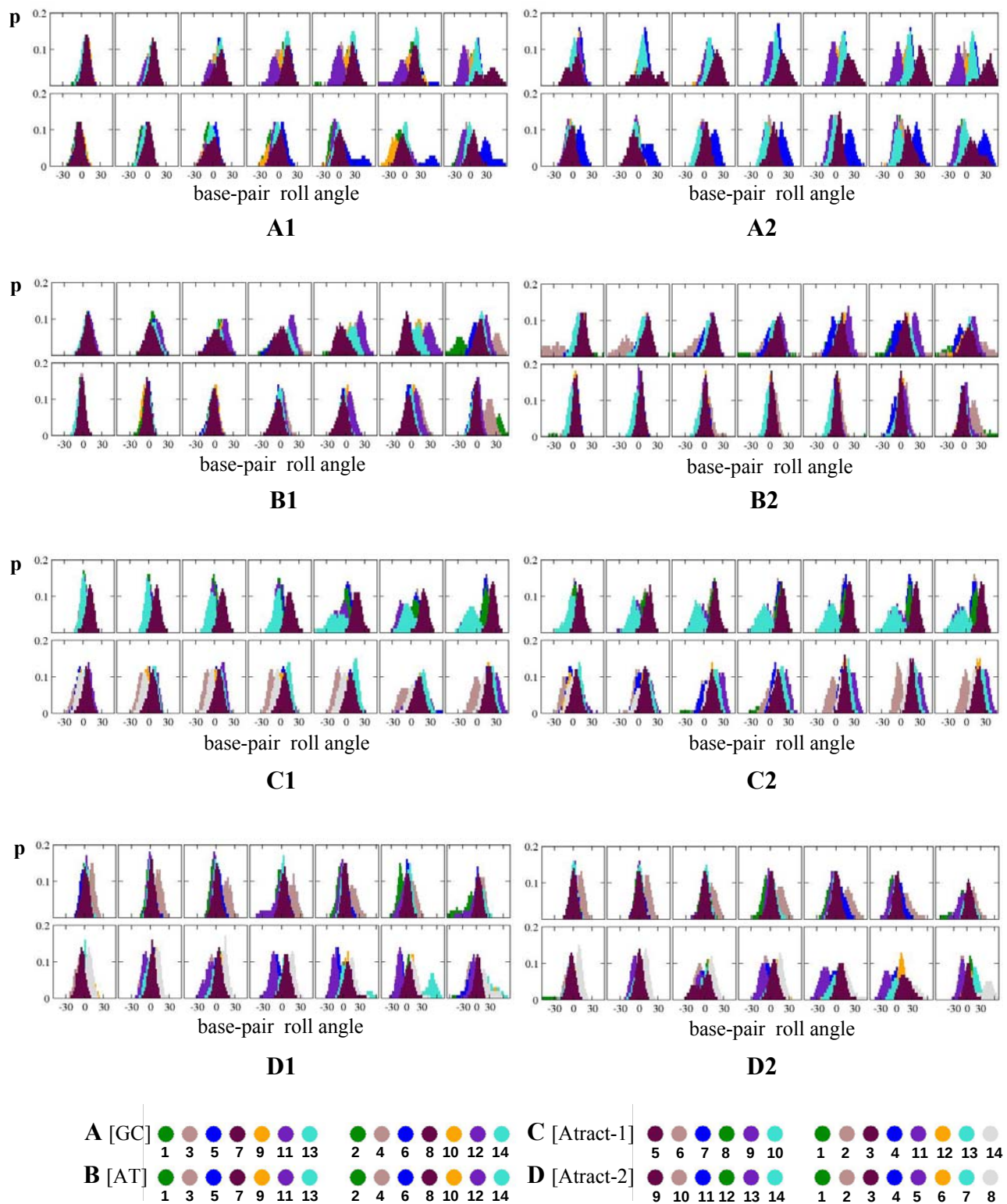
B2



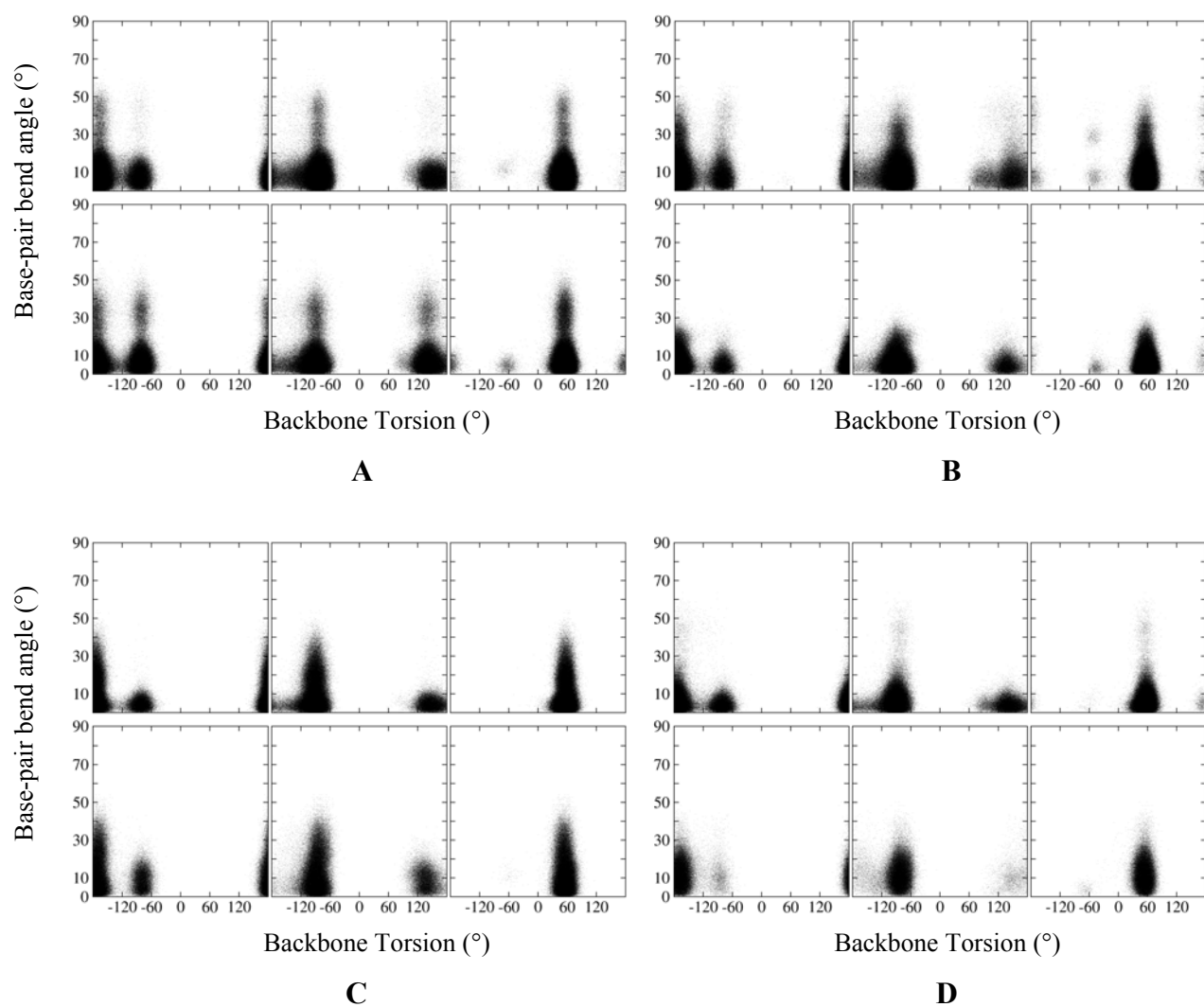
Supplementary Figure A



Supplementary Figure B



Supplementary Figure C



Supplementary Figure D

ANNEXE II

Convergence criteria and parameters of the Replica Exchange simulations (chapter V, VI)

Appendix II.1.- Mathematical details for D-BP-RexMD: the convergence criteria a_c^i

At every MD time step for current dihedral value ρ_c^i , a variance ratio a_c^i is evaluated for convergence. The ratio a_c^i divides the current deviation $\sigma_c^{i^2}$ by the limiting variance $\sigma_m^{i^2}$ of the corresponding metavariable m (=metavariable containing ρ^i), both taken from the average $\langle \rho^i \rangle_m$ over every value ρ^i sampled in m in the past:

$$a_c^i = \frac{\sigma_c^{i^2}}{\sigma_m^{i^2}} = \frac{\rho_c^i - \langle \rho^i \rangle_m}{\frac{1}{N_m} \sum \delta(m_c^i - m) (\rho_m^i - \langle \rho^i \rangle_m)} \quad \text{eq.II.1}$$

where δ is the usual Dirac delta symbol, i is one dihedral variable, m is a specific range of this dihedral (= current metavariable), N_m is the number of samples recorded in the metavariable m and c means current, updated value. The convergence to unity of a_c^i ($a_c^i < \text{or} = 1$) is tested on periods corresponding to a relaxation time chosen between two consecutive replica exchange attempts and if this convergence criterion is fulfilled the scaling potential is switched to the position of metavariable m .

An important setup is that the sampling point in the retailers-replicas does not need wandering away as long as the torsion metavariable sampled currently has never taken part in a replica exchange with the replica 0 (unscaled force field). Thus the convergence criterions a_c^i of metavariables are evaluated in the unscaled reference replica which itself is exempt of V_{NABD} term. Updated positions are communicated to every other retailers-replicas during the replica exchange attempt.

