SIMULATIONS OF NUCLEIC ACIDS UNDER STRESS, IN SOLUTION, AND COMPLEXED TO PROTEINS

by

Jeffery M. Wereszcynski

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biophysics)
in The University of Michigan
2008

Doctoral Committee:

Assistant Professor Ioan Andricioaei, Co-Chair, University of California, Irvine
Associate Professor Heather A. Carlson, Co-Chair
Professor Gordon M. Crippen
Professor Noel C. Perkins
Associate Professor Hashim M. Al-Hashimi

ACKNOWLEDGEMENTS

The task of writing this dissertation and performing the research herein could not of been performed without several people and organizations to whom I owe a debt of gratitude. First I would like to thank Professor Andricioaei for his mentoring and support. Additionally, I would like to thank the members of the Andricioaei group: Catherine Musselman, Mariah Mills, Andy Boughton, Jeremiah Nummela, Aaron Frank, and Joanne Chen for the many wonderful scientific (and non-scientific) conversations, their help, and their support. A special thank you goes to Professor Carlson and her group for offering me a desk in their lab over these last six months, and for always making me feel welcome. Additional thanks go to my committee members, Professors Crippen, Perkins, and Al-Hashimi for their assistance and support. The Rackham Graduate School, Molecular Biophysics Training Grant, and National Science Foundation provided me with financial support, and the Pittsburgh Supercomputing Center, Department of Energy National Energy Research Scientific Computing Center, and the Teragrid provided essential computing resources without which the research presented here could not of been performed. I would like to thank my family and many friends for their support, and I would like to extend an additional thank you to Joanne for her support during the writing of this work.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

CHAPTER I

Introduction

> "Certainly no subject or field is making more progress on so many fronts at the present moment, than biology, and if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, *it is that all things are made of atoms* and that everything that living things do can be understood in terms of jiggling and wiggling of atoms." [1]

Feynman's words are as true today as they were 45 years ago. In our attempt to understand the function of biological macromolecules many points of view may be taken, from a continuum model which is concerned with overall rates of reactions (possibly encompassing hundreds of proteins) to a quantum mechanical one in which the wave functions for electron clouds surrounding individual atoms of interest are solved via high level *ab initio* methods. With different methods come different questions that may be answered. With molecular dynamics simulations we attempt to answer the question how do particular macromolecules function via the approach of Feynman, that is understanding how the "jiggling and wiggling of atoms" which define a molecule dictate its structure and function. This technique has proven to be invaluable in our understanding of many of the key molecules which make life

possible, especially DNA, RNA, and proteins.

1.1 Background

1.1.1 The Origins of Molecular Dynamics Simulations

At its heart molecular dynamics (MD) is a straightforward concept: given a system of particles and the forces between them the solutions of the simultaneous equations of classical motion can provide an unparalleled view of the time evolution of a system. In practice though there are many complications, not the least of which is the number of calculations required to solve these equations even once, let alone for an entire trajectory, thus requiring computers for these simulations. In 1957 the first MD simulation was performed by Alder and Wainwright [2]. In their calculations systems of hard spheres with a square well potential were simulated in a rectangular box with periodic boundary conditions (a procedure that had previously been employed in Monte Carlo simulations of similar system [3]). While they were able to simulate system of 32, 108, 256 and 500 particles, data for systems larger than 108 particles were not "presented due to inadequate statistics." Results of the study commented on the virial coefficients for finite systems with periodic boundary conditions.

Aneesur Rahman, who would become known as one of the fathers of molecular dynamics simulations [4], presented the first study of liquid argon with a simulation of 864 particles which interacted with one another via a Lennard-Jones potential [5]. Rahman also performed interesting studies on liquid water [6, 7] which lead to a key study of water in 1974 that introduced the ST2 water model and showed that reasonable approximations to neutron scattering and thermodynamic quantities could be obtained over a wide range of temperatures [8].

The first MD simulation on a biomolecule was published in 1977 by J. Andrew McCammon, Bruce R. Gelin and Martin Karplus [9]. It involved calculations on

bovine pancreatic trypsin inhibitor (BPTI) with a "crude molecular mechanics potential" [10] and lasted a mere 9.2 ps, however it proved instrumental in advancing our understanding of protein dynamics. For the first time an atomic view existed in which motions of the protein interior could be described as "fluid-like" and "sampled highly anharmonic regions of the potential surface."

1.1.2 Molecular Dynamics Simulations of Nucleic Acids

The early years: 1983-1994

It was not until 1983 that the first two MD simulations of nucleic acids were published, one by Michael Levitt [11] and the other by Bruce Tidor, Karl Irikura, Bernard Brooks and Martin Karplus [12]. In the study by Levitt, 90 ps MD calculations were performed on helices of length 12 and 24 base pairs. Calculations that excluded electrostatic interactions (but included a hydrogen bonding term) produced stable helices with global motions consistent with those from NMR measurements. Upon the inclusions of electrostatic interaction the helices become unstable and unwound, serving as an indication of failures in the force field. In the work by Tidor *et al.* MD simulations and harmonic calculations of three DNA hexamer systems (two in B form one in Z) were performed and compared with one another and with experimental parameters such as NMR order parameters, B-factors from x-ray diffraction, and fluorescence depolarization. While the authors acknowledged that the results were more qualitatively than quantitatively correct, the study began what would become the broad field of MD simulations on nucleic acids.

This was also the year the first simulations of RNA were published in which two 12 ps simulations of tRNA, differing from one another based on their van der Waals potentials, were compared with one another [13]. It was determined that there was little difference between the simulation which had parameters altered to mimic

solvation effects and that with the original parameters. This was one of the first studies in the subfield of implicit solvent calculations.

Advances in computing power allowed for the simulation of solvated DNA systems, with the first on a five base dodecamer surrounded by 830 water molecules and 8 sodium ions for 106 ps [14]. This study, along with many that would follow, demonstrated inaccuracies in the methods and force field through events such as base pair openings, significant bending of the DNA backbone, sugar puckering straying from the *C2'-endo* conformation, and high RMS deviations from canonical B form [15, 16]. The addition of restraining potentials to mimic Watson-Crick hydrogen bonds reduced some of these problems on a short timescale [17], however the first 1 ns MD simulation of a DNA used these restraints and reported fluctuations from B- form as large as 7.5 Å [18]. Nevertheless this period provided many exciting simulations which offered a glimpse of the dynamics intrinsic to nucleic acid molecules [19].

A renaissance of nucleic acid simulations: 1995-2000

Advances in computing power, methodology and force fields (see section 1.2.2 for a discussion of the CHARMM force field) allowed for great advances in the simulation of nucleic acids between 1995 and 2000 [20, 21]. The introduction of particle mesh Ewald methods (PME, discussed further in section 1.2.6) allowed for more accurate treatments of long range electrostatics which resulted in a significant increase of nucleic acid stability. This was highlighted in a 1995 paper by Cheatham *et al.* which compared identical simulations for DNA, RNA, and ubiquitin that differed only in the use of electrostatic methods [22]. It was noted that the nucleic acid structures were unstable with the standard cutoff algorithms but sampled structures close to the helical canonical forms with PME. In the case of ubiquitin all simulations sampled similar configurational space regardless of the electrostatic method used.

With the availability of stable simulations that lasted longer than 1 ns there were several questions that could be addressed, not the least of which was did simulations accurately predict structural and dynamical properties of nucleic acids? Several studies examined this with results that differed based on the chosen force field. Simulations with the CHARMM22 force field tended to overstabilize the A- conformation of DNA helices [23] whereas those with the AMBER Param94 parameters exhibited an undertwisting (resulting in ≈12 base pairs/turn) and O4' endo puckering [24]. Additionally the AMBER force field produced trajectories with B-RNA structures stabilized for over 10 ns [20]. These results would be instrumental in further force field refinements.

Other problems of interest that were addressed in this time were the interaction of cations with DNA [25, 26], sequence dependent bending [27], and stretching of the DNA helix [28, 29] (to name only a few). These studies of DNA under a tensile load served to illustrate the promise that molecular dynamics holds, that experiments with only a few observables (in this case extension and force) can be modeled, and through the combination of techniques a more complete understanding of the system of interest is gained than through the use of either technique separately.

Molecular Dynamics Simulations of Nucleic Acid in the 21st Century

Armed with molecular dynamics methodology, software, high speed computers, and more advanced force fields the field of MD simulations on nucleic acids has significantly expanded in the first decade of the 21st century to encompass almost every aspect of nucleic acid structure, dynamics, and function. A comprehensive review of papers published in this time would exceed the scope of this dissertation, thus a sampling of interesting works is highlighted to illustrate the types of simulations which are currently state of the art.

In 2002 two papers were published which examined the flipping of a single base out of a double helix (as has been observed in crystal structures of DNA with repair enzymes [30]). One of these studies used the CHARMM27 force field [31] while the other used AMBER Parm99 [32]. Both utilized umbrella sampling techniques (see Section 1.2.4), albeit with different definitions of a reaction coordinate. Results were remarkably similar between the two studies: flipping proceeded along distinct pathways for passage through the major or minor groove, flipping towards the major groove was slightly preferred, and the energy to flip a base was 15-20 kcal/mol. There were little or no local minima present in the flipped state, indicating that proteins must not only flip the base out but stabilize it afterwords.

A collaboration of 17 researchers in 9 labs (referred to as the "Ascona B-DNA Consortium") undertook the computationally expensive task of simulating all 136 possible DNA tetranucleotide steps and analyzing them for structural information [33, 34]. Results indicated the existence of seven backbone conformational substates determined by three parameters, the backbone dihedral angles α, γ, and the difference between ϵ and ζ. Transitions between these states were not necessarily reversible. Analysis also showed that while intrinsically rigid the structural parameters of purine/purine and purine/pyrimidine steps may experience a "significantly structural impact" from adjacent pyrimidine/purine steps. A database of these results was placed online (*http://humphry.chem.wesleyan.edu:8080/MDDNA/header.html*) where users may obtain structures for DNA of any desired sequence.

Another excellent example of modern MD simulations on DNA was recently provided by Harris *et al.* in which the supercoiling of DNA nanocircles was explored [35]. Explicit solvent calculations of 90 base pair DNA circles and implicit solvent calculations of 90 to 178 base pair systems for up to 5 ns were performed with var-

ious levels of positive and negative supercoils introduced. Of particular interest is the reported phase diagrams of DNA conformations for circles of various lengths and superhelical densities in medium and high salt environments. Results show an asymmetric response to positive and negative torsional stress which is not described by similar studies using course-grained rod models [36].

1.1.3 Molecular Dynamics Simulations of Protein-Nucleic Acid Systems

Historically the simulation of systems containing both proteins and nucleic acids has lagged behind those containing a single class of biomolecules, largely due to the size of the systems required for such calculations. The first of these studies was reported in 1989 by de Vlieg *et al.* who used NOE data to constrain an MD simulation of the *lac* repressor complexed to a 14 base pair DNA sequence [37]. This work provided a wealth of information about specific protein-DNA contacts (direct and water mediated hydrogen bonds, nonpolar interactions) and was an early demonstration of the value in using experimental parameters such as NOE data in MD simulations.

Simulations of these types remained sparse, but in 1995 an interesting study was published by Eriksson *et al.* in which a dimer of the DNA binding domain of the transcription factor glucocorticoid receptor was complexed with a 20 base pair DNA segment for 325 ps [38]. Important structural changes in both the DNA and the protein were noted as compared to the crystal structure, such as a bending of the DNA helix, expansion of sections of the DNA major groove to accommodate protein helices, and distortion of some protein secondary structure. Additionally a comparison to the crystal structure for the estrogen receptor/DNA complex showed surprising similarities between hydrogen bonding networks of the two systems, suggesting common features between the two proteins that could possibly extend to the entire family.

Advances in computing power and the introduction of MD software targeted at simulating large complexes (such as NAMD, see section 1.2.3) have made simulations of protein/nucleic acids complexes much more common. In recent years work on the nucleosome/DNA [39], λ-repressor/DNA [40], and the U1A/RNA [41, 42] complexes have been reported. With the biological importance of biomolecules interacting with one another it is expected that these types of studies will be performed with a high frequency in the future.

1.2 Computational Tools and Techniques

Key problems in molecular dynamics are the definition of the force field and the implementation of code. Indeed these presented major difficulties for early simulations as not only was the researcher required to develop algorithms and code to evaluate and differentiate energies and then propagate the resulting forces into dynamics, they also had to determine the proper formula and constants for the potential energy of the system of interest. The introduction of CHARMM and other packages helped to alleviate some of these problems by providing highly transferable and easily adaptable collections of code for MD, along with general force fields.

1.2.1 Molecular Dynamics

In molecular dynamics a system of atoms is represented as spherical points with a given radius which interact with one another via a potential energy function defined via a force field (for the CHARMM specific force field see section 1.2.2). Following directly from Newton's Second Law the spatial derivative of this force field is proportional to the acceleration for each of atoms. That is, for atom i the acceleration vector is [43, 44, 45]:

$$-\nabla V_i = m_i \ddot{r}_i \tag{1.1}$$

Time is integrated via one of numerous methods, example of which are the Verlet [46], leap-frog Verlet [47], and velocity-Verlet algorithms [48], all of which are theoretically equivalent in the limit of small time steps, but differ in accuracy due to numerical details.

1.2.2 Evolution of the CHARMM Force Field

Included with the original version of CHARMM was PARAM4, a set of parameters describing the interactions of atoms in amino acids [49]. An "extended atom" representation was used in which the radii and charges of heavy atoms were adjusted to include the hydrogen atoms which were bound to it thus reducing the size of the system by removing the need to explicitly include protons. The energy function consisted of a collection of terms including one to account for hydrogen bonding interactions.

Shortly following the release of PARAM4 a refinement was developed by Reiher, Brooks, and Karplus termed Param 19 and was commonly employed for simulations following 1985 [50]. A major enhancement was the use of extended atom representations for only non-polar hydrogens, while hydrogens bonded to oxygen or nitrogen atoms were explicitly represented and parametrized. It was shown that hydrogen bonding could be effectively reproduced through Lennard-Jones and Coulombic interactions resulting in the hydrogen bonding energy function not being included. Additional parameterization was performed such that water molecules, represented by a modification to the TIP3P model [51, 50, 52], could be introduced into the simulation.

For nucleic acids the first widely accepted CHARMM force field was constructed by Nilsson and Karplus and termed EF2 (energy function 2) [53] and was associated with the original version of CHARMM [49, 54, 55]. EF2 was an improvement to

EF1, the force field utilized in the original DNA MD simulation by Tidor *et al.*, and contained both "extended atom" and "all atom" models of nucleotides. The extended atom representations combined non-polar hydrogens with their respective heavy atoms while explicitly representing polar hydrogens while the all atom explicitly modeled these protons. Additionally a hydrogen bonding term was present in the energy function. Simulations were designed to be performed in vacuo and a distance dependent dielectric constant was often required to create stable trajectories [53].

Version 22 of CHARMM saw the introduction of the first "modern" CHARMM force fields for nucleic acids [56], proteins [57], and lipids [58] and was termed CHARMM22. This parameter set was designed to be an all-atom representation that included explicit solvent models of the TIP3P type and could allow for the interaction of proteins, nucleic acids, lipids. Hydrogen bond terms were dropped from all potential energy functions, such that the functional form of the potential energy became:

$$\begin{aligned} V(r) = & \sum_{bonds} K_b \left(b - b_0\right)^2 + \sum_{angles} K_\theta \left(\theta - \theta_0\right)^2 + \sum_{impropers} K_\psi \left(\psi - \psi_0\right)^2 \\ & + \sum_{dihedrals} K_\chi \left(1 + cos\left(n\chi - \delta\right)\right) + \sum_{Urey-Bradley} K_{UB} \left(S - S_0\right)^2 \\ & + \sum_{nonbonded} \left(\epsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\epsilon_d r_{ij}} \right). \end{aligned} \tag{1.2}$$

The first five terms represent the interactions of bonded atoms with terms for bond distance, angles, improper and regular dihedrals, and a Urey-Bradley term (the distance between two atoms bonded to the same atom) respectively. The K values are the associated force constants, the values with a subscript of zero the equilibrium values, and the dihedrals term has a periodicity of n and an offset of δ. The final summation is over nonbonded interactions, specifically the Lennard-Jones and Coulombic interactions over each pair of atoms.

The CHARMM22 force field allowed for a substantial increase in the stability of nucleic acids, however several problems remained such as the overstabilziation of A-DNA compared to B-DNA helices and a nanosecond timescale fluctuation between A- and B-DNA structures [23]. The CHARMM27 force field resolved many of these problems by reoptimizing the nucleic acid parameters based on a combination of quantum mechanical and experimental data [59]. It was shown that this new force field correctly sampled the A- B-DNA equilibrium in aqueous and non-aqueous environments and provided good agreement with various structural properties such as backbone and puckering angles and helicoidal parameters [60, 61]. At this time the lipid force field was also updated [62].

While more accurate in its description of proteins than nucleic acids the CHARMM22 force field exhibited several deficiencies such as incorrect dipeptide potential energy surfaces (when compared to high-level QM calculations) and the presence of peptide π-helices in lipid and aqueous environments. With the aid of QM calculations that included electron correlation effects on alanine, glycine, and proline dipeptides MacKerrell *et al.* derived a correction to the CHARMM22 force field termed CMAP [63]. This correction added a 2-D energy correction map to the ϕ,ψ backbone dihedrals which is interpolated by the MD software. Results suggest substantial improvement in the handling of proteins such as the reproduction of α- instead of π-helices, a lower RMS to protein crystal structures, and a better fit to N-H NMR order parameters [64, 65].

1.2.3 Selected Molecular Dynamics Software

CHARMM

The CHARMM program originated in the early 1970s through the work of Martin Karplus and his graduate student Bruce Gelin [66]. The earliest version was able

to perform energy evaluations of amino acid sequences at a given set of cartesian coordinates and energy minimization. Its functionality was significantly extended to include molecular dynamics techniques, normal mode analysis, and several forms of structural analysis through several revisions. It was released to the public in the early 1980s under the name CHARMM (Chemistry at HARvard Macromolecular Mechanics) [49]. Release required naming of the previously nameless code, and the early suggestion of HARMM by Bob Bruccoleri was modified by the addition of a "C", however Martin Karplus has noted that the original name may have been appropriate [66]:

> "I sometimes wonder if Bruccoleri's original suggestion would have served as a useful warning to inexperienced scientists working with the new program."

Since its introduction CHARMM has been maintained and developed by several groups, providing the advantage of a broad range of functionality including (but not limited to) implicit and explicit solvent molecular dynamics simulations, QM/MM calculations, implicit membrane calculations, computational alchemical methods, thermodynamic integration, and a large collection of constraints and analysis tools. It is currently used worldwide in academic and commercial versions (the latter through Accelyrs).

NAMD

With the advent of large supercomputing clusters came a need for efficient MD code parallelization, and in 1996 the program NAMD (Not just Another Molecular Dynamics program) was introduced by the Schulten group to take advantage of the newest supercomputers for MD calculations [67, 68]. NAMD effectively scales to

thousands of processors and has been used for large scale simulations such as the simulation of an entire virus (the satellite tobacco mosaic virus), the first molecular dynamics simulation for a biomolecule of over 1 million atoms [69]. The CHARMM, AMBER, or GROMACS force fields may be used. Code is maintained solely by the Theoretical and Computational Biophysics Group at the University of Illinois at Urbana-Champaign which enables developers to ensure high efficiency for code scaling. While limited in its functionality compared to CHARMM, NAMD is an excellent choice for large-scale explicit solvent simulations.

1.2.4 Umbrella Sampling

Calculation of the potential of mean force (PMF) for a system along an arbitrary reaction coordinate ζ is a complex problem that, if properly solved, may provide valuable insights into the function and interactions of biomolecules. In a direct approach a probability function $P(\zeta)$ could be constructed from an unrestrained trajectory and the PMF calculated by: $W(\zeta) = -k \log(P(\zeta))$. This is however impracticable for systems studied with MD as the range of phase space sampled by the simulation will be heavily biased towards the area surrounding the local minima near the initial structure.

The need for enhanced sampling was recognized as early as 1974 when Torrie and Valleau introduced what has come to be termed "umbrella sampling" [70, 71]. In this method a restraining potential as a function of ζ is added to the Hamiltonian of the system to bias it into simulating a region that would not otherwise be sampled. A "window" is defined as a single simulation in which a restraining potential w_i is added such that sampling is enhanced in an area ζ_i. Often times a harmonic potential is used so $w_i = K_i (\zeta - \zeta_i)^2$ with the force constant K_i to be chosen, although this is not a unique choice and any restorative potential may be used (quartic for example).

For sampling of a wide range of phase space successive windows may be run with restraining potentials centered on different values of ζ such that biased sampling occurs on a much greater range then would be possible with unbiased simulations.