In summary the positions of metavariables which will be sampled are not known in advance for a given dihedral variable and a given molecular system. They are determined on the fly after convergence of their specific sub-averages in the early steps of the simulation. This automatic determination takes advantage of the enhanced rotational diffusion arising from systematic flooding of dihedral phase points.

In the limit of long-time sampling (*i.e.* every metavariable's average is converged) these metavariables are sampled alternatively and each determines an extended Hamiltonian (with specific position of the scaling potential) invariant with time. In practice the above time-dependent scaling approach violates the detailed balance criterion (equation 2) (292) but this equilibrating phase can be tracked, removed and used to determine a set of n positions where to introduce n flooding potentials in the 2D- ϵ/ζ phase space.

Appendix II.2.- Mathematical details for analysis: the convergence criteria $S(\tau)$

A rank of statistical efficiency for exploration of clusters of dihedrals was used during analysis of the trajectories. Based on the definition of the statistical inefficiency S (180) for comparing the accuracy of calculated averages with different sampling methods, we define a time limiting ratio of the observed variance of a torsion average (ϵ has been taken as indicator) to the total variance at $t = 4$ ns.

$$S(\tau) = \frac{\sigma_{\tau}^2}{\sigma_T^2} = \frac{\frac{1}{N\tau} \sum_{j=1}^{\tau} (\rho_j^i - \langle \rho^i \rangle)^2}{\frac{1}{N_T} \sum_{k=1}^T (\rho_k^i - \langle \rho^i \rangle)^2} \quad \text{eq.II.2}$$

This sub-class of autocorrelation function indicates the MD time range beyond which no new information contributes to the average torsion parameter. The assumption of uncorrelated Gaussians statistics does not hold because the NABD space divides into subaverages some of which are not accessible by all methods. However for analogous phase-space sampling (Fig.5.4 b/c, d and f) the convergence of $S(\tau)$ with time provides a reliable measure in term of relative efficiency.

Appendix II.3.- Simulation parameters and exchange ratios

Nucleotide System	Sampling Scheme	Number of Replicas	Simulation Length	Temperature / E _{max}
DNA d(AA)	1. Standard MD	1	20 ns	300K / Ø
	2. T-RexMD	26	26*20 ns	300,302,305,309,...,535K / Ø
	3. S-BP-RexMD	5	5*4ns	300K / 0,1,3,6,10 kcal/mol
	4. D-BP-RexMD	5	5*4 ns	300K / 0,1,3,6,10 kcal/mol
	5. E-BP-RexMD	5	5*4 ns	300K / 0,1,3,6,10 kcal/mol
DNA abasic system PDB-2hss	1. Standard MD	1	5 ns	300K / Ø
	2. T-RexMD	16	16*5 ns	300,302,305,309,...,402K / Ø
	3. E-BP-RexMD	5	5*5 ns	300K / 0,1,3,6,10 kcal/mol

Table II.1. Simulation parameters for the sampling method compared and the molecular system used in Chapter 5.

Method, # Replicas	Acceptance Ratio per Replica (in %)
T-RexMD AA 26	37.4 - 69.1 - 56.6 - 43.4 - 33.3 - 25.6 - 19.2 - 15.3 - 12.7 - 12.2 - 13.9 - 15.3 - 17.8 - 18.2 - 16.9 - 19.6 - 21.9 - 23.4 - 23.0 - 22.7 - 25.1 - 27.3 - 30.0 - 30.0 - 29.8 - 15.4
S-BP-RexMD AA 5	56.9 - 67.6 - 81.2 - 86.6 - 89.0
D-BP-RexMD AA 5	74.9 - 71.8 - 59.7 - 53.5 - 56.2
E-BP-RexMD AA 5	41.8 - 39.4 - 49.8 - 64.7 - 66.8
T-RexMD ABA 16	28.2 - 22.8 - 18.2 - 20.2 - 22.3 - 25.9 - 22.7 - 23.7 - 19.5 - 25.2 - 9.9 - 9.2 - 17.1 - 12.5 - 13.5 - 10.3
E-BP-RexMDABA 5	34.4 - 33.3 - 26.3 - 24.5 - 29.2

Table II.2. Final acceptance ratios *per* replica (exchange frequencies) of every Replica Exchange Molecular Dynamics simulations of Chapter 5.

Method, # Replicas	Acceptance Ratio per Replica (in %)
D-BP-RexMD 5	74.9 - 71.8 - 59.7 - 53.5 - 56.2
SCREM 5 ($\omega_G = 10^{-5}$)	71.4 - 64.3 - 60.8 - 68.6 - 72.6
SCREM 5 ($\omega_G = 10^{-3}$)	7.5 - 47.5 - 89.9 - 93.1 - 93.8
SCREM 2 ($\omega_G = 10^{-4}$)	49.2 for both
SCREM 2 ($\omega_G = 5 \cdot 10^{-4}$)	14.4 for both
SCREM 2 ($\omega_G = 10^{-3}$)	8.2 for both

Table II.3. Final acceptance ratios *per* replica (exchange frequencies) of every Replica Exchange Molecular Dynamics simulations of Chapter 6.

ANNEXE III

Fortran code for inducing bending between two screw axis in Nucleic Acids (chapter IV)

The subroutine can be introduced in the sander module of the Amber program

```
#include "dprec.h"

!+++++
      subroutine benrg(x,f,inrest,thet,ir,ntwr,e,r,rk,tlim,ntb, &
                     aave,aave0,nave,nexact,ipower,tauave,ravein,dt, &
                     nstep,navint,iave,incflg,iravin,iflag,name,ipres)

!
!
! Subroutine BEnding eNeRGy
!
! This subroutine calculates a nucleic acid bending angle restraint energy
! between two handles/axoids of 4 base-pairs each (Curuksu,J.,
! Zakrzewska,K., Zacharias,M. (2008) Nucleic Acid Research). The
! energy function is quadratic.
!
! if r is the bending angle:
!
!      r < R   : E = Rk*(r-R)**2
!      r = R   : E = 0
!      R < r   : E = Rk*(r-R)**2
!
! specific input/variables for this subroutine:
!   ir(i) : array of residues for defining two handles (2*8 nucleotides).
!   ia(i) : atom pointers array.
!   (v,w,z) : DNA/RNA base fixed coordinate system.
!   u(i) : space-fixed rotation axis between (v,w,z)i and (v,w,z)i+1.
!   ax,ay : the two bending handles (vector-sum of rotation vectors).
!   Each row contains dx,dy,dz of one atom with respect to the current
!   derivative coordinate, for which dxyz is set to 1 (other are set to 0).
!
! other input:
!   x(i) : coordinate array.
!   f(i) : force array;
!           modified on output to include forces from this restraint.
!   inrest: number of the restraint(the 1st or the 2nd) for 2d constraint
!   i1-i16 : atom pointers for the c1' (3*(i-1), where i is absolute atom
!           number. The angle is defined between two vectors that sum
!           local inter-nucleotide rotation axis u. 1st vector: i1,i2,i3,i4
!           and nucleotides on the complementary strand (i5,i6,i7,i8).
!           2nd vector: nucleotides i9,i10,i11,i12
!           and nucleotides on the complementary strand (i13,i14,i15,i16).
```

```

!      pout   : print result after every ntwr steps
!      ntb    : periodic boundary conditions flag.
!      iflag  : =0, then calculate energy and derivatives
!               =1, then only calculate current value of angle
!               =2, the calculate current value of angle and energy, but
!                   do not accumulate derivatives
!      incflg : determines whether local saved pointers are updated in aveint
!