The question then becomes how does one properly unbias the data from these windows to construct an unbiased probability distribution (which may then be used to calculate a PMF). Several methods have been devised to do this [72, 73, 74], however many of them suffered from problems such as requiring a large overlap between adjacent windows, large errors due to an arbitrary choice of force constants, and not using the entire set of windows in optimizing overlaps. Kumar *et .al* solved these problems with the introduction of the Weighted Histogram Analysis Method (WHAM) [75]. In this technique the mean unbiased probability at a given position, $\langle\rho(\zeta)\rangle$, is determined through solving the WHAM equations:

$$\langle\rho\left(\zeta\right)\rangle = \frac{\sum_{i=1}^{N_w} n_i \rho\left(\zeta\right)_{(i)}}{\sum_{i=1}^{N_w} n_i e^{-[w_i(\zeta)-F_i]/kT}}$$

$$F_i = -kT \log\left[\sum_{i=1}^{N_{bins}} \langle\rho\left(\zeta\right)\rangle e^{-[w_i(\zeta)/kT]}\right]. \tag{1.3}$$

Due to the probabilities being dependent on force constants F_i that must be determined, the above equations are solved iteratively until they are self consistent. WHAM minimizes the error associated with overlapping windows with one another and does not discard any data in the process [76]. Additionally WHAM may be extended to an arbitrary number of dimensions (although the number of simulations required to fully cover an n-dimensional phase space goes by $O(x^n)$) [77].

There are two major considerations which affect the applicability and accuracy of free energy calculations via umbrella sampling and WHAM. The first is the definition

of a reaction coordinate and the implementation of a restraining potential defined by this coordinate. While in some cases this may be straightforward, such as in the extension or twist of a DNA molecule [78, 79], in others it may be more complicated and not feasible. The second is sufficient overlap of windows and sampling within windows required for convergence of the PMF. This is solved through an interplay of the number of windows simulated, the range between windows, the force constant K_i used in each window, and the simulation time in each window.

1.2.5 Calculation of Conformational Entropy

The stability of a biological molecule has been shown to be a complex interplay between entropic and enthalpic factors. While molecular dynamics simulations provide a good representation of a system's enthalpic components, entropic contributions have proved much more difficult to determine. Methods exist for calculation of the conformational entropy for a given biomolecule, with harmonic and quasiharmonic analysis as the two methods of choice.

Normal Mode

In normal mode analysis a mass-weighted matrix of second derivatives of the energy with respect to the $3n$ cartesian coordinates of the structure (the Hessian) is diagonalized [80]. The eigenvectors and eigenvalues are calculated to give the motions and frequencies intrinsic to the molecule in that state, and the entropy of a biomolecule may be calculated from these frequencies using the same formula as derived by Andricioaei and Karplus for quasiharmonic analysis (below, equation 1.5) [81]. Calculations are limited in that they only reflect the energy minima for which the molecule is in, however unlike with quasiharmonic analysis it is not necessary to calculate trajectories for determination of thermodynamic quantities.

Quasiharmonic Analysis

An alternative approach was introduced in 1981 by Karplus and Kushick in which the covariance matrix calculated from a trajectory, C, may be used to calculate the entropic difference between two conformations as $k/2\ \ln\left(det\ C_a/det\ C_b\right)$ [82]. The covariance matrix C is defined as:

$$C_{ij} = \langle (x_i - \langle x_i \rangle)\,(x_j - \langle x_j \rangle) \rangle\,. \tag{1.4}$$

This approach however tends to be numerically unstable and relies on a conversion to internal coordinates [83]. Schlitter circumvented these problems by deriving a method to calculate the conformational entropy directly from the mass-weighted covariance matrix of cartisian coordinates [84]. This method utilizes an approximation to the quantum mechanical harmonic oscillator and provides an upper bound to the macromolecule's conformational entropy. Andricioaei and Karplus derived a formulation in which the exact QM oscillator frequency could be utilized, thus providing a formalism with a tighter upper bound to the conformational entropy [85]. In this method the eigenvalues of the mass weighted covariance matrix are calculated to determine the quasiharmonic frequencies $\omega_i = \sqrt{kT/\lambda_i}$ to calculate the entropy via:

$$S = k \sum_{i}^{3n-6} \frac{\hbar\omega_i/kT}{e^{\hbar\omega_i/kT} - 1} - \ln\left(1 - e^{-\hbar\omega_i/kT}\right). \tag{1.5}$$

The major assumption of this approach is that the fluctuations represented in the covariance matrix are a result of a multivariate Gaussian probability distribution, i.e. that the underlying potential energy may be described as a sum of harmonic terms, a reasonable assumption in molecular dynamics. Additionally there is the underlying problem of sufficient sampling of configurational space. It has been noted that as sampling time increases the entropy follows the formula [86, 87]:

$$S(t) = S_{inf} - \frac{\beta}{t^{\alpha}} \tag{1.6}$$

thus it is normally necessary to calculate the entropy for several simulation lengths to extrapolate the entropy for an infinitely long simulation. The constants α and β are fitting constants and it is not immediately clear whether they have a physical significance, although work on this is under investigation [87].

While quasiharmonic analysis is computationally more demanding then normal mode analysis in that it requires the calculation of MD trajectories, it provides the advantage of sampling a wider range of phase space and explores motions which may not be represented in the energy minima that normal modes explore.

1.2.6 Electrostatic Calculations

Energy and force calculations for large systems require substantial computational power, with non-bonded interactions dominating those calculations. Electrostatics calculations are highly demanding as a coulombic potential must be calculated between each pair of atoms, creating a problem on the $O(n^2)$ timescale. To alleviate some of this computational burden two methods have been developed: truncation of electrostatic interactions after a given length and the particle mesh Ewald method. In the first electrostatics are calculated normally between atoms i and j if the distance between them is less then a given cutoff r_1, a switching function is used if the distance between them (x) satisfies $r_1 < x < r_2$, and for a separation $x > r_2$ no interaction is calculated [88]. With older force fields this method created instabilities, especially for charged molecules such as DNA [89, 16], however truncation methods do produce stable trajectories with newer force fields [90, 91].

A more rigorous approach to electrostatic calculations is the particle mesh Ewald

method [92, 93]. In this scheme the electrostatics are separated into two terms, a short-range and a long-range term. The short range term is summed over atoms within a given cutoff and sums quickly using a standard Coulombic potential. For the long-range potentials the charge density of the periodic system is calculated on a lattice, following which a fast Fourier transform is employed to calculate the Coulombic potential at a specific location. This method operates on an $O(N \log N)$ timescale (slower than truncation methods but much faster then calculations of all electrostatic pairs) and is highly accurate. Its main assumption is that of an infinitely periodic system, and due to periodicity the net charge of the system must be zero.

1.3 Structures of Nucleic Acids Helices

Nucleic acids may exist in one of several regular helical patterns. The original structure solved by Watson and Crick [94] came to be termed "B-DNA," but at lower humidity it may transition to "A-DNA" [95, 96, 97]. Both structures have a right handed twist, however A-form has a lower rise and twist angle, creating more bases per turn, and the bases are tilted with respect to the central axis.

A quite different structure from A- and B- form was solved in 1979 by Andrew Wang, Alexander Rich and coworkers and was termed "Z-DNA" [98]. In this structure a six base DNA sequence of alternating cytosines and guanines exhibited a left-handed turn with rise greater then that of B-DNA. This structure was also unique in that it had a dinucleotide repeat (as opposed to a single repeat in A- or B- form) creating six units (or twelve bases) per turn.

Each nucleotide consists of a phosphate group (which when joined with other bases comprises the phosphate backbone), a sugar, and a base [99]. The conformation of the sugar is of particular interest and has been shown to be highly correlated with

Helix	Sense	Twist (°)	Rise (Å)	Units/turn	Base inclination(°)	Sugar Puckering
A-form	Right-Handed	32.7	2.54	11	22.6	C3'-endo
B-form	Right-Handed	36.0	3.38	10	2.8	C2'-endo
Z-form (C)	Left-Handed	-49.3	3.63	6	0.1	C2'-endo
Z-form (G)	Left-Handed	-10.3	3.63	6	0.1	C2'-exo

Table 1.1: Structural Properties of A-, B-, and Z- form nucleic acid helices.

helical structure [97]. To relieve steric clash one of the atoms in the five membered sugar is forced out of plane, and it is this atom, which is said to be "puckered," that may change between different configurations. Depending on the side of the ring it is on the atom may exhibit "endo" or "exo" puckering. Standard puckering, along with other helical parameters, are given in Table 1.1 as reported by Neidle [97].

1.4 Specific Aims

This dissertation aims at advancing our understanding of nucleic acids, their interaction with proteins, and their simulations.

1. In Chapter 2 we present, at atomic resolution, a simulation of the dynamics involved in the transitions from B-DNA and A-RNA to Pauling (P) forms [100, 101] and to denatured states driven by application of external torque and tension. We then calculate the free energy profile along a B to P transition coordinate and from it, compute a reversible pathway, i.e, an isotherm of tension and torque pairs required to maintain P-DNA in equilibrium. The reversible isotherm maps correctly onto a phase diagram derived from single molecule experiments, and yields values of elongation, twist, and twist-stretch coupling in agreement with measured values. We also show that configurational entropy compensates significantly for the large electrostatic energy increase due to closer-packed P backbones. A similar set of simulations applied to RNA are used to predict a novel structure, P-RNA, with its associated free energy, equi-

librium tension, torque and structural parameters, and to assign the location, on the phase-diagram, of a putative force-torque dependent RNA "triple point".

2. Thermoresponsive devices may be constructed based on the entropic difference between left and right handed nucleic acids and its relation to sequence and environmental conditions [102, 103]. Despite the importance of nanoscale temperature measurements a detailed understanding of the microscopic origin of the temperature induced switching activity of RNA and DNA based nanothermometers is lacking. To address it, we present in Chapter 3 MD simulations exploring the temperature-dependent structural dynamics and energetics of these devices. We present configurational, solvent, and ionic entropies calculations on a set of RNA and DNA systems that cover the landscape of sequence, salt conditions, and helical direction to determine their effects on transitions from right to left handed structures and vice versa.

3. Human topoisomerase I is thought to remove DNA supercoils by nicking one strand at a phosphodiester bond, covalently attaching to the 3' end of the nick, and allowing the downstream DNA to rotate around the intact strand [104]. In Chapter 4 we calculate free energy profiles for the rotation of downstream DNA to mimic the release of both positive and negative supercoils which confirms previous experimental and computational work suggesting these mechanisms progress along distinct pathways. Additionally, simulations performed with the ternary complex of topoisomerase, DNA, and the chemotherapeutic drug topotecan show important differences in the mechanisms for DNA relaxation in this system, possibly accounting for the decreased rate of supercoil relaxation observed in experiments. Calculations show a good agreement between

experimentally observed rate constants and those recovered from simulations. Calculations also present evidence for a semi-open state of the protein which facilitates rotations after the initial one as a result of biasing the protein into a conformation more favorable to strand rotation than the clamped state required for nicking of the DNA

4. In Chapter 5 a method exploiting the properties of an artificial (non-physical) Langevin dynamics with a negative frictional coefficient along a suitable manifold and positive friction in the perpendicular directions is presented for the enhanced calculation of time correlation functions for rare event problems. Exact time-correlation functions that describe the kinetics of the transitions for the all-positive, physical system can be calculated by reweigthing the generated trajectories according to stochastic path integral treatment involving a functional weight based on an Onsager-Machlup action functional. The method is tested on a prototypical multidimensional model system featuring the main elements of conformational space characteristic of complex condensed matter systems. Using the present method, accurate estimates of rate constants require at least three order of magnitudes fewer trajectories than regular Langevin dynamics. The method is particularly useful in calculating kinetic properties in the context of multidimensional energy landscapes characteristic of complex systems such as proteins and nucleic acids.

BIBLIOGRAPHY

[1] R.P. Feynman, R.B. Leighton, and M. Sands. *The Feynman Lectures in Physics*. Addison-Wesley, 1963.

[2] B. J. Alder and T. E. Wainwright. Phase transition for a hard sphere system. *Journal Of Chemical Physics*, 27:1208–1209, 1957.

[3] B. J. Alder, S. P. Frankel, and V. A. Lewinson. Radial distribution function calculated by the monte-carlo method for a hard sphere fluid. *Journal Of Chemical Physics*, 23:417–419, 1955.

[4] W. F. Van Gunsteren and H. Berendsen. Computer simulation of molecular dynamics methodology applications and perspectives in chemistry. *Angewandte Chemie International Edition in English*, 29:992–1023, 1990.

[5] A. Rahman. Correlations in motion of atoms in liquid argon. *Physical Review A-General Physics*, 136:A405–&, 1964.

[6] A. Rahman and F. H. Stillinger. Molecular dynamics study of liquid water. *Journal Of Chemical Physics*, 55:3336–&, 1971.

[7] F. H. Stillinger and A. Rahman. Molecular dynamics study of temperature effects on water structure and kinetics. *Journal Of Chemical Physics*, 57:1281–&, 1972.

[8] F. H. Stillinger and A. Rahman. Improved simulation of liquid water by molecular-dynamics. *Journal Of Chemical Physics*, 60:1545–1557, 1974.

[9] J. A. McCammon, B. R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, 267:585–590, 1977.

[10] M. Karplus and J. A. McCammon. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 9:646–652, 2002.

[11] M. Levitt. Computer simulation of DNA double-helix dynamics. *Cold Spring Harb Symp Quant Biol*, 47 Pt 1:251–62, 1983.

[12] B. Tidor, K. K. Irikura, B. R. Brooks, and M. Karplus. Dynamics of DNA oligomers. *Journal Of Biomolecular Structure & Dynamics*, 1:231–252, 1983.

[13] M. Prabhakaran, S. C. Harvey, B. Mao, and J. A. McCammon. Molecular-dynamics of phenylalanine transfer-RNA. *Journal Of Biomolecular Structure & Dynamics*, 1:357–369, 1983.

[14] G. L. Seibel, U. C. Singh, and P. A. Kollman. A molecular-dynamics simulation of double-helical B-DNA including counterions and water. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 82:6537–6540, 1985.

[15] F. Briki, J. Ramstein, R. Lavery, and D. Genest. Evidence for the stochastic nature of base pair opening in DNA - a brownian dynamics simulation. *Journal Of The American Chemical Society*, 113:2490–2493, 1991.

[16] K. Miaskiewicz, R. Osman, and H. Weinstein. Molecular-dynamics simulation of the hydrated d(cgcgaattcgcg)$_2$ dodecamer. *Journal Of The American Chemical Society*, 115:1526–1537, 1993.

[17] S. Swaminathan, G. Ravishanker, and D. L. Beveridge. Molecular-dynamics of B-DNA including water and counterions - a 140-ps trajectory for d(cgcgaattcgcg) based on the GROMOS force-field. *Journal Of The American Chemical Society*, 113:5027–5040, 1991.

[18] K. J. McConnell, R. Nirmala, M. A. Young, G. Ravishanker, and D. L. Beveridge. A nanosecond molecular-dynamics trajectory for a B-DNA double helix - evidence for substates. *Journal Of The American Chemical Society*, 116:4461–4462, 1994.

[19] D. L. Beveridge and G. Ravishanker. Molecular-dynamics studies of DNA. *Current Opinion In Structural Biology*, 4:246–255, 1994.

[20] T. E. Cheatham and P. A. Kollman. Molecular dynamics simulation of nucleic acids. *Annual Review Of Physical Chemistry*, 51:435–471, 2000.

[21] T. E. Cheatham and M. A. Young. Molecular dynamics simulation of nucleic acids: Successes, limitations, and promise. *Biopolymers*, 56:232–256, 2000.

[22] T. E. Cheatham, J. L. Miller, T. Fox, T. A. Darden, and P. A. Kollman. Molecular-dynamics simulations on solvated biomolecular systems - the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *Journal Of The American Chemical Society*, 117:4193–4194, 1995.

[23] M. Feig and B. M. Pettitt. Structural equilibrium of DNA represented with different force fields. *Biophysical Journal*, 75:134–149, 1998.

[24] T. E. Cheatham and P. A. Kollman. Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *Journal Of Molecular Biology*, 259:434–444, 1996.

[25] M. Feig and B. M. Pettitt. Sodium and chlorine ions as part of the DNA solvation shell. *Biophysical Journal*, 77:1769–1781, 1999.

[26] N. V. Hud and M. Polak. DNA-cation interactions: the major and minor grooves are flexible ionophores. *Current Opinion In Structural Biology*, 11:293–301, 2001.

[27] D. Strahs and T. Schlick. A-tract bending: Insights into experimental structures by computational models. *Journal Of Molecular Biology*, 301:643–663, 2000.

[28] M. W. Konrad and J. I. Bolonick. Molecular dynamics simulation of DNA stretching is consistent with the tension observed for extension and strand separation and predicts a novel ladder structure. *Journal Of The American Chemical Society*, 118:10989–10994, 1996.

[29] K. M. Kosikov, A. A. Gorin, V. B. Zhurkin, and W. K. Olson. DNA stretching and compression: Large-scale simulations of double helical structures. *Journal Of Molecular Biology*, 289:1301–1326, 1999.

[30] S. Klimasauskas, S. Kumar, R. J. Roberts, and X. D. Cheng. Hhal methyltransferase flips its target base out of the DNA helix. *Cell*, 76:357–369, 1994.

[31] N. K. Banavali and A. D. MacKerell. Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *Journal Of Molecular Biology*, 319:141–160, 2002.

[32] P. Varnai and R. Lavery. Base flipping in DNA: Pathways and energetics studied with molecular dynamic simulations. *Journal Of The American Chemical Society*, 124:7272–7273, 2002.

[33] D. L. Beveridge, G. Barreiro, K. S. Byun, D. A. Case, T. E. Cheatham, S. B. Dixit, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, E. Seibert, H. Sklenar, G. Stoll, K. M. Thayer, P. Varnai, and M. A. Young. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(C(p)G) steps. *Biophysical Journal*, 87:3799–3813, 2004.

[34] S. B. Dixit, D. L. Beveridge, D. A. Case, T. E. Cheatham, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, H. Sklenar, K. M. Thayer, and P. Varnai. 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. *Biophysical Journal*, 89:3721–3740, 2005.

[35] S. A. Harris, C. A. Laughton, and T. B. Liverpool. Mapping the phase diagram of the writhe of DNA nanocircles using atomistic molecular dynamics simulations. *Nucleic Acids Research*, 36:21–29, 2008.

[36] J. F. Marko and E. D. Siggia. Statistical-mechanics of supercoiled DNA. *Physical Review E*, 52:2912–2938, 1995.

[37] J. de Vlieg, H. J. C. Berendsen, and W. F. Van Gunsteren. An NMR-based molecular-dynamics simulation of the interaction of the lac repressor headpiece and its operator in aqueous-solution. *Proteins-Structure Function And Genetics*, 6:104–127, 1989.

[38] M. A. L. Eriksson, T. Hard, and L. Nilsson. Molecular-dynamics simulations of the glucocorticoid receptor DNA-binding domain in complex with DNA and free in solution. *Biophysical Journal*, 68:402–426, 1995.

[39] T. C. Bishop. Molecular dynamics simulations of a nucleosome and free DNA. *Journal Of Biomolecular Structure & Dynamics*, 22:673–685, 2005.

[40] M. Saito and A. Sarai. Free energy calculations for the relative binding affinity between DNA and lambda-repressor. *Proteins-Structure Function And Genetics*, 52:129–136, 2003.

[41] F. Pitici, D. L. Beveridge, and A. M. Baranger. Molecular dynamics simulation studies of induced fit and conformational capture in U1A-RNA binding: Do molecular substates code for specificity? *Biopolymers*, 65:424–435, 2002.

[42] Y. Zhao, B. L. Kormos, D. L. Beveridge, and A. M. Baranger. Molecular dynamics simulation studies of a protein-RNA complex with a selectively modified binding interface. *Biopolymers*, 81:256–269, 2006.

[43] M.P. Allen and D.J. Tildesley. *Computer Simulations of Liquids*. Oxford University Press, 1987.

[44] D. Frenkel and B. Smit. *Understanding Molecular Simulation*. Elsevier, 2002.

[45] D. C. Rapaport. *The Art of Molecular Dynamics Simulation*. Cambridge University Press, 2004.

[46] L. Verlet. Computer experiments on classical fluids .I. Thermodynamical properties of lennard-jones molecules. *Physical Review*, 159:98, 1967.

[47] R.W. Hockney. The potential calculation and some applications. *Methods of Computational Physics*, 9:136–211, 1970.

[48] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson. A computer-simulation method for the calculation of equilibrium-constants for the formation of physical clusters of molecules - application to small water clusters. *Journal Of Chemical Physics*, 76:637–649, 1982.

[49] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM - a program for macromolecular energy, minimization, and dynamics calculations. *Journal Of Computational Chemistry*, 4:187–217, 1983.

[50] E. Neria, S. Fischer, and M. Karplus. Simulation of activation free energies in molecular systems. *Journal Of Chemical Physics*, 105:1902–1921, 1996.

[51] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *Journal Of Chemical Physics*, 79:926–935, 1983.

[52] P. Mark and L. Nilsson. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *Journal Of Physical Chemistry A*, 105:9954–9960, 2001.

[53] L. Nilsson and M. Karplus. Empirical energy functions for energy minimization and dynamics of nucleic-acids. *Journal Of Computational Chemistry*, 7:591–616, 1986.