```

```

      implicit _REAL_ (a-h,o-z)
      _REAL_ rk,nu,nv,nw,nz,nax,nay
      parameter (small=1.0d-14)
      parameter (khandle = 2)
      parameter (kh4 = 4 * khandle)
      parameter (kscrew = kh4 - 4)
      integer pout
      integer iatnm
      parameter (pi = 3.141592653589793d0)
      parameter (rtod = 180.0d0 / pi)
      parameter (dtor = pi / 180.0d0)
      parameter (small2=1.0d-5)
      dimension x(1),f(1),aave(1),aave0(1),dg(7),dc(7),ind(7)
      dimension xij(3),xkj(3),s1(3),s2(3),xijs(3),t(3)
      dimension dgc(7,3),dcc(7,3),dtt(7,3),dgt(7,3),dte(7,3)
      dimension dk(7,3),dgu(7,3),dpp(7,3,3),fo(7,3),name(1),ipres(1)
      dimension
ia(7*kh4),v(3,kh4),w(3,kh4),z(3,kh4),g(3,kh4),nv(kh4),nw(kh4)
      dimension nz(kh4),ax(3),ay(3),q(3,3)
      dimension dxyz(7,3),u(3,kscrew),nu(kscrew),dnu(kscrew),d(7*kh4,3)
      dimension cosrf(kscrew),ustep(3,kscrew),ir(kh4),iatnm(4),m(7)
      character(len=4) atnam(4),string
      logical :: dbug = .true.
      rinc = 1000.0d0
      kh = khandle
      kh2 = 2 * kh
      kh3 = 3 * kh

      iflg = 1
      if (iflag == 1) iflg = 2
      rf = r * rtod

      lstep = nstep + 1
      write(6,'(1X,"step",11X,I6)') nstep

```

```

! =====
!      1) Definition of base fixed coordinate axis systems
! =====

```

```

! Find an atom specific to A, T, G or C

```

```

write(6,'(1X,"benrg",8i4)')ir(1),ir(2),ir(3),ir(4),ir(5),ir(6),ir(7),ir(8)

```

```

      atnam(1) = "N4"
      atnam(2) = "O4"
      atnam(3) = "O6"
      atnam(4) = "N6"

```

```

      do i = 1,4
        read(atnam(i),'(A4)') iatnm(i)
      enddo

```

```

! Find seven atoms for each nucleotide (there are 16)

      l = 0

residus: do j = 1,kh4
checkatom:  do i = ipres(ir(j)),ipres(ir(j)+1)-1

      if(name(i) == iatnm(1))then          ! Cytosine
        m(1) = 3*(i-8-1)    ! C1'
        m(5) = m(1) + 21    ! C4
        m(6) = m(5) + 12    ! N3
        m(7) = m(6) + 3     ! C2
        goto 100

      else if(name(i) == iatnm(2))then      ! Tyrosine or Uridine
        m(1) = 3*(i-11-1)   ! idem Cyt.
        m(5) = m(1) + 30
        m(6) = m(5) + 6
        m(7) = m(6) + 6
        goto 100

      else if(name(i) == iatnm(3))then      ! Guanine
        m(1) = 3*(i-8-1)    ! C1'
        m(5) = m(1) + 21    ! C6
        m(6) = m(5) + 6     ! N1
        m(7) = m(6) + 6     ! C2
        goto 100

      else if(name(i) == iatnm(4))then      ! Adenine
        m(1) = 3*(i-8-1)    ! idem Gua.
        m(5) = m(1) + 21
        m(6) = m(5) + 12
        m(7) = m(6) + 3
        goto 100
      endif

    enddo checkatom

100    m(2) = m(1) + 6        ! N1/N9                ! for A,T,G and C
      m(3) = m(2) + 3        ! C6/C8
      m(4) = m(3) + 6        ! C5/N7

      do i = 1,7
        l = l + 1
        ia(l) = m(i)
      enddo

    enddo residus

      write(6,'(1X,"benrg 2")')

! calculate 16 nucleotide based axis coordinate systems

refaxis: do i = 1,kh4
      j = (i - 1) * 7 + 1
      do l = 1,3
        v(l,i) = 0
        do n = 1,5
          v(l,i) = v(l,i) + x(ia(j + n + 1) + 1) - x(ia(j + n) + 1)
        enddo
      enddo
    enddo

```

```

        enddo
        v(1,i) = v(1,i) / 5
        g(1,i) = x(ia(j) + 1) - x(ia(j + 1) + 1)
    enddo

    z(1,i) = v(2,i) * g(3,i) - v(3,i) * g(2,i)
    z(2,i) = v(3,i) * g(1,i) - v(1,i) * g(3,i)
    z(3,i) = v(1,i) * g(2,i) - v(2,i) * g(1,i)
    w(1,i) = v(2,i) * z(3,i) - v(3,i) * z(2,i)
    w(2,i) = v(3,i) * z(1,i) - v(1,i) * z(3,i)
    w(3,i) = v(1,i) * z(2,i) - v(2,i) * z(1,i)

    nv(i) = sqrt(v(1,i)*v(1,i) + v(2,i)*v(2,i) + v(3,i)*v(3,i))
    nz(i) = sqrt(z(1,i)*z(1,i) + z(2,i)*z(2,i) + z(3,i)*z(3,i))
    nw(i) = sqrt(w(1,i)*w(1,i) + w(2,i)*w(2,i) + w(3,i)*w(3,i))

    do l = 1,3
        v(1,i) = v(1,i) / nv(i)
        z(1,i) = z(1,i) / nz(i)
        w(1,i) = w(1,i) / nw(i)
    enddo

! orth.basis must have identical orientation on both strand, thus:

    if ((i > kh .and. i <= kh2) .or. (i > kh3)) then
        do l = 1,3
            z(1,i) = - z(1,i)
            w(1,i) = - w(1,i)
        enddo
    endif

    det = v(1,i) * (z(2,i)*w(3,i) - z(3,i)*w(2,i)) - &
          z(1,i) * (v(2,i)*w(3,i) - v(3,i)*w(2,i)) + &
          w(1,i) * (v(2,i)*z(3,i) - z(2,i)*v(3,i))
    if(det < 0.995) write(6,('Det. not +1 at res. ",i2)') i

enddo  refaxis

! =====
!      2) Definition of two handles for global nucleic acid bending
!      =====

    do i = 1,3
        ax(i) = 0.0
        ay(i) = 0.0
    enddo

screw:  do i = 1,kscrew

        l = i + (i - 1) / (kh - 1)      ! ++1 every i ++(kh-1)

! rotation matrix q

        do k = 1,3
            do j = 1,3
                q(j,k) = v(j,l)*v(k,l+1) + z(j,l)*z(k,l+1) +
w(j,l)*w(k,l+1)
            enddo
        enddo

```

```

! rotation vectors u

u(1,i) = q(3,2) - q(2,3)
u(2,i) = q(1,3) - q(3,1)
u(3,i) = q(2,1) - q(1,2)
nu(i) = sqrt(u(1,i)*u(1,i) + u(2,i)*u(2,i) + u(3,i)*u(3,i))
do j = 1,3
    u(j,i) = u(j,i) / nu(i)
enddo

! keep orientation of u in same hemisphere as in previous step

if (nstep > 1)then
    cosrf(i) = (ustep(1,i) * u(1,i) + ustep(2,i) * u(2,i) &
                + ustep(3,i) * u(3,i))
    if (dacos(cosrf(i)) > pi/2) then
        do j = 1,3
            u(j,i) = - u(j,i)
        enddo
        write(6,('vector u opposed'))
    endif
endif

do j = 1,3
    ustep(j,i) = u(j,i)
enddo

! handles (ax==----...---==ay), sum of rotation vectors

do j = 1,3
    if (i <= (kh2 - 2)) ax(j) = ax(j) + u(j,i)
    if (i > (kh2 - 2)) ay(j) = ay(j) + u(j,i)
enddo

enddo screw

nax = sqrt(ax(1)*ax(1) + ax(2)*ax(2) + ax(3)*ax(3))
nay = sqrt(ay(1)*ay(1) + ay(2)*ay(2) + ay(3)*ay(3))

do j = 1,3
    ax(j) = ax(j) / nax
    ay(j) = ay(j) / nay
enddo

! calculate bending angle

cosi = ax(1)*ay(1) + ax(2)*ay(2) + ax(3)*ay(3)
sinu = sqrt(1 - (cosi*cosi))

if (sinu == 0.) sinu = 0.011
if (sinu == 0.) write(6,('sine of global bending angle was 0'))
if (cosi > 1.0d0) cosi = 1.0d0

if (cosi < -1.0d0) cosi = -1.0d0

thet = dacos(cosu)
val = thet * rtod

! debug tool

if (dbug) then          !debug is set in the header

```

```

        write(6, '("ax coord",3f8.3)') ax(1), ax(2), ax(3)
        write(6, '("ay coord",3f8.3)') ay(1), ay(2), ay(3)
        do j = 1,kh4
            write(6, '(" v axis",3f8.3)') v(1,j), v(2,j), v(3,j)
            write(6, '(" g axis",3f8.3)') g(1,j), g(2,j), g(3,j)
            write(6, '(" z axis",3f8.3)') z(1,j), z(2,j), z(3,j)
            write(6, '(" w axis",3f8.3)') w(1,j), w(2,j), w(3,j)
            if(j <= kscrew) write(6, '(" u screw",3f8.3)') u(1,j), u(2,j),
u(3,j)
        enddo
    endif

    if (iflag == 1) return

! Convert rk from deg to rad (note rk unit is kcal.mol-1.deg-2)

    rkr = rk * rtod * rtod                ! rtod = dtor-1

! calculate energy (e)

    if (thet == r) then
        e = 0.0d0
        df = 0.0d0
    else
        dif = thet - r
        e = rkr * dif * dif
        df = -2.0d0 * rkr * dif / sinu
    endif

    write(6, '("bending angle:",f7.2,/, "EBENDING =",4X,f7.2)')val,e

    if (iflag == 2) return

! =====
!      3) Analytic derivatives (xyz, 112 atoms) of the two handles
! =====

    m = 0

residu:  do n = 1,kh4
        coord:  do j = 1,3

            atom:  do i = 1,7

! set every coordinate derivative to 0 (see header comment)

                do jj = 1,3
                    do ii = 1,7
                        dxyz(ii,jj) = 0.0
                    enddo
                enddo

! set the current coordinate derivative to 1

                dxyz(i,j) = 1.

! restore initial (unnormalized) axis system

                do l = 1,3

```

```

v(1,n) = v(1,n) * nv(n)
z(1,n) = z(1,n) * nz(n)
w(1,n) = w(1,n) * nw(n)
enddo

if ((n > kh .and. n <= kh2).or.(n > kh3))then
do l = 1,3
z(1,n) = - z(1,n)
w(1,n) = - w(1,n)
enddo
endif