[54] A. D. MacKerell, N. Banavali, and N. Foloppe. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, 56:257–265, 2000.

[55] A. D. MacKerell and L. Nilsson. *Computational Studies of RNA and DNA*, chapter 3. Springer, 2006.

[56] A. D. MacKerell, J. Wiorkiewiczkuczera, and M. Karplus. An all-atom empirical energy function for the simulation of nucleic-acids. *Journal Of The American Chemical Society*, 117:11946–11975, 1995.

[57] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal Of Physical Chemistry B*, 102:3586–3616, 1998.

[58] M. Schlenkrich, J. Brickmann, A. D. MacKerell, and M. Karplus. *In Biological Membranes: A Molecular Perspective from Computation and Experiment*, pages 31–81. Boston, 1996.

[59] N. Foloppe and A. D. MacKerell. All-atom empirical force field for nucleic acids: I. parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal Of Computational Chemistry*, 21:86–104, 2000.

[60] A. D. MacKerell and N. K. Banavali. All-atom empirical force field for nucleic acids: II. application to molecular dynamics simulations of DNA and RNA in solution. *Journal Of Computational Chemistry*, 21:105–120, 2000.

[61] S. Y. Reddy, F. Leclerc, and M. Karplus. DNA polymorphism: A comparison of force fields for nucleic acids. *Biophysical Journal*, 84:1421–1449, 2003.

[62] S. E. Feller and A. D. MacKerell. An improved empirical potential energy function for molecular simulations of phospholipids. *Journal Of Physical Chemistry B*, 104:7510–7515, 2000.

[63] A. D. Mackerell, M. Feig, and C. L. Brooks. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal Of Computational Chemistry*, 25:1400–1415, 2004.

[64] D. I. Freedberg, R. M. Venable, A. Rossi, T. E. Bull, and R. W. Pastor. Discriminating the helical forms of peptides by NMR and molecular dynamics simulation. *Journal Of The American Chemical Society*, 126:10478–10484, 2004.

[65] M. Buck, S. Bouguet-Bonnet, R. W. Pastor, and A. D. MacKerell. Importance of the CMAP correction to the CHARMM22 protein force field: Dynamics of hen lysozyme. *Biophysical Journal*, 90:L36–L38, 2006.

[66] M. Karplus. Spinach on the ceiling: A theoretical chemist's return to biology. *Annual Review Of Biophysics And Biomolecular Structure*, 35:1–47, 2006.

[67] M. T. Nelson, W. Humphrey, A. Gursoy, A. Dalke, L. V. Kale, R. D. Skeel, and K. Schulten. NAMD: A parallel, object oriented molecular dynamics program. *International Journal Of Supercomputer Applications And High Performance Computing*, 10:251–268, 1996.

[68] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with NAMD. *Journal Of Computational Chemistry*, 26:1781–1802, 2005.

[69] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure*, 14:437–449, 2006.

[70] G. M. Torrie and J. P. Valleau. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the subcritical Lennard-Jones fluid. *Chem. Phys. Lett.*, 28:578–581, 1974.

[71] G. M. Torrie and J. P. Valleau. Non-physical sampling distributions in Monte-Carlo free-energy estimation - Umbrella sampling. *Journal of Computational Physics*, 23:187–199, 1977.

[72] S. H. Northrup, M. R. Pear, C. Y. Lee, J. A. McCammon, and M. Karplus. Dynamical theory of activated processes in globular-proteins. *Proceedings Of The National Academy Of Sciences Of The United States Of America-Biological Sciences*, 79:4035–4039, 1982.

[73] C. Haydock, J. C. Sharp, and F. G. Prendergast. Tryptophan-47 rotational isomerization in variant-3 scorpion neurotoxin - a combination thermodynamic perturbation and umbrella sampling study. *Biophysical Journal*, 57:1269–1279, 1990.

[74] J. Shen and J. A. McCammon. Molecular-dynamics simulation of superoxide interacting with superoxide-dismutase. *Chemical Physics*, 158:191–198, 1991.

[75] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. *J. Comput. Chem*, 13:1011–1021, 1992.

[76] B. Roux. The calculation of the potential of mean force using computer-simulations. *Computer Physics Communications*, 91:275–282, 1995.

[77] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. Multidimensional free-energy calculations using the weighted histogram analysis method. *Journal Of Computational Chemistry*, 16:1339–1350, 1995.

[78] S. Piana. Structure and energy of a DNA dodecamer under tensile load. *Nucleic Acids Research*, 33:7029–7038, 2005.

[79] S. Kannan, K. Kohlhoff, and M. Zacharias. B-DNA under stress: over-and untwisting of DNA during molecular dynamics simulations. *Biophysical Journal*, 91:2956–2965, 2006.

[80] B. Brooks and M. Karplus. Harmonic dynamics of proteins - normal-modes and fluctuations in bovine pancreatic trypsin-inhibitor. *Proceedings Of The National Academy Of Sciences Of The United States Of America-Biological Sciences*, 80:6571–6575, 1983.

[81] M. Levitt, C. Sander, and P. S. Stern. Protein normal-mode dynamics - trypsin-inhibitor, crambin, ribonuclease and lysozyme. *Journal Of Molecular Biology*, 181:423–447, 1985.

[82] M. Karplus and J. N. Kushick. Method for estimating the configurational entropy of macromolecules. *Macromolecules*, 14:325–332, 1981.

[83] A. Di Nola, H. J. C. Berendsen, and O. Edholm. Free-energy determination of polypeptide conformations generated by molecular-dynamics. *Macromolecules*, 17:2044–2050, 1984.

[84] J. Schlitter. Estimation of absolute and relative entropies of macromolecules using the covariance-matrix. *Chemical Physics Letters*, 215:617–621, 1993.

[85] I. Andricioaei and M. Karplus. On the calculation of entropy from covariance matrices of the atomic fluctuations. *Journal of Chemical Physics*, 115:6289–6292, 2001.

[86] S. A. Harris, E. Gavathiotis, M. S. Searle, M. Orozco, and C. A. Laughton. Cooperativity in drug-DNA recognition: A molecular dynamics study. *Journal Of The American Chemical Society*, 123:12658–12663, 2001.

[87] S. A. Harris and C. A. Laughton. A simple physical description of DNA dynamics: quasi-harmonic analysis as a route to the configurational entropy. *Journal Of Physics-Condensed Matter*, 19, 2007.

[88] P. J. Steinbach and B. R. Brooks. New spherical-cutoff methods for long-range forces in macromolecular simulation. *Journal Of Computational Chemistry*, 15:667–683, 1994.

[89] S. C. Harvey. Treatment of electrostatic effects in macromolecular modeling. *Proteins-Structure Function And Genetics*, 5:78–92, 1989.

[90] J. Norberg and L. Nilsson. On the truncation of long-range electrostatic interactions in DNA. *Biophysical Journal*, 79:1537–1553, 2000.

[91] E. Giudice and R. Lavery. Simulations of nucleic acids and their complexes. *Accounts Of Chemical Research*, 35:350–357, 2002.

[92] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald - an $N \cdot log(N)$ method for Ewald sums in large systems. *Journal Of Chemical Physics*, 98:10089–10092, 1993.

[93] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh Ewald method. *Journal Of Chemical Physics*, 103:8577–8593, 1995.

[94] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.

[95] W. Saenger. *Principles of nucleic acid structure*. Springer-Verlag, New York, 1984.

[96] S. Neidle. *Oxford Handbook of Nucleic Acid Structure*. Oxford University Press, 1999.

[97] S. Neidle. *Principles of Nucleic Acid Structure*. Oxford University Press, 2008.

[98] A. H. J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. Vanboom, G. Vandermarel, and A. Rich. Molecular-structure of a left-handed double helical DNA fragment at atomic resolution. *Nature*, 282:680–686, 1979.

[99] Lodish et al. *Molecular Cell Biology*. W.H. Freeman and Company, 2004.

[100] J. F. Allemand, D. Bensimon, R. Lavery, and V. Croquette. Stretched and overwound DNA forms a pauling-like structure with exposed bases. *Proceedings of The National Academy of Sciences of The United States of America*, 95:14152–14157, 1998.

[101] Z. Bryant, M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante. Structural transitions and elasticity from torque measurements on DNA. *Nature*, 424:338–341, 2003.

[102] R. Tashiro and H. Sugiyama. A nanothermometer based on the different pi stackings of B- and Z-DNA. *Angewandte Chemie-International Edition*, 42:6018–6020, 2003.

[103] R. Tashiro and H. Sugiyama. Biomolecule-based switching devices that respond inversely to thermal stimuli. *Journal Of The American Chemical Society*, 127:2094–2097, 2005.

[104] D. A. Koster, K. Palle, E. S. M. Bot, M. A. Bjornsti, and N. H. Dekker. Antitumour drugs impede DNA uncoiling by topoisomerase I. *Nature*, 448:213–217, 2007.

CHAPTER II

On Structural Transitions, Thermodynamic Equilibrium and the Phase Diagram of DNA and RNA Duplexes under Torque and Tension

2.1 Introduction

In several instances involving DNA, nucleoprotein complexes exert, *in vivo*, forces or torques that distort it appreciably e.g., by stretching [1], wrapping-around [2], or looping [3]. Furthermore, a quantitative assessment of double stranded DNA deformation can aid in designing novel nanomechanical devices [4], and in perfecting rapid genetic mapping techniques for stretched, surface-immobilized DNA [5, 6]. While less studied in this respect, double stranded RNA, too, can experience significant structural perturbations. This is likely to play a role in the context of RNA interference [7] and viral RNA capsid compaction [8], as well as in modulating the specific interaction of the RNA duplex with proteins such as the RNA helicases [9], polymerases [10] and nucleases [11].

The development of single-molecule manipulation techniques has spurred a good number of exciting studies on the mechanical response of nucleic acids to tension and torque [12, 13]. They have revealed elastic properties otherwise hidden in bulk assays [14], have shown how stretching supercoiled DNA may activate homologous pairing [15], and have assessed the force-dependence of RNA folding [16]. They

have also demonstrated a unique ability to generate novel macromolecular forms. Outstanding examples are the studies of Cluzel et al. [17] and Smith et al. [18] on stretched DNA (S-DNA), a form 70% longer than B-DNA. The transition was subsequently modeled, with varied abilities to reproduce experimental observables, by computer simulations [19, 20, 21, 22, 23]. Additionally, an S-RNA form has also been measured and compared to S-DNA [24].

The present study focuses on another novel form of nucleic acid duplexes, recently revealed in single-molecule experiments that twisted and stretched double-stranded DNA with magnetic or optical beads attached to the ends [25, 26, 14]. In the over-twisting case, these experiments produced a structure that was hypothesized to be akin to, (and thereby to somewhat vindicate), an early model of DNA proposed by Pauling [27] (P-DNA). Pauling had modeled three helical backbones inside and the bases flipped outside. Soon after its publication, the P-DNA structure appeared to be untenable in light of the Watson-Crick model [28] for DNA under physiological conditions. Subsequently, some evidence existed to assume the presence of double-stranded P-DNA, but only under very particular conditions in dry DNA [29] or ethanol solutions [30] (see also [31]).

While the twisting and stretching single-molecule manipulations were instrumental in renewing the interest in P-DNA models, such experiments can only report a limited number of "configurational" observables (e.g., an extension or a bend angle). It is therefore important to complement them by simulations that can reveal atomic details of any assumed structural transition. The experimental report of double-stranded P-DNA did include a detailed structural model for P-DNA, generated using molecular mechanics in helical coordinates [25]. It involved minimization, in the absence of water and counterions, of a helically symmetric, periodic duplex

with twist constraints (the number of degrees of freedom was significantly reduced by fixing bond lengths and many of the bond angles). However, the actual all-atom dynamics and thermodynamics of the transition have not been previously calculated.

Here we report a study of dynamics, structures, energies, entropies, longitudinal stretching forces (from here on interchangeably referred to as "forces," or "tensions") and transversal torques calculated from atomic molecular dynamics simulations. For DNA, the calculated extension, rise, and the underlying forces-torques that effect P-DNA transitions match well with experimental data. For RNA, we produce the first model for how a P-RNA structure might look, and calculate the forces and torques that could produce it.

Logistically, this chapter is organized in two parts. The first one is largely a non-equilibrium study: we apply what we call *driving torques and forces* to observe conformational transitions on the nanosecond time scale. The second part deals with equilibrium; using thermodynamic averaging, we compute the *equilibrium torques and forces* needed to maintain the P forms. In the first part, we start with an analysis of overtwisting of the B-DNA state using a large driving torque and three driving forces of different magnitudes, and follow that with the undertwisting case under the same forces. We then change the starting structure to A-RNA, and apply similar driving force-torque combinations. We showcase the dynamics of how torque of either sign and stretching forces of various magnitudes applied on the helix create a variety of forms such as P, supercoiled P (scP), left-handed P, and denatured. In the second part, the Equilibrium Calculations section, we then shift towards our central result, the free energy profile and the underlying equilibrium forces and torques along the B- to P-DNA and A- to P-RNA conformational transitions, and map a reversible transformation on a phase diagram. A qualitative dissection of the enthalpy and

entropy changes from B to P is also attempted, and detailed structural parameters for P-DNA and the hypothetical P-RNA are presented. We end with concluding discussions.

2.2 Methods

Molecular dynamics simulations were performed with the program CHARMM [32], using version 27 of the nucleic acid force field parameters [33, 34]. A canonical, double-helical B-DNA of the Drew-Dickerson [35] dodecamer, d(CGCGAATTCGCG)$_2$ and a canonical, double helical A-form RNA of sequence (CGCGAAUUCGCG)$_2$ were generated. Their initial structures were aligned so that their primary axis were the x-axis. They were overlaid with a previously equilibrated water box, containing TIP3P water molecules [36] and sodium ions, with dimensions 100 Å $\times$ 36 Å $\times$ 36 Å in the case of B-DNA and 100 Å $\times$ 40 Å $\times$ 40 Å for A-RNA. Any solvent molecules within 1.6 Å of nucleic acid heavy atoms were deleted and the appropriate number of sodium atoms farthest from the nucleic acid were also deleted to create an electrically neutral solution. Periodic boundary conditions were used, electrostatic interactions were calculated with the particle-mesh Ewald method [37], and Lennard-Jones interactions were truncated at 14 Å with a switch-smoothing function from 12 Å to 14 Å. For equilibration purposes the system underwent 500 steps of steepest-descent minimization, followed by 2000 steps of adopted basis Newton-Raphson minimization. The system was rapidly heated to 300 K over 0.6 ps, followed by an equilibration period of 1 ns with weak harmonic restraints applied to the C3' carbon atoms at the 5' and 3' ends of each DNA strand, to prevent the helical axis from becoming unaligned with the x-axis. The leapfrog Verlet algorithm was used with Nosé-Hoover dynamics [38, 39] to keep the temperature constant throughout the simulations. During equi-

libration, a 2 fs timestep was used with SHAKE [40] to constrain all covalent bonds involving hydrogens, while during all other simulations the timestep was reduced to 1 fs and SHAKE was not used.

To keep the nucleic acid aligned along the x-axis throughout the simulation, the center of mass of the C3' atoms of the terminal bases were restrained within a cylinder of radius 0.5 Å aligned along the x-axis using a separate, flat-bottom, geometrical mean field harmonic potentials for each end with force constant of 10 kcal/mol/Å^2. To simulate the experimental set-up, a pulling force was applied in the x direction to the C3' atoms of the two bases at one end of the duplex, while the C3' atoms of the bases at the opposite end were harmonically restrained in the x-direction, but were otherwise free to move in the y-z plane. A torsional force was coded in the CHARMM source (see Appendix A). Its implementation followed directly from the definition of torque, $\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}$, in that, at each step, a transversal force was calculated that had the magnitude of the desired torque divided by the distance from the x-axis, and the direction was perpendicular to the vector distance from the x-axis to the atom. This torque was applied to the C3' atoms of the terminal bases of each strand such that each end was torqued in an opposite direction. Each individual torque had a magnitude of 300 pN·nm, resulting in a driving torque of 600 pN·nm. We defined the driving torque as positive if it acts in the direction that would increase the twist of the helix, while a negative driving torque lies in the direction that would decreases it. For each nucleic acid we presented the results of six independent simulations, three with overtwisting of the DNA helix with a positive driving torque and pulling forces of 10 pN, 100 pN, and 1000 pN, and three with a negative driving torque and pulling forces of 10 pN, 100 pN, and 1000 pN. We analyzed the backbone spacing of the DNA, defined as the distance from a phosphorus atom to the closest backbone atom

on the opposite strand, averaged over all bases, and the relative extension as the ratio of the instantaneous length to the initial one. We also examined the puckering phase of the sugar groups, with the definitions of puckering types based on the phase angle [41]. Entropy calculations were performed with quasiharmonic analysis [42] using the last 250 ps of the simulation to determine the effective frequencies. The enthalpy change of various energetic contributions were obtained by averaging over the last 250 ps of the trajectories in the P states and subtracting values obtained from averaging over the last 500 ps of the B (or A) trajectories. Helical parameters for twist-stretch coupling were calculated with CURVES [43] and the program VMD [44] was used for the creation of movies and figure images.

The RMSD to the final (reference) structure, $\rho = (\sum_{i=1}^{N}(\mathbf{r}_i - \mathbf{r}_i^{\mathrm{ref}})^2/N)^{1/2}$, with i indexing the backbone atoms (P, O5', C5', C4', C3' and O3', for a total of 142 atoms) was used as a transition coordinate for the conformational change. For both DNA and RNA, the reference structures were created by the simulations for which the driving torque overtwisted the respective duplex and 1000 pN of tension were applied. The equilibrium forces and torques involved in the transitions were computed from the mean force $\langle \mathbf{f}(\rho) \rangle$ on the system, where brackets denote canonical-ensemble averaging. The mean force was in turn derived from the potential of mean force (PMF), $W(\rho) \equiv -k_B T \ln \int \exp(-V(\mathbf{r})\delta(\rho(\mathbf{r}) - \rho))d\mathbf{r}$. Umbrella sampling [45] in combination with the weighted-histogram analysis method [46] as implemented by Grossfield [47] was used to calculate the potential of mean force W for the transition of B-DNA to P-DNA and A-RNA to P-RNA. Each window began with a snapshot from the overtwisting high/tension driving simulation at a corresponding ρ value in increments of 0.1 Å from their initial value (11.5 in DNA, 13.5 in RNA) to 0 (the reference state). Each window was run at two restraining potentials, one at 50

kcal/(mol·Å) and the other at 150 kcal/(mol·Å). For DNA these were run for 325 ps and for RNA for 305 ps. In the 150 kcal/mol·Å simulations the first 25 ps was used as an initial equilibrium period and for the 50 kcal/(mol·Å) the first 50 ps was used as an equilibration period. When combined, the windows thereby sampled 66.7 ns in the B-DNA case and 72.8 ns in the A-RNA case. The mean force was calculated by taking the numerical derivative of the free energy with a step size of 0.05 Å. To determine the tension on a single backbone atom i the mean force along the x direction was computed by taking the derivative of $W(\rho)$ with respect to x_i direction:

$$\langle f_{x_i} \rangle \equiv -\frac{dW}{dx_i} = -\frac{dW}{d\rho}\frac{d\rho}{dx_i} = -\frac{dW}{d\rho}\frac{(x_i - x_{ref})}{\rho \cdot N} \tag{2.1}$$

Tension in the duplex was then computed by adding the two mean forces that acted longitudinally in each strand. The mean force $\langle f_{x_i} \rangle$ was calculated for each heavy backbone atom i of non-terminal base pairs at 1 ps increments throughout the sampling periods for the PMF, and all tension values within a 0.125 Å window of the PMF were averaged, followed by averaging over all i atoms, to give the equilibrium tension on the nucleic acid. Similarly, we calculated the equilibrium torque by finding the backbone forces in the y and z directions and adding, for the backbone atoms of each non-terminal base pair (i.e., the kinematic unit involved in nucleobase rotations [48]), the cross products with the radii vectors of the helix:

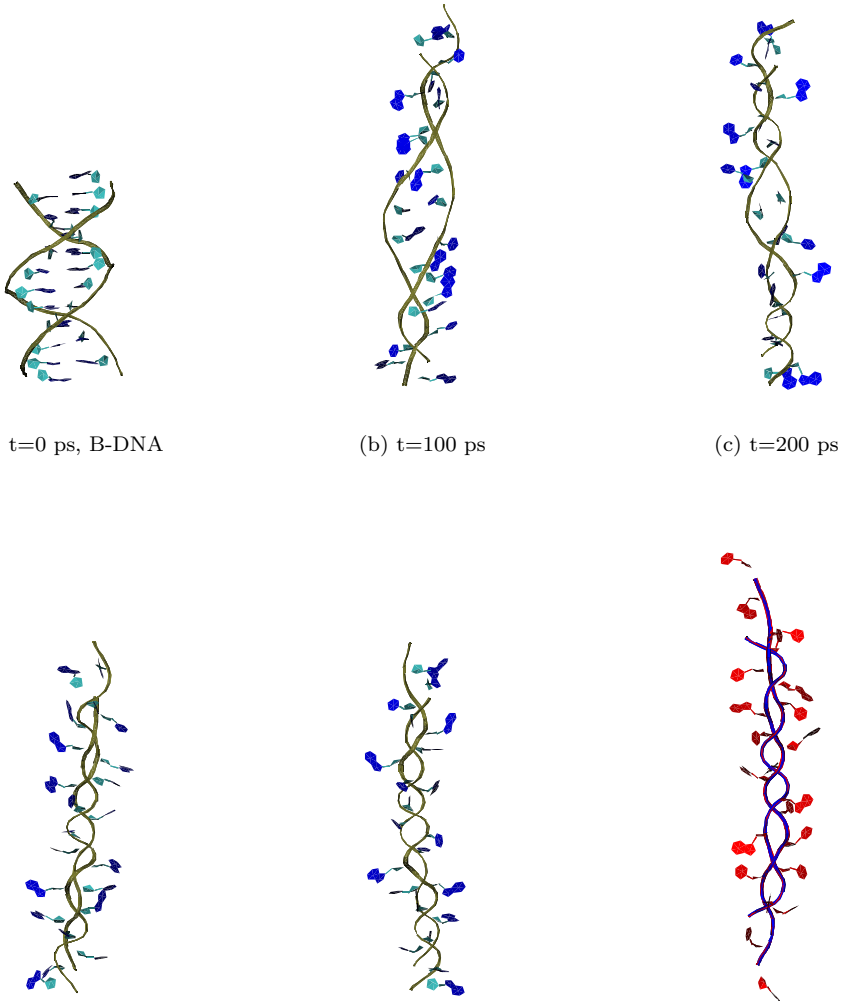
$$\begin{aligned} \langle \tau_{x_i} \rangle &= -\mathbf{r} \times (\frac{dW}{dy_i}\hat{j} + \frac{dW}{dz_i}\hat{k}) \\ &= -\frac{dW}{d\rho} \cdot \frac{(y\hat{j} + z\hat{k}) \times ((y_i - y_{\text{ref}})\hat{j} + (z_i - z_{\text{ref}})\hat{k})}{\rho \cdot N} \end{aligned} \tag{2.2}$$

2.3 Non-equilibrium structural transitions

2.3.1 Overtwisting B-DNA leads to P-DNA.

A large driving torque and tensions of three magnitudes were simulated at both strands at the end of a dodecamer duplex (as described above). In all three cases detailed below, the torque induced overtwisting of DNA and a subsequent transition to a structure with flipped-out bases. However, the magnitude of the pulling force had a significant impact on the details of the transition and the final shape of the DNA backbone.