```

! derivative of coordinate axis system

```

dvx = 0.
dvy = 0.
dvz = 0.
do k = 2,6
dvx = dvx + dxyz(k + 1,1) - dxyz(k,1)
dvy = dvy + dxyz(k + 1,2) - dxyz(k,2)
dvz = dvz + dxyz(k + 1,3) - dxyz(k,3)
enddo
dvx = dvx / 5
dvy = dvy / 5
dvz = dvz / 5

dgx = dxyz(1,1) - dxyz(2,1)
dgy = dxyz(1,2) - dxyz(2,2)
dgz = dxyz(1,3) - dxyz(2,3)

dzx = (dvy * g(3,n) + v(2,n) * dgz) - &
(dvz * g(2,n) + v(3,n) * dgy)
dzy = (dvz * g(1,n) + v(3,n) * dgx) - &
(dvx * g(3,n) + v(1,n) * dgz)
dzz = (dvx * g(2,n) + v(1,n) * dgy) - &
(dvy * g(1,n) + v(2,n) * dgx)

dwx = (dvy * z(3,n) + v(2,n) * dzz) - &
(dvz * z(2,n) + v(3,n) * dzy)
dwy = (dvz * z(1,n) + v(3,n) * dzx) - &
(dvx * z(3,n) + v(1,n) * dzz)
dwz = (dvx * z(2,n) + v(1,n) * dzy) - &
(dvy * z(1,n) + v(2,n) * dzx)

dnv = (v(1,n) * dvx + v(2,n) * dvy + v(3,n) * dvz) / nv(n)
s = nv(n) * nv(n)
dvx = (dvx * nv(n) - v(1,n) * dnv) / s
dvy = (dvy * nv(n) - v(2,n) * dnv) / s
dvz = (dvz * nv(n) - v(3,n) * dnv) / s

dnz = (z(1,n) * dzx + z(2,n) * dzy + z(3,n) * dzz) / nz(n)
s = nz(n) * nz(n)
d zx = (dzx * nz(n) - z(1,n) * dnv) / s
d zy = (dzy * nz(n) - z(2,n) * dnv) / s
d zz = (dzz * nz(n) - z(3,n) * dnv) / s

dnw = (w(1,n) * dwx + w(2,n) * dwy + w(3,n) * dwz) / nw(n)
s = nw(n) * nw(n)
dwx = (dwx * nw(n) - w(1,n) * dnw) / s
dwy = (dwy * nw(n) - w(2,n) * dnw) / s

```

```

dwz = (dwz * nw(n) - w(3,n) * dnw) / s

if (n > kh .and. n <= kh2 .or. n > kh3)then
  dzx = -dzx
  dzy = -dzy
  dzz = -dzz
  dwx = -dwx
  dwy = -dwy
  dwz = -dwz
endif

```

! restore normalized axis system

```

do l = 1,3
  v(l,n) = v(l,n) / nv(n)
  z(l,n) = z(l,n) / nz(n)
  w(l,n) = w(l,n) / nw(n)
enddo

if (n > kh .and. n <= kh2 .or. n > kh3)then
  do l = 1,3
    z(l,n) = -z(l,n)
    w(l,n) = -w(l,n)
  enddo
endif

```

! derivative of rotation vectors 'u'

```

x1 = dvz * v(2,n+1) + dzz * z(2,n+1) + &
     dwz * w(2,n+1) - ( dvy * v(3,n+1) + &
     dzy * z(3,n+1) + dwy * w(3,n+1) )
y1 = dvx * v(3,n+1) + dzx * z(3,n+1) + &
     dwx * w(3,n+1) - ( dvz * v(1,n+1) + &
     dzz * z(1,n+1) + dwz * w(1,n+1) )
z1 = dvy * v(1,n+1) + dzy * z(1,n+1) + &
     dwy * w(1,n+1) - ( dvx * v(2,n+1) + &
     dzx * z(2,n+1) + dwx * w(2,n+1) )

x2 = v(3,n-1) * dvy + z(3,n-1) * dzy + &
     w(3,n-1) * dwy - ( v(2,n-1) * dvz + &
     z(2,n-1) * dzz + w(2,n-1) * dwz )
y2 = v(1,n-1) * dvz + z(1,n-1) * dzz + &
     w(1,n-1) * dwz - ( v(3,n-1) * dvx + &
     z(3,n-1) * dzx + w(3,n-1) * dwx )
z2 = v(2,n-1) * dvx + z(2,n-1) * dzx + &
     w(2,n-1) * dwx - ( v(1,n-1) * dvy + &
     z(1,n-1) * dzy + w(1,n-1) * dwy )

k = n - (n - 1) / kh      ! --1 every n ++kh

if (n < kh4) then
  do l = 1,3
    u(l,k) = u(l,k) * nu(k)
  enddo
  dnu(k) = (u(1,k) * x1 + u(2,k) * y1 + u(3,k) * z1)
            / nu(k)
  s = nu(k) * nu(k)
  x1 = (x1 * nu(k) - u(1,k) * dnu(k)) / s
  y1 = (y1 * nu(k) - u(2,k) * dnu(k)) / s
  z1 = (z1 * nu(k) - u(3,k) * dnu(k)) / s

```

```

        do l = 1,3
            u(l,k) = u(l,k) / nu(k)
        enddo
    endif

    if(n > 1)then
        do l = 1,3
            u(l,k-1) = u(l,k-1) * nu(k-1)
        enddo
        dnu(k-1) = (u(1,k-1) * x2 + u(2,k-1) * y2 +
                    u(3,k-1) * z2) / nu(k-1)
        s = nu(k-1) * nu(k-1)
        x2 = (x2 * nu(k-1) - u(1,k-1) * dnu(k-1)) / s
        y2 = (y2 * nu(k-1) - u(2,k-1) * dnu(k-1)) / s
        z2 = (z2 * nu(k-1) - u(3,k-1) * dnu(k-1)) / s
        do l=1,3
            u(l,k-1) = u(l,k-1) / nu(k-1)
        enddo
    endif

    if(dacos(cosrf(k)) > pi/2)then
        x1 = -x1
        y1 = -y1
        z1 = -z1
    endif

    if(dacos(cosrf(k-1)) > pi/2)then
        x2 = -x2
        y2 = -y2
        z2 = -z2
    endif