(1) With 1000 pN of force applied over a nanosecond, the initial behavior of the backbone was somewhat similar to that observed in the B- to S-DNA transition [17, 18] in that the backbone is elongated and the bases begin to tilt. A notable exception is that the sugars are forced closer together, while the bases flip out of the helix almost concomitantly (Figure 2.1(b)). As the simulation proceeds the twist propagates through the helix and the bases are exposed to solution, but not flipped all the way out of the helix (Figure 2.1(c)). By 500 ps, the backbone regularly winds around itself and the bases are all the way exposed to solution. Between 500 ps and 1 ns there is no discernible change as the structure settles in a P-form helix. Figure 2.4(a) shows the relative extension of the DNA as a function of time. It confirms that the helix is longest at 100 ps with a relative extension of 1.8, before shortening as it winds around itself to an extension of 1.6. The twist of the DNA increases from its initial value to a value fluctuating around 148° between adjacent base pairs, yielding a twist value of 2.25 base pairs/turn. Examination of the backbone spacing in Figure 2.4(b) also shows that most of the transition is completed within the first 400 ps. The final spacing between strands is approximately 3.2 Å, which brings the negative charges on the two backbones in much closer proximity than when in the



(a) t=0 ps, B-DNA (b) t=100 ps (c) t=200 ps

(d) t=500 ps (e) t=1000 ps, P-DNA (f) P-RNA

Figure 2.1: (a.)-(e.): Snapshots from temporal evolution of B-P transition induced by overtwisting B-DNA with positive driving torque and 1000 pN tension. For comparison, we present a model of P-RNA in frame (f.), produced by a similar driving simulation on A-RNA. Backbone for DNA is in gold with blue bases and RNA has a blue backbone with red bases.

B-DNA form. An atomic view of the backbone structure (see Figure 2.2) shows that sugar rings become approximately parallel to the helical axis, and that the phosphate anionic oxygens radiate outward to minimize their electrostatic repulsion. Simulation also reveals a longitudinal shift of one of the strands relative to the

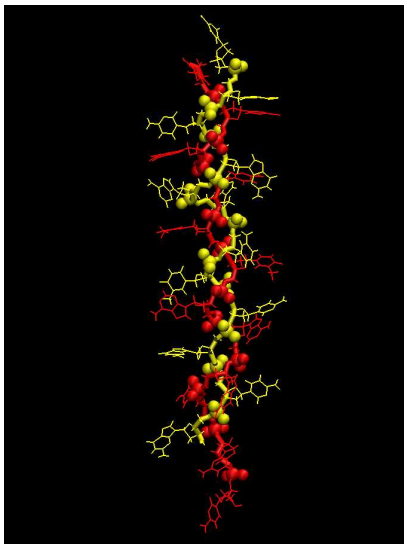


Figure 2.2: Atomic view of P-DNA backbone (in bond representation). The phosphate anionic oxygens (smaller spheres) radiate outwards from the DNA axis to minimize their electrostatic interactions, while the phosphorus atoms (large spheres) on opposite strand (colored yellow or red) line up in a more staggered position than when in B-DNA. The ribose rings of both strands assume a perpendicular position relative to the helical axis. Phosphorus and anionic oxygen atoms are explicitly shown according to their van der Waals radius.

other by up to one nucleotide unit. While this staggered configuration might occur in the experiment (base complementarity in P-DNA is lost and staggering could lower phosphate repulsions), it could also be caused by edge effects in our finite-size system. Therefore, calculations of equilibrium properties excluded the terminal bases (see below), and we expect the shift to not alter the structure or energy of the non-terminal bases.

Given the values of our calculated extension (1.6) and twist (2.25 base pairs/turn), we believe that the simulation has created a P-DNA structure close to that formed in the single molecule experiments. Those reports indicate extension values of 1.6-1.75 and twists of 2.4-2.6 base pairs/turn, depending on the study [25, 26, 14].

(2) When only 10 pN of tension are applied to the molecule, the DNA transition to P form proceed at a slower rate than in the 1000 pN case. This can be seen in the backbone spacing, presented in Figure 2.4(b). The final structure has a

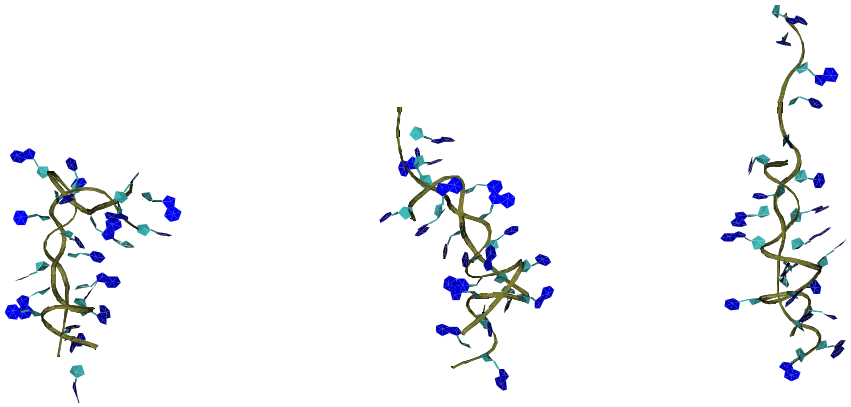
buckled backbone (Figure 2.3(a)) with an extension of only 0.76 of that of B-DNA (Figure 2.4(a)). Due to its writhe, this structure is reminiscent of the super-coiled P-DNA reported in high torque/low pulling force single molecule DNA experiments in [14].

(3) When 100 pN of tension are applied, the backbone still buckles, however not nearly to the extent seen in the 10 pN tension case, generating a middle section of super coiled P-DNA and a section that consists of single strands extended into solution (Figure 2.3(b)). When only the middle six bases are included in the analysis, the backbone spacing is 3.2 Å and the extension is 1.03, consistent with slightly super-coiled P-DNA. While the edge fraying observed is a finite-size artifact of simulation, the structural data obtained from the middle base pairs provide reasonable putative models for the experimentally generated supercoiled structures.

2.3.2 Undertwisting B-DNA leads to alternative structures.

We also simulated a *negative* driving torque (-600 pN·nm) on the dodecamer, with the three driving forces.

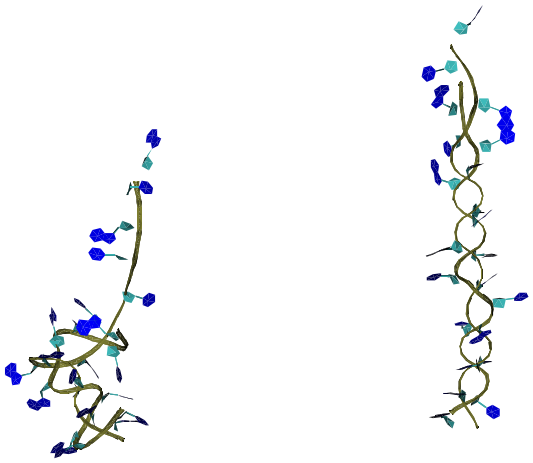
(1) Experiments have indicated [25] that undertwisting DNA leads to its denaturation. In a simulation with the negative driving torque and 10 pN of tension applied on B-DNA, this was observed. As the simulation proceeds, the torque first untwists the helix as the Watson-Crick hydrogen bonds become strained and begin to break. By 200 ps, the strands begin to translate relative to one another, and at 500 ps one strand is extended while the other coils around it with no obvious structure. The average backbone separation (Figure 2.4(d)) is approximately 4.6 Å, significantly larger than that for P-DNA, but not close to other types of organized DNA. This simulation indicates that undertwisting with a low tension leads to denatured DNA that could involve an intermediate with one strand twisted around the other in a



(a) Overtwisting with low tension

(b) Overtwisting with medium tension

(c) Underwisting with low tension



(d) Undertwisting with medium tension

(e) Undertwisting with high tension

Figure 2.3: Final DNA structures created in simulations with a positive or negative driving torque and various tensions.

disorganized fashion (Figure 2.3(c)).

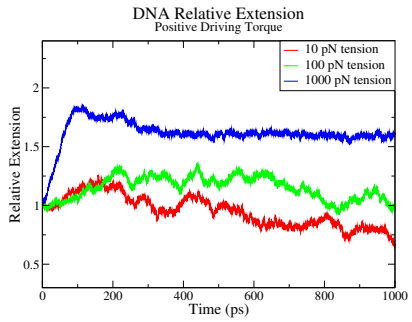
(2) In contrast to the low-tension case, undertwisting while applying a tension of 1000 pN produces a highly regular structure similar to P-DNA, but with a left-handed helical orientation (shown in Figure 2.3(e)). In the transition to this structure the relative extension (Figure 2.4(c)) and the backbone spacing (Figure 2.4(d)) follow

paths that are similar to those values obtained for by overtwisting with high tension. DNA passes through a state of zero twist before it re-twists into a left handed conformation with a twist of 158° and 2.28 pairs/turn. The undertwisting experiments, performed with low pulling forces, report denatured states and find no evidence for left-handed P states [25, 14]. It is formally possible that the left-handed P forms we have generated in the simulations are transient states, and that a longer sampling time would lead to denaturation. It is also possible, however, that in particular pulling geometries that apply (as done here) tension on *both* strands, these structures could be generated. We note that the simplified modeling of P-DNA in the original single-molecule report [25] had also proposed the existence of a left-handed P that was close to a mirror-like image of right-handed P-DNA.

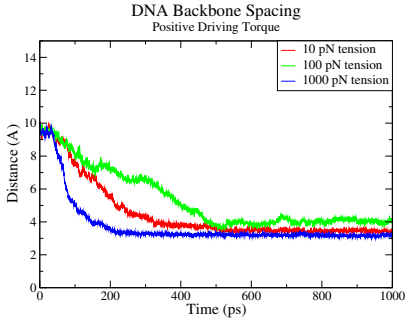
(3) When the medium pulling force (100 pN) is applied, the structure of DNA becomes a combination of denatured DNA and left-handed P-DNA (possibly supercoiled). The super-coiling reduces the relative extension (Figure 2.4(c)) to 0.8 while the average backbone separation (Figure 2.4(d)) is at approximately the same level as in the 10 pN tension case, this is a result of the bases in the denatured region; the backbone in the left handed P-DNA region have a separation of approximately 3.4 Å (Figure 2.3(d)).

2.3.3 Model for a novel RNA form: P-RNA.

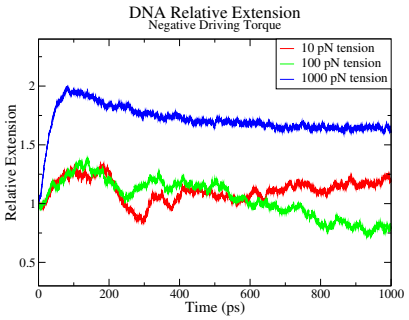
We have performed simulations starting from an A-RNA structure using the same driving forces and torques employed for DNA, i.e., 600 pN·nm positive and negative torque, with 10, 100 and 1000 pN pulling forces (see Figure 2.5). As was the case for DNA, overtwisting RNA with a large driving force produced a P-form structure that we refer to as P-RNA (see Figure 2.1(f)). While similar in overall shape with P-DNA, it differed in absolute extension, which was greater than P-DNA by about



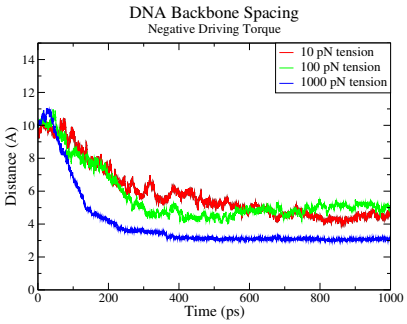
(a) Relative extension of DNA under a driving torque as a function of time



(b) Backbone spacing of DNA under a driving torque as a function of time



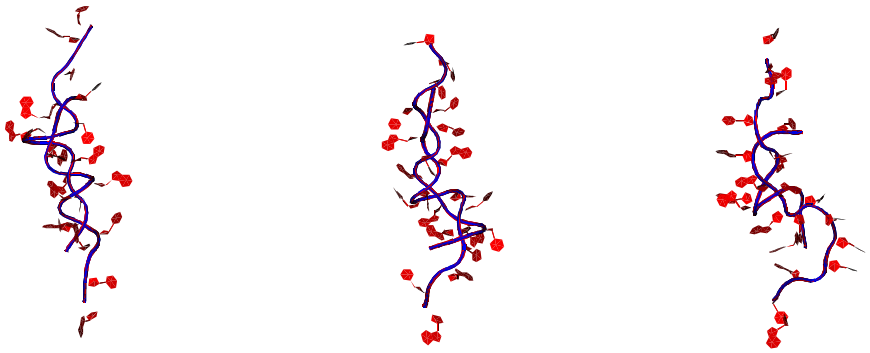
(c) Relative extension of DNA under a driving torque as a function of time



(d) Backbone spacing of DNA under a driving torque as a function of time

Figure 2.4: Structural properties of B-to-P DNA transitions.

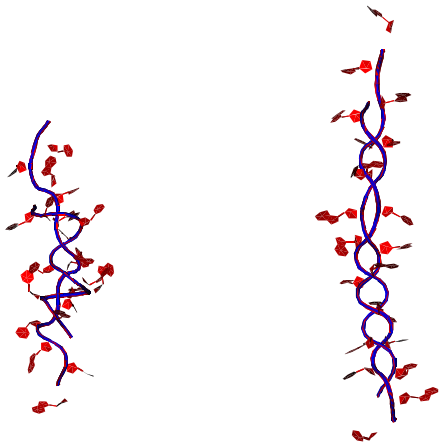
10% (see Figure 2.6). Another distinction from the P-DNA simulations was that the backbone spacing, and consequently the atomic structure on the whole, approached the final P-RNA conformation on a slower time scale than in the DNA cases (i.e., on the order of 300 ps for most of the bases and 900 ps for all of them as opposed to 150-300 ps for all DNA bases) for the collapse of the backbone spacing below 4 Å. This observation is in accord with results from umbrella sampling, (see next sections and Figure 2.8) which indicate a slightly steeper energy profile along the transition to P-RNA. Simulations with other torques and tensions created structures similar to those in the DNA case (left handed P-form, right handed supercoiled P, and denatured states) but also on a slower time scale when compared to DNA.



(a) Overtwisting with low tension

(b) Overtwisting with medium tension

(c) Underwisting with low tension



(d) Undertwisting with medium tension

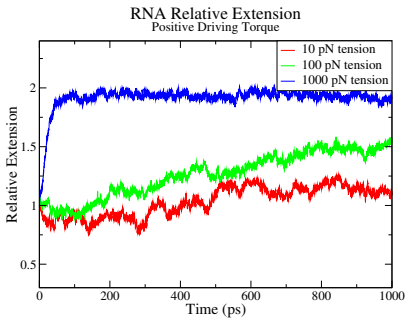
(e) Undertwisting with high tension

Figure 2.5: Final RNA structures created in simulations with a positive or negative driving torque and various tensions.

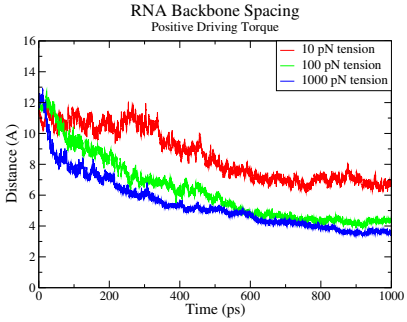
2.4 Equilibrium Calculations

2.4.1 Free Energy and Equilibrium Forces and Torques along a B- to P-DNA Transition Isotherm.

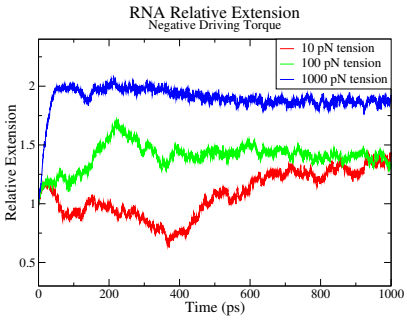
The previous section presented *driving* simulations, i.e., simulations that induced conformational transitions to P forms on a rapid (irreversible) time scale, and that,



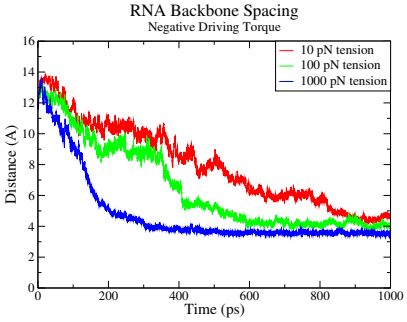
(a) Relative extension of RNA under a positive driving torque as a function of time



(b) Backbone spacing of RNA under a positive driving torque as a function of time



(c) Relative extension of RNA under a negative torque as a function of time



(d) Backbone spacing of RNA under a negative driving torque as a function of time

Figure 2.6: Structural properties of B-to-P RNA transitions.

therefore, produced trajectories amenable only to qualitative assessment. In the present section, we use the end-point structures from those simulations to rigorously perform an equilibrium, reversible transition. The underlying free energies are presented in quantitative terms by performing an extensive calculation of the potential of mean force (PMF) along the conformational transitions to the P forms. The transition coordinate, ρ, is chosen as the root-mean squared displacement (RMSD) of backbone atoms relative to a final P structure (see Methods for details). Results obtained using umbrella sampling in combination with the weighted-histogram analysis method are presented in Figure 2.8(a). Both B-DNA and A-RNA exhibit a large increase in free energy as structures approach the P forms.

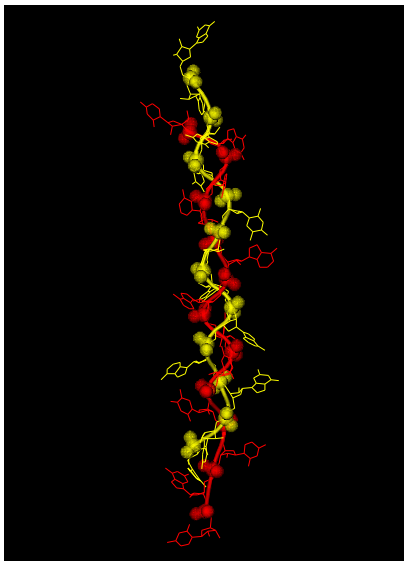
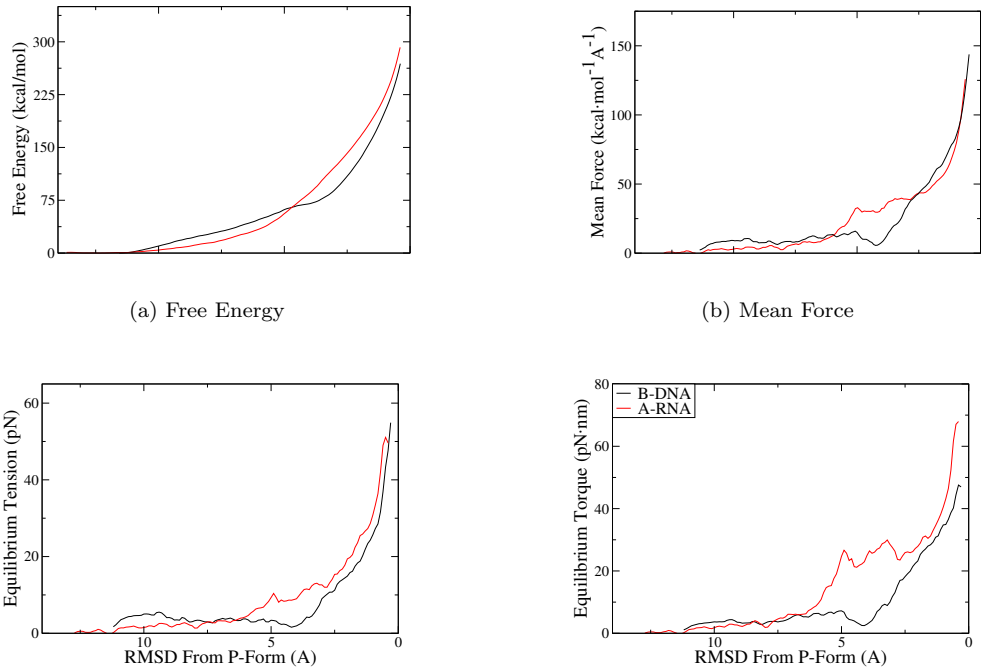


Figure 2.7: Atomic view of the P-RNA backbone. Same representation as in Figure 2.2

For DNA, it is observed that the initial deformation from the B state to a state with $\rho = 5$ Å away from the P reference requires a relatively weak, linear increase in free energy. Structurally, for DNA conformations with a value of ρ greater than 5 Å, the primary deformation is a lengthening of the axis, while the (non-terminal) bases remain at the center of the helix. As ρ decreases towards P values, the bases begin to flip outwards and the backbone twists around itself, causing a large increase in the electrostatic energies of the negatively charged backbones. This sets in over the interval $\rho = 4.5 \rightarrow 0.8$ Å, and is characterized by a stronger, quadratic increase in the free energy. While a gradual, rather than a sharp transition is observed over this interval (as expected due the finite, small size of the simulated dodecamer), visual inspection reveals that a P-DNA state is fully formed at a value of $\rho \approx 1$ Å.

The mean forces obtained from the gradient of the PMF (in Figure 2.8(b), see Methods for details) are used to calculate the *equilibrium* tensions and torques (Figs. 2.8(c) and 2.8(d), respectively). I.e., for each value of ρ, we obtained the thermodynamically-averaged force $\mathbf{f}(\rho)$ and torque $\boldsymbol{\tau}(\rho)$ (where the average was over



(a) Free Energy

(b) Mean Force

(c) Calculated Equilibrium Tension

(d) Calculated Equilibrium Torque

Figure 2.8: (a.) Free energy profiles of B- to P-DNA and A- to P-RNA transitions. (b.) Mean forces calculated from the derivative of the free energies. (c.) Calculated equilibrium tensions for formation of P-DNA and P-RNA. (d.) Calculated equilibrium torques for formation of P-DNA and P-RNA.

a constant-temperature ensemble of configurations) that would be needed to maintain the structure in equilibrium at that value of ρ. The reversible transition pathway such produced by umbrella sampling along ρ (in effect, a reversible isotherm in $\mathbf{f}$-$\boldsymbol{\tau}$ coordinates with ρ as an order parameter) allowed us to map the calculated force-torque pairs onto an experimentally-derived [14] phase diagram of force-torque dependent DNA states [26, 49] (see Figure 2.9). Because the pathway is a result of umbrella sampling calculation that yields the lowest force needed to maintain in equilibrium a P-like structure, the isotherm passes almost through the "triple point" between B, P and scP. This indicates that the lower bound for tensions required to induce P-forms is about 25 pN, which validates, on one hand, the points on the phase diagram measured experimentally, and, on the other hand, the overall fea-

tures of the borderlines between B, P and scP derived theoretically (cf. Figure 4a in reference [49]). Moving along the ρ pathway, structures created around ρ =1.0 Å reveal the transition of DNA into P form. For them, we calculate an equilibrium tension of 25.7±9.7 pN and an equilibrium torque of 34.8±3.2 pN·nm (with the error estimates being the standard deviation of all structures within 0.25 Å of ρ =1.0). This compares quite well with the torque-force triple-point experimental values of 25 pN and 34 pN·nm, respectively. In Figure 2.9 also notable is the fact that our predicted $\mathbf{f}$-$\boldsymbol{\tau}$ DNA state at ρ =1.0 Å lies on the border between B- and P-DNA phases exactly where mapped experimentally in reference [14] (see Figure 2.9), while structures with a lower ρ are deeper inside the P region. It is worth noting that in the region near the initial B-DNA structure (ρ >9.5 Å), for which DNA is still in the entropic regime (F≤10 pN), we were also able to calculate (see Section 2.4.5 for details) a twist/stretch coupling constant of -15 nm. It agrees, both in sign and absolute value, with recent single molecule experiments indicating that, near the B form of DNA, an increase in twist leads to an increase in extension [50, 51].

2.4.2 The A- to P-RNA transition: Free Energy, Forces, Torques and a Hypothetical Triple Point

As for DNA, the P-RNA structure generated by the driving simulation was used as reference in defining the RMSD transition coordinate, ρ, and umbrella sampling of ρ was performed to calculate free energies and equilibrium forces and torques. The free energy profile of the transition from A- to P-RNA is initially lower than for the DNA case, but then begins to increase more rapidly at approximately ρ =6.5 Å, which we attribute to the additional inter-strand repulsion between the 2' oxygen atoms. As RNA extends into P form, this effect becomes less significant but it does cause P-RNA to have a larger free energy than P-DNA. Throughout the second half of the

transition, the equilibrium torques and tensions are higher than in the DNA case and we calculate that, at the "borderline" ρ =1.0 Å, P-RNA has an equilibrium tension and torque of 30.8±12.8 pN and 40.5±6.4 pN·nm. By analogy with the DNA case, we predict that in an RNA phase diagram (which has not been experimentally mapped yet) this would correspond to a "triple point" of B, P, and scP RNA with higher forces and torques (see inset to Figure 2.9). While it is formally possible that this is valid only for the particular sequence we studied, the "up-shift" of the force-torque values for the triple point we predict for RNA is consistent with measurements on a variety of stretched RNA sequences [24], which report an increase in the value of the force required to effect the A- to S-RNA transition by about 10 pN relative to DNA. To be additionally noted are two observations. Firstly, the larger extension for RNA that we compute (see Table 2.2) is also in accord with the larger RNA stretching factor (1.0 relative to 0.7 cf. reference [24]) measured in those experiments. Secondly, when compared to P-DNA, P-RNA has a larger variance in our calculated forces and torques, similarly to the experimental observation [24] of a larger variance in the plateau force for S-RNA.

2.4.3 Qualitative Decomposition of Energy and Entropy Changes.

The DNA internal energy contributions to the enthalpy change (see Methods for calculation details) for various structural transitions are presented in Table 2.1. Large positive change contributions arise from the van der Waals (Int_{VDW}) and electrostatic internal (Int_E) energy terms, which are due to the close proximity of the backbones. The other internal energy terms, Int_{oth}, also contributed significantly to the energy increase, as bonds and angles in DNA were rotated and stretched away from their B-form equilibrium values.

Our conformational entropy calculations (see Methods for details) have revealed

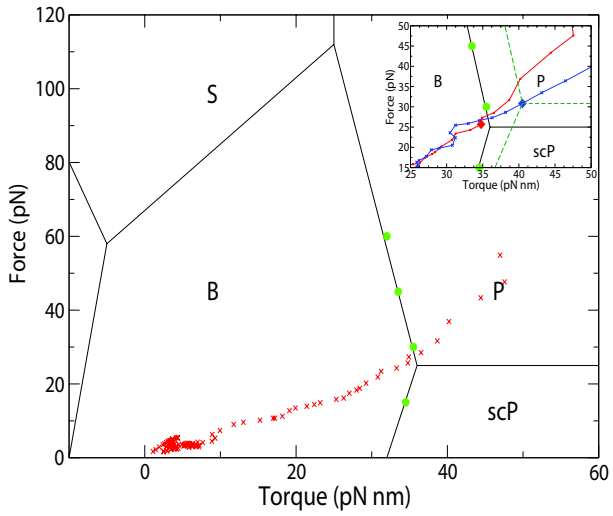


Figure 2.9: The global force-torque phase diagram of DNA overlayed with points in force-torque space from our B to P, reversible, umbrella sampling pathway (or "isotherm"). Points in green have been previosuly measured experimentally [14]. Note that, because the calculated equilibrium pathway yields the lowest force needed to maintain in equilibrium a P-like structure, the isotherm passes almost through the lower right "triple point" between B, P and scP. This indicates that tensions lower than about 25 pN cannot induce P-forms and validates the experimentally-derived phase diagram. The inset shows a zoomed in region of the phase diagram with the value at $\rho = 1$, an approximate triple point, emphasized with a diamond, and a sketch of what a phase digram with a triple point at this value may look like with green dashes. For comparison, the DNA isotherm is in red, while the RNA isotherm is in blue.

an interesting compensation effect. The entropy change from canonical to twisted forms is serving as a significant counterbalance to the large enthalpy change estimated above. All simulations showed an entropy increase in the range of 0.19-0.72 kcal/(K·mol) (corresponding to a free energy decrease ranging from -76.0 kcal/mol to -258 kcal/mol, Figure 2.10) . This relatively significant increase in entropy is ascribed to the flipped bases being free to move due to their lack of stacking, and is in accord with studies indicating that torsional rigidity of DNA correlates with the stacking energies and not with the melting temperatures [52].

It is important to re-emphasize that the numbers in Table 2.1 are highly approximate. The sampling time of the driving simulations is not sufficient for accurate

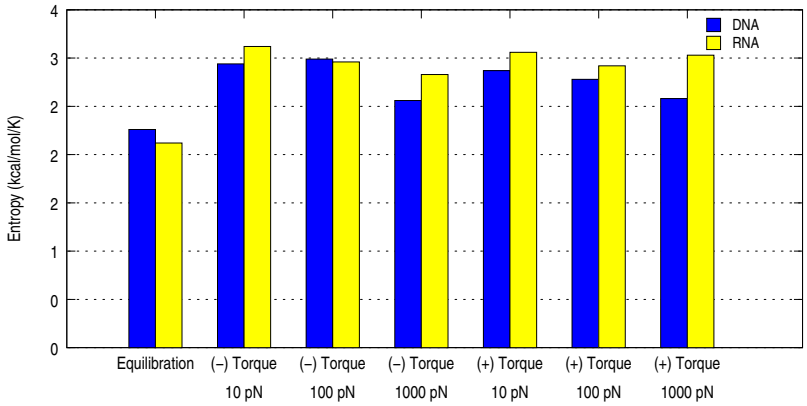


Figure 2.10: Conformational entropy from quasiharmonic analysis of the ten non-terminal base pairs for the various DNA and RNA structures discussed in the main text. A temperature of 300 K and a scaling factor of 12/10 (accounting for the quasi-extensivity of entropy) were used to calculate the $-T\Delta S$ values reported in Table 1 in the text.

f_{driving} =	10 pN (-) Torque	100 pN (-) Torque	1000 pN (-) Torque	10 pN (+) Torque	100 pN (+) Torque	1000 pN (+) Torque
ΔInt_{VDW}	112	120	109	105	96	140
ΔInt_{E}	554	515	628	681	551	586
ΔInt_{oth}	396	521	489	490	415	475
$-T\Delta S_{DNA}$	-258	-223	-68.4	-176	-170	-76.0

Table 2.1: DNA internal energy and entropy changes (in kcal/mol) for structural transition from B-DNA upon simulations with driving torques and forces.

convergence. As described above, the quantitative description was brought in by the additional umbrella sampling simulations for the calculation of free energy profiles. While admittedly the approximate nature of the decomposition here allows only a qualitative picture, an unequivocally large increase of the enthalpy for the system is expected to exist. In compensation, the increase in entropy helps to offset the enthalpic cost of the new structures created, but the additional, decisive compensation for the (still) high cost of creation of the structures is provided by the external twisting and pulling forces.

2.4.4 Structural comparison of P-DNA and P-RNA.

Table 2.2 presents a comparison of forces and structural parameters for DNA and RNA from the configurations gathered during umbrella sampling. This equilibrium

Structure	f[pN]	τ[pN·nm]	l/l_0	twist [bp/turn]
$B-DNA$	-	-	1.0	10.5
$P-DNA$	25.74±9.7 (30)	34.8±3.2 (34)	1.57(1.6-1.75)	2.35 (2.4-2.6)
$A-RNA$	-	-	0.86	10.7
$P-RNA$	30.8±12.8	40.5±6.4	1.71	3.00

Table 2.2: Equilibrium forces (f), torques (τ), relative extensions (l/l_0), and twist for P-form nucleic acids calculated from umbrella sampling simulations at $\rho = 1$ (see text for details). The numbers in parentheses are corresponding experimental data, where available, from references [25, 26, 14]

Structure	α	β	γ	δ	ϵ	ζ
$B-DNA$	-60°	165°	53°	132°	-163°	-112°
$P-DNA$	-55°	176°	-156°	129°	-168°	160°
$A-RNA$	-88°	172°	64°	78°	-156°	-70°
$P-RNA$	175°	180°	116°	138°	-168°	160°

Table 2.3: Equilibrium backbone torsion angles ($\alpha - \zeta$) for P-form nucleic acids calculated from umbrella sampling simulations at ρ =1 (see text for details).

sampling has allowed us to create a more accurate calculation of extension and twist for P-DNA and P-RNA. We calculate that P-DNA has an average rise of 5.3 Å and an extension of 1.57 relative to the B-DNA initial structure. For P-RNA there is a slightly higher rise of 5.8 Å with an extension of 1.94 relative to the A-RNA initial structure (and 1.71 relative to the initial B-DNA structure). The additional electrostatic repulsion from the 2' hydroxyl oxygen with the backbone could be forcing the backbone into a straighter conformation, creating the larger rise for P-RNA. The backbone torsion angles α to ζ were calculated for structures with ρ <0.4 Å for P forms, ρ =11.4-11.6 Å for B-DNA, and ρ =13-13.2 Å for A-RNA (see Figure 2.3). Angles α, β, δ, and ϵ did not vary significantly between the initial and final conformations. Rotations about the glycosidic bond χ appeared to be unaffected by the P-form backbone and were distributed according to the energy of steric hindrance between base and sugar. We also examined the puckering phase of the sugars in both P-DNA and P-RNA. For both overtwisting and undertwisting, we found that backbone P conformations did not influence the distribution of puckering at various edges in the sugar rings, which were again determined solely by the energy difference

between north and south puckers [53, 54]. The average counterion-phosphate distance does not change appreciably throughout the transition; however because the backbone is more condensed, the *local* sodium concentration does increase from its initial value.

2.4.5 Calculation of B-DNA Twist/Stretch Coupling Constant.

The energy of DNA per unit length, in the linear response limit, can be approximated as

$$E/L_0 = \frac{4\pi^2 k_B T}{p^2}(C\sigma^2 + B\epsilon^2 + 2D\epsilon\sigma), \tag{2.3}$$

where $\epsilon = L/L_0 - 1$ is the relative change in extension L, $\sigma = Tw/Tw_0 - 1$ is the relative change in twist Tw, p =3.6 nm is the DNA pitch, and C =100 nm, B =78 nm represent the torsional, and, respectively, stretch moduli for DNA (with values from reference [50]). D is the twist-stretch coupling constant we seek to compute from our simulations.

The average DNA structure created in each umbrella sampling window that had an RMS >9.5 Å (i.e., within a linear-response, small twist-angle region of up to approximately 2 Å away from B-DNA) was analyzed with CURVES 5.0 [43] to determine the extension and twist of the central 6 base pairs. The coupling constant D was determined by fitting the numerical extension-twist dependence (see Figure 2.11) to the equation:

$$\epsilon - \epsilon_{\sigma=0} = -D\sigma/B, \tag{2.4}$$

where L_0 and Tw_0 are the equilibrium extension and twist, determined from the ρ =11.5 Å window (the canonical B-DNA form, with values 3.43 nm and 35.18°, respectively). The value of $\epsilon_{\sigma=0}$ was derived from a linear fit of ϵ vs. σ values

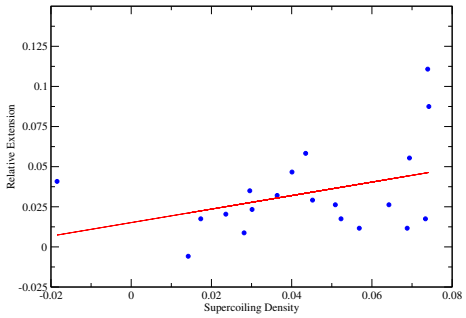


Figure 2.11: Twist-stretch dependence from umbrella sampling simulations (blue points) and the linear fit to Eq. (2.4) (red line)

(giving $\epsilon_{\sigma=0} = 0.0152$). Averaging of all ϵ values determined by this method gives D =-15.4±53.5 nm.

Although the variance is large (because of the small, finite size of the system, relative fluctuations do not decay sufficiently) the negative character of twist-stretch coupling is indisputable (see Figure 2.11). The average coupling constant D =-15.4 nm compares favorably with the result of Lionnet *et al.*, who report D =-9.1±4 nm, and with that of Gore *et al.* [51], who measure D =-11.1±2.5 nm. (The value for the coupling constant g = -90pN·nm reported by Gore *et al.* was transformed to D, i.e., in the units used by Lionnet *et al.*, by employing the formula $g/S = Dp/(2\pi B)$ with S =1100 pN the stretch modulus used in Gore *et al.*)

It is also worth noting that our calculated value is close to those reported in the modeling section of the Lionnet *et al.* reference, who computed, using a different force field (see reference [55] therein) and helical-symmetry energy minimization, values of D =-13 to D =-20 nm, depending on the applied tension.

2.5 Concluding Discussions

We have presented a series of all-atom simulations concerning the effect of torsion and tension on double-stranded nucleic acids and have shown that that transitions to a P form or denatured states are possible. These forms are energetically disfavored

under equilibrium conditions, but become favored when DNA is under the relatively high torque and tension that are applied by single-molecule manipulations or that can, in some instances, arise in nucleoprotein complexes.

We have presented the dynamics of how, in the high tension cases, ordered right-handed or left-handed P structures arise, and we have shown that, in the lower tension cases, the nucleic acid can either go toward a denatured state or a supercoiled P form.

While the driving torques and forces used in our simulation to induce the over-torqued states were one to two orders of magnitude greater than those used in experiments, we have employed them merely to generate the end structures (which agreed, in structural terms, i.e., extension and twist, with the experiments). In other words, we have used the irreversible trajectory generated with large torques/tensions simply as a "transportation" means to overcome the energetic cost to getting to P-DNA (and P-RNA). With the P structure as a target, we have then generated, using umbrella sampling, a reversible, equilibrium transformation pathway, and have calculated its free energy profile. From the free energy profile (a potential of mean force), we have derived the theoretically-exact, lower bound, equilibrium forces and torques, thereby showing that these structures may be created by the forces and torques reported in single-molecule experiments. The good agreement between the calculated and measured parameters (force-torque, extension, rise) for P-DNA, and the passage of our calculated $\mathbf{f}$-$\boldsymbol{\tau}$ isotherm through the triple point of the experimentally-derived phase diagram suggest strongly that the simulated structures correspond to those generated in the corresponding experiments, and lend credence to a model for the novel P-RNA structure we generated using similar conditions in the simulation of an A-RNA duplex.

The calculation of the free energy profile in the vicinity of B-DNA has additionally

provided an equilibrium twist-elongation dependence that enabled the calculation, in the low-twist limit, of a negative twist-stretch coupling constant of -15 nm, in accord with recent experiments [50, 51].

The fact that not only the driving torque, but also the driving tension (i.e., the forces applied along the helical axis) needed to be an order of magnitude larger in our nanosecond-time simulation to induce the microsecond-time (or longer) transitions to P forms reported by the experiments indicates that the conformational pathways to P are not perpendicular to the helical axis. Significant energy barriers are expected to exist in directions along the axis. This is not totally unexpected given twist-stretch coupling in DNA [56, 57, 50, 51], and is in accord with the fact that stretching longitudinally undertwisted DNA induces a flipping out of bases that can activate homologous pairing in physiological conditions [15].

This work has been published in The Proceedings Of The National Academy Of Sciences Of The United States of America [58].

BIBLIOGRAPHY

[1] A. Rich. The rise of single-molecule DNA biochemistry. *Proceedings of The National Academy of Sciences of The United States of America*, 95:13999–14000, 1998.