```

! derivative of the handles (ax,ay)

```

    if(mod((n-1),kh4) == 0.)then
        axoidx = x1
        axoidy = y1
        axoidz = z1
    elseif(mod(n,kh4) == 0.)then
        axoidx = x2
        axoidy = y2
        axoidz = z2
    else
        axoidx = x2 + x1
        axoidy = y2 + y1
        axoidz = z2 + z1
    endif

    if (n <= kh2)then
        dnax = ax(1) * axoidx + ax(2) * axoidy + ax(3) * axoidz
        axoidx = (axoidx - ax(1) * dnax) / nax
        axoidy = (axoidy - ax(2) * dnax) / nax
        axoidz = (axoidz - ax(3) * dnax) / nax
        b = axoidx * ay(1) + axoidy * ay(2) + axoidz * ay(3)
        a = ax(1) * axoidx + ax(2) * axoidy + ax(3) * axoidz
    else
        dnay = ay(1) * axoidx + ay(2) * axoidy + ay(3) * axoidz
        axoidx = (axoidx - ay(1) * dnay) / nay
    endif

```

```

        axoidy = (axoidy - ay(2) * dnay) / nay
        axoidz = (axoidz - ay(3) * dnay) / nay
        b = axoidx * ax(1) + axoidy * ax(2) + axoidz * ax(3)
        a = ay(1) * axoidx + ay(2) * axoidy + ay(3) * axoidz
    endif

! derivative of the penalty term k(thet - r)2

        d(i + m, j) = df * (b - cosi * a)

        enddo atom
    enddo coord
    m = m + 7

enddo residu

! array 'd' contains 112 rows (16 res. * 7 ato.) and 3 columns (xyz coord.)

    do i = 1, kh4 * 7
        na = ia(i)
        do j = 1, 3
            f(na + j) = f(na + j) - d(i, j)
        enddo
    enddo

! ===== the end
    return
end subroutine benrg

```

ANNEXE IV

Fortran code for server client updates in Hamiltonian Replica Exchange applied to Nucleic Acid backbone dihedrals (SCREM, chapter VI)

Script to introduce in the tornrg.f subroutine of sander module in the Amber program

```
elseif (ialtd == 2) then ! option to use the BREM/SCREM protocols

! read memory function parameter sets
! at t=0, client1.dat and client2.dat (array of 0) need be present in
! current path

    if(nstep == 0)then

        ! client 1 starting requests

        rg1(ir) = r1
        rg2(ir) = r2
        rg3(ir) = r3
        rg4(ir) = r4
        idiff(ir) = 0

        if(k2 /= 0.) write(6,('("destabilizing range= ",4f5.1)')r1,r2,r3,r4

        line = 0
        call amopen(55,"/username/path/client1.dat",'O','F','R')
43      line = line + 1
        read(55,
' (12f8.4,6i10)',end=44)(b1(j,line),j=1,6),(b2(j,line),j=1,6),(ns(j,line),j=
1,6)
        goto 43
44      close(55)

        ira = ir !total number of requests(client1)

        ! client 2 starting requests

        line = 0
        call amopen(56,"/username/path/client2.dat",'O','F','R')
45      line = line + 1
        read(56,'(16i10)',end=46) (n16(j,line),j=0,15)
        if(k2 /= 0.) write(6,('("gaussian memory numbers=
",16i10)')(n16(j,line),j=0,15)
        goto 45
46      close(56)
```

```

    cg(0) = 11.25 !gaussian centers
    do i=1,15
        cg(i) = cg(i-1) + 22.5
    enddo
    do i=0,15
        cg(i) = cg(i) * dtor
    enddo

endif !step 0 vs. step >= 1

if(ir > ira) ira = ir    !just a guarantee...

! Translate the value of the torsion (by +- N*360) to bring it as close
as possible
! to the interval [0-360]. Use this ap value from now on.

28 if (ap > pi2) then
    ap = ap - pi2
    goto 28
else if (ap < 0.) then
    ap = ap + pi2
    goto 28
endif

!=====
!   Servers/Client Interactions
!=====

ap = ap / dtor

if (k2 == 0.0)then !CLIENT TRANSACTIONS

    ! Update Client 2 requests (Gaussian parameters set)

    r = ap / 22.5
    ig = int(r)
    n16(ig,ir) = n16(ig,ir) + 1

    ! Update Client 1 requests (Global Potential parameters set)

    if ((ap >= 30.) .and. (ap < 330.))then
        d = (ap - 30.) / 60.
        m = int(d) + 1
        rge1 = (m - 1) * 60.
    elseif((ap >= 330.) .and. (ap < 360.) .or. (ap >= 0) .and. (ap <
30))then
        m = 6
        rge1 = -60.
    else
        write(6,*) "ap not assigned in global ranges loop"
    endif

    write(6,*) "dt =",nstep
    write(6,*) "torsion =",ap

    ns(m,ir) = ns(m,ir) + 1
    rge1 = rge1 * dtor
    ap = ap * dtor

    b1(m,ir) = ( b1(m,ir) * (ns(m,ir) - 1) + ap ) / ns(m,ir)