[2] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond. Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. *Nature*, 389:251–260, 1997.

[3] R. Schleif. DNA looping. *Annual Review of Biochemistry*, 61:199–223, 1992.

[4] C. D. Mao, W. Q. Sun, Z. Y. Shen, and N. C. Seeman. A nanomechanical device based on the B-Z transition of DNA. *Nature*, 397:144–146, 1999.

[5] A. Lim, E. T. Dimalanta, K. D. Potamousis, G. Yen, J. Apodoca, C. H. Tao, J. Y. Lin, R. Qi, J. Skiadas, A. Ramanathan, N. T. Perna, G. Plunkett, V. Burland, B. Mau, J. Hackett, F. R. Blattner, T. S. Anantharaman, B. Mishra, and D. C. Schwartz. Shotgun optical maps of the whole escherichia coli o157 : H7 genome. *Genome Research*, 11:1584–1593, 2001.

[6] K. M. Phillips, J. W. Larson, G. R. Yantz, C. M. D'Antoni, M. V. Gallo, K. A. Gillis, N. M. Goncalves, L. A. Neely, S. R. Gullans, and R. Gilmanshin. Application of single molecule technology to rapidly map long DNA and study the conformation of stretched DNA. *Nucleic Acids Research*, 33:5829–5837, 2005.

[7] A. Fire, S. Q. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello. Potent and specific genetic interference by double-stranded RNA in caenorhabditis elegans. *Nature*, 391:806–811, 1998.

[8] Lodish et al. *Molecular Cell Biology*. W.H. Freeman and Company, 2004.

[9] J. de la Cruz, D. Kressler, and P. Linder. Unwinding rna in saccharomyces cerevisiae: Dead-box proteins and related families. *Trends In Biochemical Sciences*, 24:192–198, 1999.

[10] S. J. Butcher, J. M. Grimes, E. V. Makeyev, D. H. Bamford, and D. L. Stuart. A mechanism for initiating RNA-dependent RNA polymerization. *Nature*, 410:235–240, 2001.

[11] E. Bernstein, A. A. Caudy, S. M. Hammond, and G. J. Hannon. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409:363–366, 2001.

[12] T. Strick, J. F. O. Allemand, V. Croquette, and D. Bensimon. The manipulation of single biomolecules. *Physics Today*, 54:46–51, 2001.

[13] C. Bustamante, Z. Bryant, and S. B. Smith. Ten years of tension: single-molecule DNA mechanics. *Nature*, 421:423–427, 2003.

[14] Z. Bryant, M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante. Structural transitions and elasticity from torque measurements on DNA. *Nature*, 424:338–341, 2003.

[15] T. R. Strick, V. Croquette, and D. Bensimon. Homologous pairing in stretched supercoiled dna. *Proceedings of The National Academy of Sciences of The United States of America*, 95:10579–10583, 1998.

[16] J. Liphardt, B. Onoa, S. Smith, I. Jr. Tinoco, and C. Bustamante. Reversible unfolding of single RNA molecules by mechanical force. *Science*, 292:733–737, 2001.

[17] P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J. L. Viovy, D. Chatenay, and F. Caron. DNA: An extensible molecule. *Science*, 271:792–794, 1996.

[18] S. B. Smith, Y. J. Cui, and C. Bustamante. Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded dna molecules. *Science*, 271:795–799, 1996.

[19] M. W. Konrad and J. I. Bolonick. Molecular dynamics simulation of DNA stretching is consistent with the tension observed for extension and strand separation and predicts a novel ladder structure. *Journal of The American Chemical Society*, 118:10989–10994, 1996.

[20] A. Lebrun and R. Lavery. Modelling extreme stretching of DNA. *Nucleic Acids Research*, 24:2260–2267, 1996.

[21] A. D. MacKerell and G. U. Lee. Structure, force, and energy of a double-stranded DNA oligonucleotide under tensile loads. *European Biophysics Journal With Biophysics Letters*, 28:415–426, 1999.

[22] R. Lavery and A. Lebrun. Modelling DNA stretching for physics and biology. *Genetica*, 106:75–84, 1999.

[23] K. M. Kosikov, A. A. Gorin, V. B. Zhurkin, and W. K. Olson. DNA stretching and compression: Large-scale simulations of double helical structures. *Journal of Molecular Biology*, 289:1301–1326, 1999.

[24] M. Bonin, R. Zhu, Y. Klaue, J. Oberstrass, E. Oesterschulze, and W. Nellen. Analysis of RNA flexibility by scanning force spectroscopy. *Nucleic Acids Research*, 30, 2002.

[25] J. F. Allemand, D. Bensimon, R. Lavery, and V. Croquette. Stretched and overwound DNA forms a Pauling-like structure with exposed bases. *Proceedings of The National Academy of Sciences of The United States of America*, 95:14152–14157, 1998.

[26] J. F. Leger, G. Romano, A. Sarkar, J. Robert, L. Bourdieu, D. Chatenay, and J. F. Marko. Structural transitions of a twisted and stretched DNA molecule. *Physical Review Letters*, 83:1066–1069, 1999.

[27] L. Pauling and R. B. Corey. A Proposed Structure for the Nucleic Acids. *Proceedings of the National Academy of Science*, 39:84–97, February 1953.

[28] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.

[29] M. Falk, K. A. Hartman, and R. C. Lord. Hydration of deoxyribonucleic acid .3. A spectroscopic study of effect of hydration on structure of deoxyribonucleic acid. *Journal of The American Chemical Society*, 85:391–&, 1963.

[30] J. T. Bokma, W. C. Johnson, and J. Blok. Cd of the li-salt of DNA in ethanol water mixtures - evidence for the B-form to C-form transition in solution. *Biopolymers*, 26:893–909, 1987.

[31] J. Piskur and A. Rupprecht. Aggregated DNA in ethanol solution. *Febs Letters*, 375:174–178, 1995.

[32] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM - a program for macromolecular energy, minimization, and dynamics calculations. *Journal Of Computational Chemistry*, 4:187–217, 1983.

[33] N. Foloppe and A. D. MacKerell. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal Of Computational Chemistry*, 21:86–104, 2000.

[34] A. D. MacKerell, N. Banavali, and N. Foloppe. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, 56:257–265, 2000.

[35] H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson. Structure of a B-DNA dodecamer - Conformation and dynamics .1. *Proceedings of The National Academy of Sciences of The United States of America-Biological Sciences*, 78:2179–2183, 1981.

[36] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.

[37] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh Ewald method. *Journal Of Chemical Physics*, 103:8577–8593, 1995.

[38] S. Nosé. A unified formulation of the constant temperature molecular-dynamics methods. *J. Chem. Phys.*, 81:511–519, 1984.

[39] W. G. Hoover. Canonical dynamics—equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985.

[40] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes. *Journal Of Computational Physics*, 23:327–341, 1977.

[41] W. Saenger. *Principles of Nucleic Acid Structure*. Springer-Verlang: New York, 1984.

[42] I. Andricioaei and M. Karplus. On the calculation of entropy from covariance matrices of the atomic fluctuations. *Journal of Chemical Physics*, 115:6289–6292, 2001.

[43] R. Lavery and H. Sklenar. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic-acids. *Journal Of Biomolecular Structure & Dynamics*, 6:63–91, 1988.

[44] W. Humphrey, A. Dalke, and K. Schulten. Vmd: Visual molecular dynamics. *Journal Of Molecular Graphics*, 14:33–&, 1996.

[45] G. M. Torrie and J. P. Valleau. Non-physical sampling distributions in Monte-Carlo free-energy estimation - umbrella sampling. *Journal of Computational Physics*, 23:187–199, 1977.

[46] S. Kumar, D. Bouzida, RH. Swendsen, PA. Kollman, and JM. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. *J. Comput. Chem*, 13:1011–1021, 1992.

[47] A. Grossfield. Source code from http://dasher.wustl.edu/alan/.

[48] R. E. Dickerson and H. R. Drew. Kinematic model for B-DNA. *Proceedings of The National Academy of Sciences of The United States of America-Biological Sciences*, 78:7318–7322, 1981.

[49] A. Sarkar, J. F. Leger, D. Chatenay, and J. F. Marko. Structural transitions in DNA driven by external force and torque. *Physical Review E*, 63:051903, 2001.

[50] T Lionnet, S. Joubaud, R. Lavery, D. Bensimon, and V. Croquette. Wringing out DNA. *Phys Rev Lett*, 96:178102, 2006.

[51] J. Gore, Z. Bryant, M. Nollmann, M. U. Le, N. R. Cozzarelli, and C. Bustamante. DNA overwinds when stretched. *Nature*, 442:836–839, 2006.

[52] D. P. Millar, R. J. Robbins, and A. H. Zewail. Direct observation of the torsional dynamics of DNA and RNA by picosecond spectroscopy. *Proceedings of The National Academy of Sciences of The United States of America-Physical Sciences*, 77:5593–5597, 1980.

[53] W. Olson. How flexible is the furanose ring? 2. An updated potential energy estimate. *Journal of the American Chemical Society*, 104:278–286, 1980.

[54] N. Foloppe and A. D. MacKerell. Conformational properties of the deoxyribose and ribose moieties of nucleic acids: A quantum mechanical study. *J. Phys. Chem. B*, 102:6669–6678, 1998.

[55] T. E. Cheatham, P. Cieplak, and P. A. Kollman. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *Journal Of Biomolecular Structure & Dynamics*, 16:845–862, 1999.

[56] J. F. Marko. Stretching must twist DNA. *Europhysics Letters*, 38:183–188, 1997.

[57] R. D. Kamien, T. C. Lubensky, P. Nelson, and C. S. OHern. Direct determination of DNA twist-stretch coupling. *Europhysics Letters*, 38:237–242, 1997.

[58] J. Wereszczynski and I. Andricioaei. On structural transitions, thermodynamic equilibrium, and the phase diagram of dna and rna duplexes under torque and tension. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 103:16200–16205, 2006.

CHAPTER III

Conformational and Solvent Entropy Contributes to Switching Activity in Nucleic Acid-Based Nanothermometers

3.1 Introduction

There exists substantial excitement in the development of methods to control mechanical movement in molecular-scale devices composed of nucleic acids through triggering conformational changes of the respective molecules by various stimuli [1, 2, 3]. With increasing miniaturization in nanodevices, and given the fundamental difference between physicochemical properties in the macro and nano regimes, there is also substantial interest (and significant challenge) in accurately measuring temperature changes over nanometer scales in diverse fields spanning from nanofluidics to computer chip design to hyperthermal cancer treatment. In the opposite direction of inducing motion by local stimuli, measuring the conformational change has been recently proposed as a means to detect local changes in temperature on the nanometer scale, which would qualify these devices as "nanothermometers," and would make them very useful in taking temperature measurements in nanoscale devices or in mapping temperature differences within single cells [4]. In a set of recently reported exciting fluorescence experiments Tashiro and Sugiyama have shown that, by taking advantage of the characteristic properties of nucleic acids in solution, nanothermome-

ters may be constructed based on the temperature sensitivity of helical transitions in DNA and RNA duplexes [5, 6]. In these experiments a fluorescent adenine analogue that is sensitive to the direction of helicity was inserted into an alternating purine-pyrimidine sequence, and in high ionic conditions, upon increasing the temperature, the RNA transitioned from A- to Z-form (as gauged from the increase in the fluorescence of the base), while the DNA transitioned from Z- to B-form (reflected by a decrease in fluorescence). This is a fantastic feat because by designing a device with an appropriate composition of RNA and DNA duplexes one may envision devices that could accurately measure temperatures on a fine length scale.

It has become clear from recent experimental and theoretical studies (see Auffinger [7] for a review) that water molecules and ions within complex nucleic acid structures can display long residency times, and assist drug binding and catalytic reactions. For example, an MD study showed that the hydration shell was affected by the helical conformation (A- or B- form), the sequence, and the ionic concentration of the environment [8]. It is therefore expected that these considerations should play a role in conformational switching between B, Z, or A forms. For example, the ability of canonical B-DNA to undergo switching to left handed Z-DNA is strongly salt and sequence dependent [9, 10]. While for this B- to Z- transition the electrostatic effect of the solvent including ions is well calibrated [11], less understood is the contribution of the entropy change upon switching. A key factor in the relative stabilization of Z- and B- form DNA is described by the observation that the closest approach distance between phosphates on opposites strands decreases from 11.7 Å in B-DNA to 7.7 Å in Z-DNA [12]. It is thus not surprising that a high salt concentration screens this electrostatic repulsion and helps to stabilize Z-DNA [9]. Using nonlinear Poisson-Boltzmann calculations, Misra and Honig [11] have analyzed the electrostatic

contributions to the B- to Z- transition in DNA and have concluded that electrostatic forces are favoring Z-DNA, and that the dominant role in the relative stabilization of B- is by consequence dictated by non-electrostatic forces, out of which entropy was suggested to be significant. Moreover, while the DNA B-Z switching is well studied and simulated [13], relatively less is known about RNA switching. It is known that the stabilization of Z-RNA tends to require an even higher salt concentration and temperature than Z-DNA, potentially due to the increased energy required to force pyrimidine residues into a C2'-endo puckering conformation [14].

Despite its importance in nanoscale thermometric technology, a detailed understanding of the microscopic origin of the temperature induced switching activity of nucleic acid based nanothermometers is lacking. To investigate these questions we have performed a series of MD simulations on RNA and DNA duplexes and calculated their conformational, solvent, and ionic entropies. By comparing entropies of different helical structures with different sequences and in varying ionic conditions we show that conformational entropy plays an important role in explaining the thermodynamic basis for these devices and may partially account for the sequence dependence and ionic conditions required by these experiments. Calculations of solvent and ionic entropies have traditionally been infeasible due to the large volume of conformational space sampled by solvent molecules in MD simulations, however we employ a recently developed method of permutation reduction which serves to lessen this complication and provide estimates for these values [15].

3.2 Methods

We have constructed 16 separate systems of nucleic acids in a water box and simulated them with periodic boundary conditions and constant pressure and temperature

through the Nose-Hoover algorithm [16, 17, 18] with the program CHARMM [19], using version 27 of the nucleic acid force field parameters [20, 21]. For each nucleic acid form (i.e., A, B or Z), half the simulations were performed with a mixed sequence $(GCGAAATTTGCG)_2$, with uridine replacing thymine in RNA molecules, while the other half were done with a C-G sequence analogous to what was used in experiments $(CGCGCGCGCGCG)_2$. Of the total number of simulations with both mixed and C-G sequences, half were performed with a right handed helix (A-RNA or B-DNA) and half in the Z- form. Furthermore, half of the simulations were performed with just enough sodium counterions to neutralize the charge of the system, while the other half included a number of sodium and chloride ions that created an approximately 4M NaCl environment (with excess sodiums to neutralize the system). A 4M NaCl environment was used for direct comparison to the DNA data presented by Tashiro and Sugiyama (Figure 4 of [6]) and was maintained for the RNA simulations for consistency. Therefore, each combination of nucleic acid, helix conformation, sequence, and ionic conditions was simulated. For each system following an equilibration period of 200 ps with harmonic restraints on the nucleic acid heavy atoms and extensive minimization, five 2-ns simulations were calculated, yielding a total of 80 2-ns simulations.

Conformational entropies were calculated by the use of quasiharmonic frequency analysis [22], a method in which the eigenvalues of the mass weighted covariance matrix are calculated to determine the quasiharmonic frequencies, $\omega_i = \sqrt{kT/\lambda_i}$, which are then used to calculate the conformational entropy via:

$$S = k \sum_{i}^{3n-6} \frac{\hbar\omega_i/kT}{e^{\hbar\omega_i/kT} - 1} - ln\left(1 - e^{-\hbar\omega_i/kT}\right). \tag{3.1}$$

Frequencies calculated from the nucleic acid's fluctuations beginning at 500 ps of simulation time (to allow equilibration of the nucleic acid) and ending at each 100

ps interval until the end of the simulation time were used.

The entropy at infinite sampling time S_{inf} was extrapolated from the time dependent entropies at the time points described above by using the formula [23, 24]:

$$S(t) = S_{inf} - \frac{\beta}{t^{\alpha}} \tag{3.2}$$

where α was chosen to best fit the data (usually a value between 1.5 and 1.7). The mean conformational entropies at infinite sampling time were calculated and are shown in Tables 3.1 and 3.2, with the error bars representing the standard deviations between each of the five simulations for each set of conditions. The driving force for the transitions are however the *change* in free energy between the right and left handed helical states, and the free energy change as a results of the sum of conformational, solvent, and ionic entropies (as well as from just the conformational part) at 293 K are shown in Table 3.3,

Recently a method of permutation reduction in which water molecules are "relabeled" was developed. This method aims at minimizing the difference between a permuted frame of a trajectory and a reference structure. That is, a permutation (or "relabeling") π_i on a frame $x(t_i)$ is chosen so that the difference between the new frame and a reference structure x_0 is minimized (Equation 3 in [15]):

$$min\left(|\pi_i \cdot x(t_i) - x_0|^2\right). \tag{3.3}$$

This method was applied to solvent and ionic molecule (in high salt simulations) with code provided by the authors. Following this quasiharmonic analysis was performed and entropies at infinite sampling were calculated as before. For water molecules entropies were only calculated for those with an average oxygen position over the first 100 ps less then 3 Å away from a nucleic acid heavy atom (although all were

	Mixed Sequence		C-G Sequence	
	A-RNA	Z-RNA	A-RNA	Z-RNA
0 M [NaCl]	$2.418 \pm .069$	$2.319 \pm .044$	$2.308 \pm .054$	$2.409 \pm .023$
4 M [NaCl]	$2.293 \pm .027$	$2.280 \pm .036$	$2.199 \pm .034$	$2.299 \pm .035$

Table 3.1: RNA conformational entropy values in units of kcal·mol^{-1}·K^{-1}

	Mixed Sequence		C-G Sequence	
	B-DNA	Z-DNA	B-DNA	Z-DNA
0 M [NaCl]	$2.339 \pm .009$	$2.583 \pm .109$	$2.255 \pm .033$	$2.515 \pm .040$
4 M [NaCl]	$2.273 \pm .020$	$2.440 \pm .042$	$2.173 \pm .062$	$2.316 \pm .038$

Table 3.2: DNA conformational entropy values in units of kcal·mol^{-1}·K^{-1}

permuted) as this minimized the sum of the variances from water entropies. For ions all atoms in the system were used, with permutations and entropies from sodium and chloride atoms calculated separately and combined for presentation.

3.3 Results

RNA Switching experiments suggest Z-RNA to be entropically favored over A-RNA, at least in the case of a C-G sequence in high ionic environments. Conformational entropy calculations (Table 3.1) clearly demonstrate this as C-G Z-RNA dodecemers in low and high salt concentrations show a stabilization of 23 to 29 kcal/mol. Mixed sequence RNA molecules show an inverted response, that is A-RNA is the favored form in low salt whereas in high salt the two conformations have almost identical conformational entropies.

Entropy differences of the solvent shells are large and overwhelm conformational entropies in low ionic environments (see Figure 3.1(a)) such that the total entropic

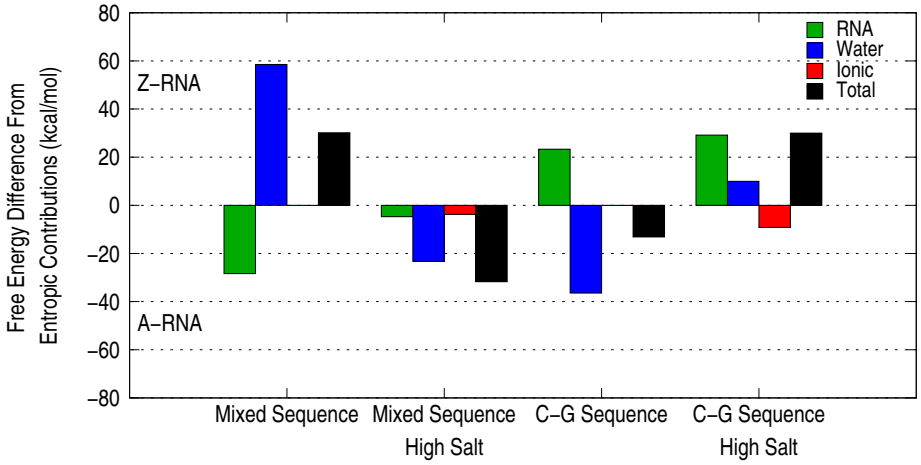
	Z→A-RNA		Z→B-DNA	
	Mixed Sequence	C-G Sequence	Mixed Sequence	C-G Sequence
0 M [NaCl]	30.1 ± 93.0 (-28.4 ± 20.8)	-12.2 ± 81.4 (23.3 ± 21.4)	-2.1 ± 85.4 (70.0 ± 27.8)	32.6 ± 87.8 (76.2 ± 13.8)
4 M [NaCl]	-3.5 ± 26.9 (-4.7 ± 12.6)	30.0 ± 30.9 (29.2 ± 12.9)	56.5 ± 36.1 (49.4 ± 12.0)	-28.2 ± 61.6 (40.4 ± 20.2)

Table 3.3: Total (and conformational) free energy change, -TΔS, of transition from left to right handed forms in units of kcal/mol at T=293 K.

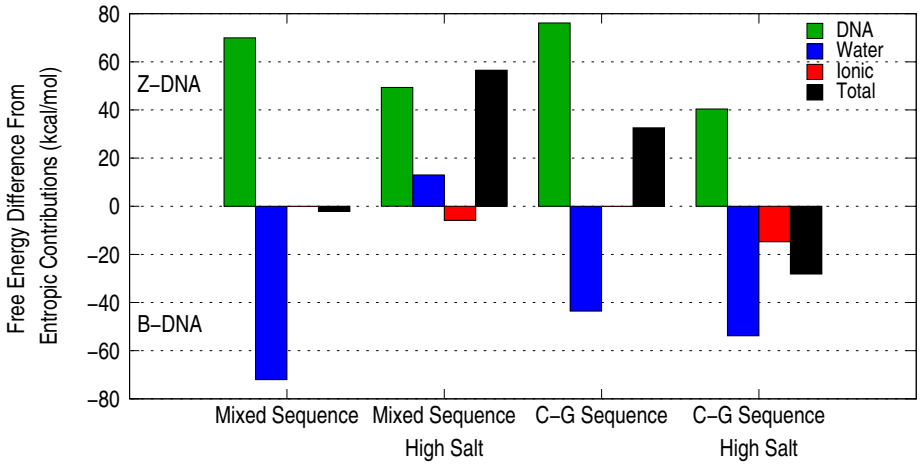
value of the system favors Z- form in the case of a mixed sequence and A- form in C-G sequences (Table 3.3). High ionic environments tend to suppress solvent entropy differences, and in the C-G simulations ionic and solvent entropies nearly negate one another resulting in conformational entropy dominating the entropic differences between left and right handed systems.

DNA Comparisons between entropies for DNA and RNA systems show interesting differences from one another. While conformational entropy differences in RNA are largely sequence dependent, DNA conformational entropy differences show a high correlation with ionic environment and little sequence dependence (see Table 3.2). Absolute conformational entropies do show a decrease of 0.094±.024 kcal/mol/K for C-G sequences relative to mixed ones, possibly a result of more Watson-Crick hydrogen bonds in C-G sequences serving to stabilize the dodecmer and thus reducing motions. The presence of ions dampens conformational entropies in both sequences, however the reduction for Z-DNA molecules is over twice as large as for B-DNA ones. Nevertheless in all situations left handed helices are strongly favored by conformational entropy over right handed ones.

Similar to results on RNA molecules solvent and ionic entropies tend to oppose the stabilization effect of conformational entropy differences, with the exception being the mixed sequence in high salt environment where a slight reinforcement of the conformational entropy difference is observed (Figure 3.1(b)). Mixed sequences in low salt environments show an almost complete cancellation of entropy differences between Z- and B- form, whereas with the C-G sequence the solvent difference is much less in low ionic environment and there is still a preference for Z- form. The only situation where a B- form is entropically preferable is in the C-G high ionic



(a) RNA



(b) DNA

Figure 3.1: Contributions to entropic switching between left and right handed helices for (a.) RNA and (b.) DNA. Free energy differences from nucleic acid (green), water (blue), salt ions (red, only in high salt simulations), and their sum (black). Left handed helices are favored by positive values while those favoring right handed ones are negative.

conditions simulations where both the solvent and ionic entropies combine to over-compensate for the conformational entropy favoring Z- form. This is in agreement with experimental results which suggest that this system has a higher entropy in B-form due to the fact that the Z- to B- transition occurs as temperature increase.

3.4 Discussion

The results presented here serve to explain the switching seen between left and right handed nucleic acid devices as a complex interplay between conformational, solvent, and ionic entropy contributions. In the RNA case sequence dictates the direction of conformational entropy preference, whereas an increase in ionic environments tend to flip solvent entropy from opposing to reinforcing the direction favored by conformational entropy. For DNA the strength of the ionic environment is the determining factor for the magnitude of the conformational entropy difference (which always favors Z- form) while solvent and ionic entropies tend to favor B-form. In the case of a C-G sequence in high salt, solvent and ionic entropies are enough to overcome the conformational entropy and favor B-form. For both RNA and DNA C-G sequence in high salt environments these calculations are in agreement with experiments which indicate an entropic preference for Z-RNA and B-DNA (by detecting a switching to these forms as the temperature increases). A study of high-resolution crystal structures showed water molecules uniquely hydrate Z-DNA with an alternating C-G sequence [25], thus it is not surprising that an entropic solvent difference would exist between forms of nucleic acid sequences.

The method of permutation reduction used here shows promise as a tool in calculations of solvent entropies. Absolute entropies calculated from this method corresponded to free energies ranging from -5.30 to -6.21 kcal/mol/molecule, similar to values calculated from another recent theoretical scheme based on velocity and rotational autocorrelation methods that predicted entropies of individual water molecules surrounding DNA to be between -6.41 and -7.27 kcal/mol [26]. The large number of solvent molecules results in the sum of water entropies being much higher than

those for the nucleic acid sequences, causing large error bars when differences are taken between systems. These errors were substantially reduced by examining only a subset of the solvent (those within 3 Å of the macromolecule) however they are still quite large in some calculations (as can be seen by the difference between the errors in conformational and total entropy differences in Table 3.3). Despite these errors it is the sign of the difference which is most important and provides insight into nucleic acid dynamics.

It is interesting to make connection with a comparative study of DNA and RNA performed. Nillson and coworkers, and, independently, Mackerrel and coworkers report results from MD simulations indicating different structural fluctuations of duplex RNA relative to DNA, leading to increased base pair opening events on the nanosecond time scale in the former (being favored by as much as 12 kcal/mol for a single opening) [27, 28]. Comparisons between A-RNA and B-DNA consistently show a higher conformational entropy in the case of RNA (in agreement with results presented here) which suggest that the increased base pair fluctuations may be due to the increase in conformational entropy. It is evident the conformational entropy of left and right handed nucleic acid helices play an important role in their thermal transitions, and that the driving force resulting from these difference can explain the sequence and salt dependence of these nanothermometers devices. These calculations are akin to MD simulations on proteins which demonstrate that temperature dependence of ionic conditions are related to amino acid sequence [29], and further demonstrate the complex interplay between local structures, ionic conditions, and conformational entropy in biomolecules.

A pioneering study of Irikura, Tidor and Karplus [30] had indicated an opposite result, that normal mode analysis of B-DNA showed higher configurational entropy

then Z- form. Our normal mode analysis on the C-G sequence/high salt B- and Z-DNA molecules showed a similar result when a distance dependent dielectric constant was used, however when calculations were performed with a Generalized Born solvation model [31] the configurational entropy switched to favoring the Z-form (as in the quasiharmonic results). While it is difficult to make conclusions from these differences they do provide further evidence that solvent conditions have a significant effect on the entropy difference between helical states.

BIBLIOGRAPHY

[1] H. Yan, X. P. Zhang, Z. Y. Shen, and N. C. Seeman. A robust DNA mechanical device controlled by hybridization topology. *Nature*, 415:62–65, 2002.

[2] C. D. Mao, W. Q. Sun, Z. Y. Shen, and N. C. Seeman. A nanomechanical device based on the B-Z transition of DNA. *Nature*, 397:144–146, 1999.

[3] B. Yurke, A. J. Turberfield, A. P. Mills, F. C. Simmel, and J. L. Neumann. A DNA-fuelled molecular machine made of DNA. *Nature*, 406:605–608, 2000.

[4] J. Lee and N. A. Kotov. Thermometer design at the nanoscale. *Nano Today*, 2:48–51, 2007.

[5] R. Tashiro and H. Sugiyama. A nanothermometer based on the different pi stackings of B- and Z-DNA. *Angewandte Chemie-International Edition*, 42:6018–6020, 2003.

[6] R. Tashiro and H. Sugiyama. Biomolecule-based switching devices that respond inversely to thermal stimuli. *Journal Of The American Chemical Society*, 127:2094–2097, 2005.

[7] P. Auffinger and Y. Hashem. Nucleic acid solvation: from outside to insight. *Current Opinion In Structural Biology*, 17:325–333, 2007.

[8] M. Feig and B. M. Pettitt. A molecular simulation picture of DNA hydration around A- and B-DNA. *Biopolymers*, 48:199–209, 1998.

[9] F. M. Pohl and T. M. Jovin. Salt-induced cooperative conformational change of a synthetic DNA - equilibrium and kinetic studies with poly(dg-dc). *Journal Of Molecular Biology*, 67:375–&, 1972.

[10] A. Rich, A. Nordheim, and A. H. J. Wang. The chemistry and biology of left-handed Z-DNA. *Annual Review Of Biochemistry*, 53:791–846, 1984.

[11] V. K. Misra and B. Honig. The electrostatic contribution to the B to Z transition of DNA. *Biochemistry*, 35:1115–1124, 1996.

[12] A. H. J. Wang, G. J. Quigley, F. J. Kolpak, G. Vandermarel, J. H. Vanboom, and A. Rich. Left-handed double helical DNA - variations in the backbone conformation. *Science*, 211:171–176, 1981.

[13] M. A. Kastenholz, T. U. Schwartz, and P. H. Hunenberger. The transition between the B and Z conformations of DNA investigated by targeted molecular dynamics simulations with explicit solvation. *Biophysical Journal*, 91:2976–2990, 2006.

[14] M. K. Teng, Y. C. Liaw, G. A. Vandermarel, J. H. Vanboom, and A. H. J. Wang. Effects of the o2' hydroxyl group on Z-DNA conformation - structure of Z-RNA and (arac)-[Z-DNA]. *Biochemistry*, 28:4923–4928, 1989.

[15] F. Reinhard and H. Grubmuller. Estimation of absolute solvent and solvation shell entropies via permutation reduction. *Journal Of Chemical Physics*, 126:014102, 2007.

[16] S. Nosé. A unified formulation of the constant temperature molecular-dynamics methods. *J. Chem. Phys.*, 81:511–519, 1984.

[17] W. G. Hoover. Canonical dynamics—equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985.

[18] S. Nose and M. L. Klein. Constant pressure molecular-dynamics for molecular-systems. *Molecular Physics*, 50:1055–1076, 1983.

[19] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM - a program for macromolecular energy, minimization, and dynamics calculations. *Journal Of Computational Chemistry*, 4:187–217, 1983.

[20] N. Foloppe and A. D. MacKerell. All-atom empirical force field for nucleic acids: I. parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal Of Computational Chemistry*, 21:86–104, 2000.

[21] A. D. MacKerell, N. Banavali, and N. Foloppe. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, 56:257–265, 2000.

[22] I. Andricioaei and M. Karplus. On the calculation of entropy from covariance matrices of the atomic fluctuations. *Journal of Chemical Physics*, 115:6289–6292, 2001.

[23] S. A. Harris, E. Gavathiotis, M. S. Searle, M. Orozco, and C. A. Laughton. Cooperativity in drug-DNA recognition: A molecular dynamics study. *Journal Of The American Chemical Society*, 123:12658–12663, 2001.

[24] S. A. Harris and C. A. Laughton. A simple physical description of DNA dynamics: quasi-harmonic analysis as a route to the configurational entropy. *Journal Of Physics-Condensed Matter*, 19, 2007.

[25] R. V. Gessner, G. J. Quigley, and M. Egli. Comparative-studies of high-resolution Z-DNA crystal-structures .1. common hydration patterns of alternating dc-dg. *Journal Of Molecular Biology*, 236:1154–1168, 1994.

[26] B. Jana, S. Pal, P. K. Maiti, S. T. Lin, J. T. Hynes, and B. Bagchi. Entropy of water in the hydration layer of major and minor grooves of DNA. *Journal Of Physical Chemistry B*, 110:19611–19618, 2006.

[27] K. Hart, B. Nystrom, M. Ohman, and L. Nilsson. Molecular dynamics simulations and free energy calculations of base flipping in dsRNA. *Rna-A Publication Of The Rna Society*, 11:609–618, 2005.

[28] N. K. Banavali and A. D. MacKerell. Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *Journal Of Molecular Biology*, 319:141–160, 2002.

[29] Ma. L and Cui. Q. The temperature dependence of salt-protein association is sequence specific. *Biochemistry*, 45:14466–14472, 2006.

[30] K. K. Irikura, B. Tidor, B. R. Brooks, and M. Karplus. Transition from B-DNA to Z-DNA - contribution of internal fluctuations to the configurational entropy difference. *Science*, 229:571–572, 1985.

[31] M. S. Lee, F. R. Salsbury, and C. L. Brooks. Novel generalized born methods. *Journal Of Chemical Physics*, 116:10606–10614, 2002.

CHAPTER IV

The Mechanism and Energetics of DNA Supercoil Relaxation by Human Topoisomerase I

4.1 Introduction

Topoisomerases are enzymes which play a crucial role in cellular functions by altering the supercoiling topology of DNA, which in turn affects transcription, replication, recombination, nucleosome interactions, and a host of other processes vital to cell proliferation [1, 2, 3, 4]. They are divided into two classes: type I enzymes which nick one strand of the phosphodiester backbone and passively allow the relaxation of supercoils, and type II enzymes which break both strands and, with energy provided through ATP hydrolysis, may increase or decrease the DNA's linking number [5]. Both classes are further subdivided into types A or B. Type IA enzymes covalently attach to the 5' phosphate whereas type IB enzymes covalently attach to the 3' end [6].

Human topoisomerase I (topo1), a type IB enzyme, consists of a single chain of 765 amino acids. Its structure in complex with duplex DNA was obtained in a crystallographic tour de force in 1998 [7], which revealed the structural basis of a swivel mechanism proposed as early as 1972 [8]. The enzyme functions by clamping around DNA, nicking it, and forming a covalent bond between the active-site Tyrosine and the 3' end of the DNA upstream of the cut while capping the 5' end of the down-

stream DNA with an OH group. Following this step supercoils are relaxed, the DNA backbone is religated, and the protein releases the DNA.

The protein structure consists of six distinct domains: an unordered N-terminal end that extends until residue 214, Core Subdomain (C.S.) I which consists of amino acids 215-232 and 320-433, C.S. II going from 233-319, C.S. III running from 434-635, the Linker domain which consists of positively charged amino acids from 636-712, and the C-terminal domain. C.S. I and II form the "cap" of the enzyme and hydrogen bonds between C.S. I and III form the "lips" of the protein [7, 9]. The "lips" region must dynamically open and closes to allow for clamping and release of the DNA before and after supercoil release. These various structural components (the linker, lips and hinge) function such that they contribute to a "controlled rotation" of the DNA inside the grip of the protein [10].

In our previous computational work we suggested, based on evidence from molecular dynamics simulations, that the relaxation of positive and negative supercoils processed along different pathways. When positive supercoils were present, the DNA forced the "lips" open by 10-14 Å to create room for the rotation of the downstream DNA, while for the relaxation of negative supercoils the "hinge" region stretched by $\sim$12 Å [11].

The importance of human topoisomerase I for cell replication has made it a prime chemotherapeutic target, with the camptothecin class of compounds being found to selectively inhibit topo1. A crystal structure was solved in which the camptothecin analogue Topotecan (TPT) was complexed with a topo1 molecule and a 22 base pair DNA segment[7]. This work showed that TPT, a five-ringed molecule, mimics a base pair and intercalates between the downstream and upstream DNA segments, creating base-stacking interactions with the -1 and +1 basepairs and forcing the free

5'-OH group further away from the active-site tyrosine. It was proposed that this increased separation would substantially decrease the probability for religation of the backbone, trapping the DNA/topo/TPT ternary complex in a covalently attached state that would prohibit interaction of the DNA with cellular machinery for replication and transcription. A combination of single-molecule and *in vitro* experiments suggested another mechanism for the cytotoxic behavior: TPT selectively inhibits the relaxation of positive supercoils substantially more than negative ones by topo1 [12]. In fact the researchers saw an inhibition in the rate of positive supercoil relaxation of approximately 20 fold as compared to negative supercoils being inhibited by 4 fold, resulting in not only the retardation of enzymatic function but also the buildup of positive supercoils in poisoned cells. The single molecule observation, the fact that negative uncoiling is faster than the positive one, was confirmed by in vivo studies that showed a build-up of positive supercoils in yeast cells during G1 and S phase as topo1 preferentially removed negative supercoils.

Related single molecule studies by the same group on another IB variant, vaccina topoisomerase, demonstrated that multiple supercoils were relieved per nicking event in these enzymes, and that the distribution for the change in linking number (ΔLk) has a probability distribution that can be modeled as a decaying exponential [13]. Additionally, this distribution is also based on the tension in the downstream DNA, suggesting the mechanism for rotation is a swivel motion, and the relaxation rate appeared to be similar for positive (overtwisted DNA) and negative (undertwisted) supercoils. The human topo single molecule study also concluded that the positive and negative relaxation rates are similar in the absence of topotecan.

Why do the rates differ significantly when TPT is bound? How can we reconcile the hypothesis put forth in our previous molecular dynamics work (the existence of

different pathways for positive and negative supercoiling) with the measured rates of rotations? In this paper we significantly extend our simulation work by calculating free energy profiles for the rotation of DNA downstream of the cut in the central pore of human topoisomerase I to mimic the relaxation of positive and negative supercoils. These calculations are then repeated on a ternary complex that contains DNA, topoisomerase, and TPT, and the results are used to determine how TPT alters the mechanisms and energies of supercoil relaxation.

4.2 Methods

A model of a covalently bound DNA/topoisomerase binary complex was made using the structure solved by Champoux, Hol et al., with PDB code 1A31 [7], with the linker region atom positions coming from the non-covalent structure (PDB code 1A36). This system was placed in a box of TIP3P water molecules [14] of dimensions 105 x 120 x 90 $Å^3$ and ionized with sodium ions to create an electrically neutral system of 117,000 atoms. Constant temperature and pressure algorithms [15] along with periodic boundary conditions were employed for all calculations. Following a 1 ns equilibration period umbrella sampling was used to calculate a free energy profile of the rotation of the downstream DNA around an axis that is defined by a vector that passed through the original positions of the phosphorus atom of the -1 adenine and the phosphorus atom of the +10 adenine. Initial conditions were created by rotating the downstream DNA in increments of 10°, harmonically restraining all of the DNA heavy atoms, and allowing the protein to undergo minimization followed by a period of equilibration before a subsequent rotation. For the topo/DNA binary system mimicking the relaxation of positive supercoils a 50 ps equilibration time was taken while for the other three systems this was increased to 500 ps. The results

of the shorter equilibration in this phase was a longer time for equilibration in the next phase, however the final results were not significantly altered. Each window of the reaction coordinate was chosen as the dihedral between the center of mass of the downstream DNA, the cartesian coordinates previously mentioned, and the position where the center of mass of the downstream DNA was in the un-rotated structure. A harmonic restraint was placed on this dihedral with a minimum corresponding to the window being simulated. Windows were sampled from -10° to 560° (10° to -560° in the case of negative supercoils) and were simulated for a minimum of 10 ns using NAMD [16] (longer for those windows which had not fully equilibrated in the binary/positive supercoil system) with the data from the final 4 ns being used in the WHAM algorithm [17] (with code provided by Grossfield [18]) for free energy calculations (for a total of 232 ns of sampling in each free energy profile). The totality of the four potential of mean force mappings took approximately 176 CPU years on three supercomputing clusters. The CHARMM force field version 27 was used for all calculations [19, 20] and the CHARMM program was used for system setup and analysis[21]. Error bars were estimated using a bootstrap algorithm with the mean statistical inefficiency taken as the decorrelation time [22].

For simulations with topotecan the 1K4T crystal structure was used [23], and the protein was truncated so that residues not present in the 1A31 structure would not be present in this system (so that both systems began at amino acid 215). Topotecan was built using the Insight II program (Molecular Simulations Inc.) and parameters were inserted into the CHARMM force field. The same procedure as detailed above was performed with the exception that during the creation of initial structures the drug molecule was fixed so that the abrupt disruptions between the DNA segments did not artificially move the topotecan, but during equilibration of each window this

constraint was released.

While performing these calculations the authors noted a paper which had results indicating that for the relaxation of negative supercoil the presence of Tryptophan 205 played an important role[24], thus residues 201-215 (which were not present in the original crystal structure by Champoux, Hol et al. but were present in the structure that included TPT) were added to both system for the relaxation of negative supercoils. Their positions were taken from the 1K4T structure, and all other procedures remained similar to those for the relaxation of positive supercoils.

Calculations of rate constants were performed by numerically integrating the Smoluchowski equation to determine the mean first passage time from a location x to one revolution beyond x, $x \pm 360$ (depending on the sign of the supercoils being released) and taking the inverse [25]:

$$k = (\tau(x))^{-1} = \left(\int_{x}^{x_{max}} dy e^{U(y)/kT} / D(y) \int_{-\infty}^{y} dz e^{U(z)/kT} \right)^{-1} \tag{4.1}$$

where $U(x)$ represents the free energy at x and D the rotational diffusion coefficient, which was position-dependent and calculated from the individual umbrella windows [26, 27]:

$$D_0(Q) = \frac{\Delta(Q)^2}{\tau_{corr}} \tag{4.2}$$

It has been demonstrated that the rotational diffusional coefficient for DNA is dependent on the length of the DNA sequence [28, 29], thus while simulations were performed with 9 constrained basepairs a correction factor for D_0 was required to calculate an experimentally observable value for D:

$$D(x) = \frac{D_{BN}(l)}{D_{BN}(l_0)} D_0(x), D_{BN}(L) \propto \frac{ln(L/2r) - 0.76 + 7.5\left[1/ln(L/r) - 0.27\right]^2}{L^3} \tag{4.3}$$

with l being chosen as the persistence length (50 nm), l_0 the length of the 9 constrained basepairs in the reaction coordinate (2.97 nm), and r the DNA radius (1.4 nm).