```

```

b2now = ( ap - b1(m,ir) ) * ( ap - b1(m,ir) )
b2(m,ir) = ( b2(m,ir) * (ns(m,ir) - 1) + b2now ) / ns(m,ir)
diff = b2now - b2(m,ir)

if(diff > 0.01) then  !!!OR. (diff < -0.01))then
    idiff(ir) = 0
else
    idiff(ir) = idiff(ir) + 1
endif

! below is some code for an alternative criterion (mainly
! var(t)==var(0->t) instead of <)
!!      b2old = b2(m,ir)      ! new criterion
!!      b1(m,ir) = ( b1(m,ir) * (ns(m,ir) - 1) + ap ) / ns(m,ir)
!!      b2now = ( ap - b1(m,ir) ) * ( ap - b1(m,ir) )
!!      b2(m,ir) = ( b2(m,ir) * (ns(m,ir) - 1) + b2now ) /
ns(m,ir)
!!      diff = b2now - b2(m,ir) ! this was the old criterion
!!      diff = b2old - b2(m,ir) ! new criterion
!!
!!      if((diff > 0.01) .OR. (diff < -0.01))then
!!          idiff(ir) = 0
!!      else
!!          idiff(ir) = idiff(ir) + 1
!!      endif

else !SERVER RESPONSES

    ! Use Client 2 requests (Gaussian parameters set)

    r = ap / 22.5
    ig = int(r) !current gaussian range
    ap = ap * dtor

    w = 22.5 * dtor
    w2 = w * w
    gvar = (ap - cg(ig)) * (ap - cg(ig))
    ng = n16(ig,ir) * 0.0001      ! may be fine tuned
    g = - gvar / (2 * w2)
    g = exp(g)
    eg = ng * g
    df_g = - eg * (ap - cg(ig)) / w2

endif ! CLIENT vs. SERVERS

! Use Client 1 requests (Global Potential parameters set)

if(idiff(ir) >= 100)then

    rg1(ir) = rge1
    rg2(ir) = rge1 + (30 * dtor)
    rg3(ir) = rg2(ir) + (60 * dtor)
    rg4(ir) = rg3(ir) + (30 * dtor)
    write(6,*) "=== global substates potential move ==="

endif

r1 = rg1(ir)
r2 = rg2(ir)
r3 = rg3(ir)
r4 = rg4(ir)

```

```

! Save updated paramater sets just before the REx attempt

if((nstep == 499) .and. (k2 == 0.0))then

    write(6, 'BPranges ', (8f7.1)) (rg1(k)/dtor, k=1, ira) !need adapt
format and rupdateo.pl for ira torsions (here 8)

    temp = rg1(ir)/dtor
    call amopen(56, "/username/path/client2.dat", 'R', 'F', 'W')

    do i = 1, ira
        write(56, '(16i10)') (n16(j,i), j=0, 15)
    enddo
    close(56)

    call amopen(55, "/username/path/client1.dat", 'R', 'F', 'W')
    do i=1, ira
        write(55, '(12f8.4, 6i10)')
(b1(j,i), j=1, 6), (b2(j,i), j=1, 6), (ns(j,i), j=1, 6)
    enddo
    close(55)

endif

!=====
!   energy and forces calculations
!=====

! Translate the value of the torsion (by +- N*360) to bring it as close
as possible
! to one of the two central "cutoff" points (r2,r3). Use this ap value
from now on.

apmean = (r2+r3)*0.5d0
18 if (ap-apmean > pi) then
    ap = ap - pi2
    goto 18
else if (apmean-ap > pi) then
    ap = ap + pi2
    goto 18
endif

app(ir) = ap !store ap for use at the next ir torsion

Emax = k2
k2 = Emax / ((r1-r2)**4)
k3 = Emax / ((r3-r4)**4)

! Based on the original scenario: ir=1-4 correspond to Epsilon-Zeta-
Zeta-Epsilon

if (mod(ir,2) > 0.) return ! calculate penalty only for even
values of ir

apmp = (rg2(ir-1)+rg3(ir-1))*0.5d0 ! apmean at ir-1
crad = (r3-r2)*0.5d0 ! radius of bump center
brad = crad + (r2-r1) ! radius of bump

if ((ap-apmean)**2 > brad**2 .or. (app(ir-1)-apmp)**2 > brad**2)then
    e = 0.0d0

```

```

        if (k2 /= 0.0) e = eg
        df = 0.0d0
        write(6,*) "e=",e,"-Outside"
        !return
    else if (ap >= r2 .and. ap <= r3 .and. app(ir-1) >= rg2(ir-1) .and.
app(ir-1) <= rg3(ir-1)) then
        e = k2*(r1-r2)**4
        if (k2 /= 0.0) e = e + eg
        df = 0.0d0
        write(6,*) "e=",e,"-Emax"
        !return
    else
        difc = ap-apmean
        if(difc < 0.0d0) difc = - difc
        difc = (difc - crad)

        difp = app(ir-1) - apmp
        if(difp < 0.0d0) difp = - difp
        difp = (difp - crad)

        !assume symmetrical potential (r1-r2=rg1(ir-1)-rg2(ir-1))
        dif1 = (r1-r2)**2
        dif1 = (difc * difc + difp * difp) - dif1 !square of a square-root

        e = k2*dif1*dif1
        if (k2 /= 0.0) e = e + eg
        df = 4.0d0 * k2 * dif1 * (difc)
        if(ap < apmean) df = -df
        write(6,*) "e=",e,"-Tails"
    endif

    if (iflag == 2) return

    ! Modify DF to be DE/D(cos(tau)); also multiply by DRAVDR, which is
    ! dtau_ave/dtau(t) if IAVE=1, and 1.0 otherwise:

    if (k2 /= 0.0) df = df + df_g
    df = -dravdr*df/sphi

    ! Now calculate other dihedral derivatives as usual...

```