The Marko model for torque was derived through a statistical mechanical approach in which the free energy of the plectonemic and elongated states was minimized for a given tension. It reports that the critical torque is [30]:

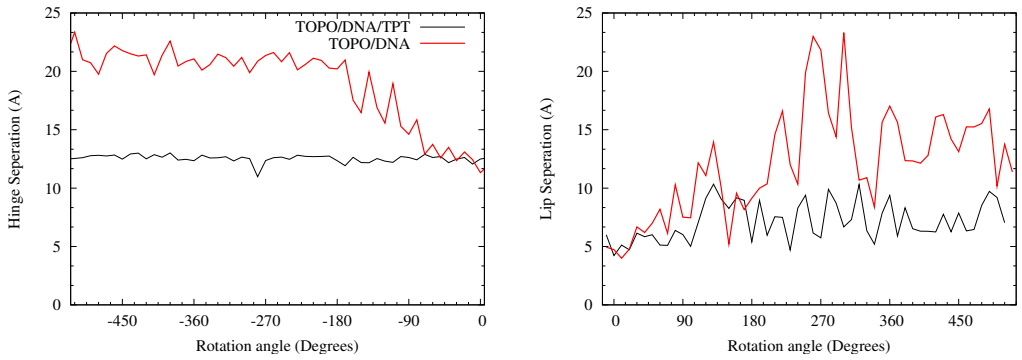
$$\Gamma_c(F) = \sqrt{\frac{2k_bTPg}{1 - P/C_s}} \tag{4.4}$$

where g is the free energy of the extended-state and, C_s the twist stiffness of the extended state, and P the stiffness of the plectonemic state. The first two quantities are given as functions of the tension, and an estimate is made for the third. To fit the data of Forth *et al.* [31] we have introduced a linearly variable P (linear as it provided the best fit) of the form $P(f) = 35.3 - 3.4f$. While this creates values for P beyond the estimated 21-27 nm the overall fit is improved for the range where experimental data exists.

4.3 Results

4.3.1 Relaxation of Positive Supercoils

Simulations of the relaxation of positive supercoils by the topo/DNA complex showed an opening of the "lips" region by as much as 18.5 Åin some windows, similar to previous results (Figure 4.1(b)). Surprisingly the calculations for the ternary topotecan complex show that the hydrogen bonds between the "lips" are not broken



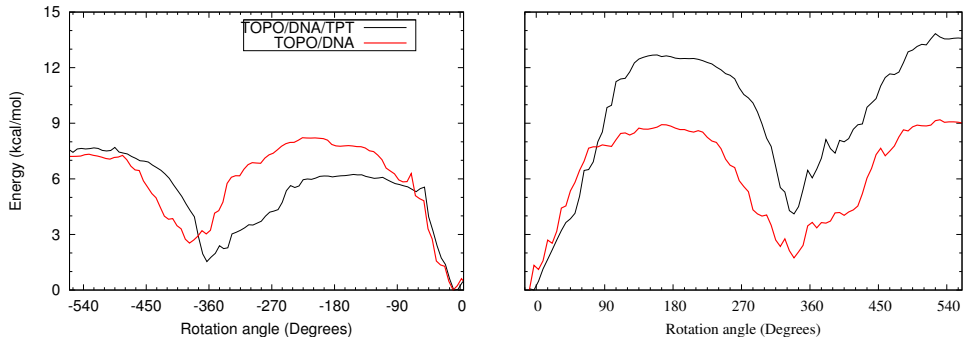
(a) Hinge Separation Negative Supercoil Relaxation

(b) Lip Separation for Positive Supercoil Relaxation

Figure 4.1: Separation of the "hinge" (a.) and "lips" (b.) regions of the topoisomerase as a function of rotation angle.

by rotation of the downstream DNA. The PMF for these processes (Figure 4.2(b)) show less energy is initially required to rotate the DNA in the ternary complex, but at approximately 80° rotations become more prohibitive due to the lack of opening by the "lips." Since TPT forces DNA out of the binding pore, there is less protein that the DNA must interact with to rotate initially (hence the lower energy), but this results in the DNA not creating enough force on the protein to open the "lips" which results in increasing the energy barrier for the initial rotations by 3.8 kcal/mol.

When the DNA has completed its first revolution, the protein does not return to its original clamped conformation, rather in the binary complex it assumes a state where the "lips" are still separated from one another by ∼8 Å (Figure 4.1(b)) that we term the "semi-open" state (Figure 4.3(c)). This state has a free energy 1.7 kcal/mol higher than that of the clamped state, and calculations show that the second DNA rotation (begun in the semi-open state) has an energy barrier of 7.2 kcal/mol. The ternary complex also has a higher energy after one revolution (by 4.1 kcal/mol), however the protein's conformation has not changed dramatically (the protein RMSD when not including the linker region is ∼3.25 Å from its initial state



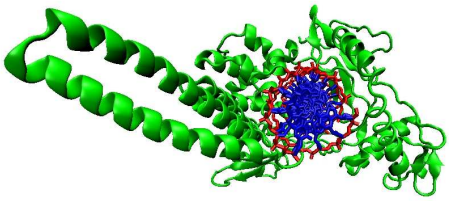
(a) Free Energy Profile for Negative Supercoil Relaxation

(b) Free Energy Profile for Positive Supercoil Relaxation

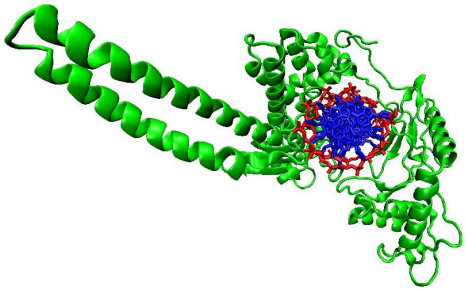
Figure 4.2: Comparison of free energies of rotation for negative and positive supercoils.

compared to ∼5.3 Å for the binary complex). The higher energy is a result of slight modifications to the system structure (such as the downstream DNA having an RMSD of 3 Å rather then 2 Å), however in general we note that the state is close to that of the original clamped conformation. These slight deformations do however decrease the energy barrier for a second rotation to 9.2 kcal/mol.

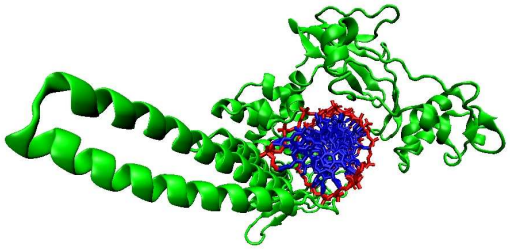
The cross-correlation functions for the C-alpha atoms of the amino acids are presented in Figure 4.4(b) with the top half representing the motions of the atoms in the native system and the bottom half those in the inhibited one (with TPT). These show that there are three regions which have motions highly correlated to one another: C.S. I and II (the cap) move as one unit, C.S. III and the C-terminal, as another and the Linker domain as a third. There is also a negative correlation between the motions of the cap and that of the C.S. III/C-terminal region indicating that as the "lips" separate these two regions, which are on opposite sides of the protein, move apart from one another. Interestingly we note a slight positive correlation between the motions of the cap and the linker domain. The presence of TPT (bottom half of Figure 4.4(b)) shows a dampening of correlations between amino acids that are not



(a) Clamped Conformation



(b) Semi-Open Negative Conformation



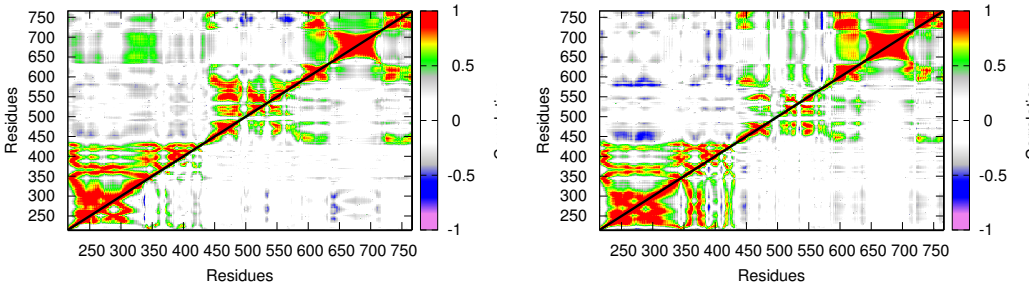
(c) Semi-Open Positive Conformation

Figure 4.3: Topoisomerase structures. (a.) The initial clamped conformation as observed in crystal structures. (b.) A semi-open conformation observed for the relaxation of negative supercoils in which the "hinge" of the protein has opened to allow room for DNA rotation. (c.) A semi-open conformation for the relaxation of positive supercoils in which the "lips" remain open after the first rotation. These semi-open conformations lower the free energy barrier for rotation compared to the initial clamped conformation, thus providing a more favorable mechanism for supercoil relaxation.

in the same domain as each other.

4.3.2 Relaxation of Negative Supercoils

Calculations of negative supercoil relaxation reveal that the "hinge" region of topoisomerase appears to open by ∼11 Å in the absence of TPT, and akin to the case with positive supercoils the presence of TPT prevents the "hinge" from opening (Figure 4.1(a)). Unlike the case of positive supercoils the lack of opening by the enzyme does not increase the energy barrier to a point higher than that of the



(a) Cross-Correlation Functions for Negative Supercoil Relaxation

(b) Cross-Correlation Functions for Positive Supercoil Relaxation

Figure 4.4: Correlation functions for positive and negative supercoil relaxation, both for the free (top half) and inhibited (with TPT-bottom half) simulations.

binary free energy barrier (Figure 4.2(a)).

Also akin to the case of positive supercoil relaxation the "hinge" region of the protein remains open after a single rotation, creating a different "semi-open" state (as shown in Figure 4.3(b)) with energy 2.6 kcal/mol and a barrier for a second rotation of 5.0 kcal/mol. There is also a deformed state for the ternary complex with an energy of 1.5 kcal/mol and which has an energy for a second rotation of 5.8 kcal/mol. Calculations show that in this deformed state the cap region has an RMSD of $\sim$ 5.5 Å (compared to $\sim$ 3 Å in the binary case) but that the internal structure of C.S. I and C.S. II do not change significantly (their RMSD are <3 Å), indicating that it is the interaction geometry between the two subdomains that creates this deformed state.

The cross-correlation maps for negative supercoils (Figure 4.4(a)) show similar structure to that of positive supercoils: the cap region has a strong internal correlation as do C.S. III and the C-terminal domains. The motions of the cap are much more strongly correlated to those of the linker (being mostly positive whereas in the positive supercoil case it is mostly negative) and the motions are still negatively cor-

related to those of C.S. III and the C-terminal domains. Topotecan appears to have a much stronger effect on decorrelating the motions between domains then in the positive supercoil case, and in fact it even decorrelates some of the motions between C.S. I and II.

4.4 Discussion

The calculated free energy profiles lack one important aspect of the *in vitro* system: the driving torque from the torsional strain built up through supercoils. We may modify the calculated PMFs to account for this by the addition of a linear term proportional to this torque, that is:

$$U'(\theta) = U(\theta) - \Gamma_c \theta, \tag{4.5}$$

where θ is the angle and Γ_c the applied torque. An analytic formula relating the tension applied to a supercoiled DNA molecule to the torque was proposed by Marko based on the stiffness of the plectonemic and extended states [30]. Recently this torque was experimentally measured for tensions >1 pN and compared to the Marko model [31], and in the Methods section we discuss a modification to the Marko model which was performed to fit the experimental data.

While a range of 0.2 pN to 3 pN have been applied in studies on vaccinia Topo1B the work with human TopoIB was performed with a tension of 0.2 pN, for an approximate free energy gain of 5.8 kcal/mol/supercoil relieved. Table 4.1 shows how this affects the energy barriers of the systems, decreasing the energy barriers by ∼2-4 kcal/mol and lowering the energy of the first minimums below that of the initial clamped state. Figure 4.5 shows the effects on the overall free energy profiles for all four systems.

To calculate rate constants for these processes we integrate the Smoluchowski

	Torque-Free System			Tilted System		
System	Barrier 1	Minimum 1	Barrier 2	Barrier 1	Minimum 1	Barrier 2
Topo/DNA (+)	$8.9 \pm .27$	$1.7 \pm .35$	$7.4 \pm .29$	$6.8 \pm .27$	$-3.9 \pm .35$	$4.57 \pm .29$
Topo/DNA/TPT (+)	$12.6 \pm .23$	$4.0 \pm .35$	$9.6 \pm .25$	$10.2 \pm .23$	$-1.3 \pm .35$	$6.7 \pm .25$
Topo/DNA (-)	$8.3 \pm .28$	$2.6 \pm .37$	$5.1 \pm .24$	$5.6 \pm .28$	$-3.5 \pm .37$	$3.1 \pm .24$
Topo/DNA/TPT (-)	$6.2 \pm .25$	$1.5 \pm .37$	$6.1 \pm .27$	$4.7 \pm .25$	$-4.2 \pm .37$	$4.0 \pm .25$

Table 4.1: Free energy barriers and the local minima after the first revolution for all four topoisomerase systems, and the effect of adding a constant torque corresponding to a free energy gain of 5.8 kcal/mol/supercoil released (corresponding to a tension of 0.2 pN according to the modified Marko formula used). While the semi-open and deformed states are unfavorable in relaxed systems the energy provided by the release of supercoils makes them favorable and also reduces the second energy barriers to a thermally accessible range of 3-7 kcal/mol.

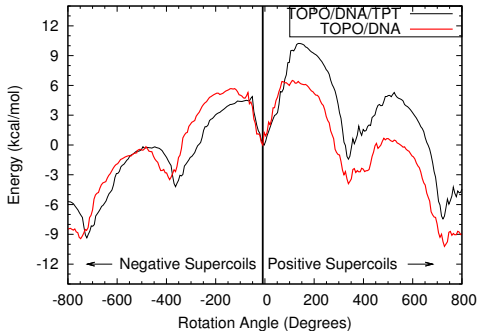


Figure 4.5: Free energy profiles for the binary and ternary complexes to show the effect of a torque resulting in 5.8 kcal/mol/supercoil (corresponding to a tension of 0.2 pN as given by the modified Marko formula).

Rate Constants	Experimental Value	First Rotation	Second Rotation
Topo/DNA (+)	∼75	6.4	34.8
Topo/DNA/TPT (+)	∼3.6	.012	1.4
Topo/DNA (-)	∼61	4.7	147
Topo/DNA/TPT (-)	∼12	268	192

Table 4.2: Calculated and experimental rate constants in Hz for an applied tension of .2 pN (corresponding to a free energy cost of 5.8 kcal/mol/supercoil). The rates for the relaxation of the second supercoil (from a semi-open state) show good agreement for positive supercoil relaxation while negative supercoils show less of a quantitative agreement but appear to not be as strongly affected by the presence of TPT, in accord with experiments. Rates for the first relaxation show a poor agreement with experiments furthering our hypothesis that rotations from a semi-open state predominate.

equation (as discussed in the Methods section) and derive rate constants for both the first and the second rotations, which are shown in Table 4.2 (along with the values from experiments [12]). The rate constants for relaxing the first supercoil show a very strong inhibition by topotecan of positive supercoils (232 fold) and an increase in rate by 38 fold for the addition of TPT to systems that relax negative supercoils. For systems beginning in a semi-open (or deformed) state the effect of TPT on rotation rates is much closer to experimental values, TPT provides an inhibition of 25.5 fold to positive supercoils (compared to 20 in experiments) and for negative supercoils the rates are similar to one another (topotecan rates are higher by 30%) as is seen in experiments where the inhibition by TPT is minimal [12]. These calculations provide further evidence that it is the second barrier, where the system starts in a semi-open or deformed state, that is repeated multiple times during *in vivo* and *in vitro* experiments that have on the order of 10-20 supercoils removed per nicking event.

Rate calculations for negative supercoil relaxation do show a higher quantitative value then those for positive supercoils by 4.2 fold for a given tension while experiment show little discrepancy between the two. This is most likely due to an incomplete understanding of the correlation between tension and torque for negative

supercoils. While formalisms such as those by Marko provide a fair description for positively supercoiled molecules there is a strong possibility that, for a given choice of parameters, they do not provide an accurate description for negative supercoils (in fact it is well known that positive and negative supercoils are quite distinct from one another [32, 33, 34]). As an example if one lowers the plectonemic stiffness to 21 nm (from the variable one discussed in the Methods section) the free energy gain is lowered by 2.2 kcal/mol/supercoil and the rate of negative supercoil relaxation is reduced by 2.4 fold (to 62 Hz).

It is interesting to note that in all four PMFs the state at 360° has a higher energy then the initial clamped state, and that the second energy barrier is approximately the same height as the first when measured relative to this clamped state. To put it another way, after one rotation the system is in a high energy intermediate that is kinetically favored but thermodynamically unfavored by an energy x, and then the barrier for subsequent rotations is lowered by the energy of the intermediate state x. If the system was controlled purely by thermodynamics, then after the initial rotation the molecule would return to its initially clamped state, however our simulations suggest that it is more favorable from a kinetic viewpoint to instead go to a high-energy intermediate which turns out to be useful from the protein's viewpoint in relaxing rotations subsequent the initial one.

4.5 Conclusions

Results suggest the following mechanism for DNA supercoil relaxation by human topoisomerase IB (which is shown in Figure 4.6). Initially DNA exists in a supercoiled state, and topoisomerase in an open state recognizes and binds to the double helix (for a further discussion of these steps see [35]). Following clamping around the DNA a

nicking step must occur to create the cut DNA backbone, after which strand rotation may occur. Torsional stress drives the DNA over the first (high) energy barrier to a transition state, whereby the protein may enter the semi-open conformation, or return to the clamped one (while relieving one supercoil). If the protein enters the semi-open conformation the DNA would not be in position for religation to occur, and assuming sufficient torsional stress remains the DNA will again rotate, coming again to a transition state after which it may sample the semi-open or closed state. There exists some probability (of which we can not comment) that the system samples a different pathway from those we observe in which the topoisomerase enters a clamped conformation after release of a supercoil, and if this occurs then religation of the backbone may result, with the release of the DNA by the protein following. Alternatively it is conceivable that there exist pathways between the clamped and semi-open states which do not require the release of supercoils. In the presence of topotecan the semi-open state is replaced with a deformed conformation and the rate of religation is substantially decreased [23]. Additionally for positive supercoils the rates to the transition state are substantially increased, creating a strong inhibition of supercoil relaxation.

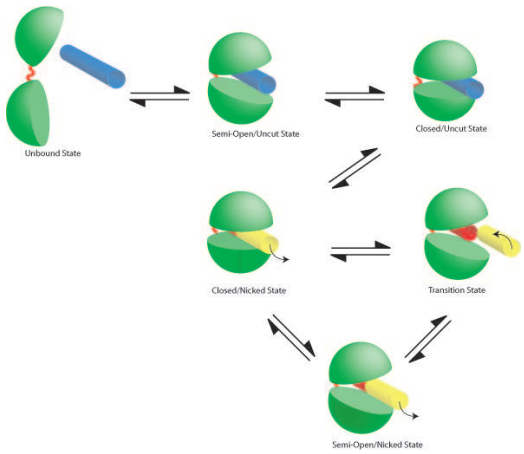


Figure 4.6: Mechanism for the relaxation of DNA supercoils by human topoisomerase I which incorporates a semi-open configuration of the protein. For the relaxation of negative supercoils the overall mechanism remains the same, however the transition and semi-open states are replaced by configurations in which the "hinge" is opened as opposed to the "lips."

BIBLIOGRAPHY

[1] K. D. Corbett and J. M. Berger. Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases. *Annual Review Of Biophysics And Biomolecular Structure*, 33:95–118, 2004.

[2] J. C. Wang. Dna topoisomerases. *Annual Review Of Biochemistry*, 65:635–692, 1996.

[3] J. B. Leppard and J. J. Champoux. Human DNA topoisomerase I: relaxation, roles, and damage control. *Chromosoma*, 114:75–85, 2005.

[4] J. J. Champoux. Dna topoisomerases: Structure, function, and mechanism. *Annual Review Of Biochemistry*, 70:369–413, 2001.

[5] J. M. Berger, S. J. Gamblin, S. C. Harrison, and J. C. Wang. Structure and mechanism of DNA topoisomerase II. *Nature*, 379:225–232, 1996.

[6] N. H. Dekker, V. V. Rybenkov, M. Duguet, N. J. Crisona, N. R. Cozzarelli, D. Bensimon, and V. Croquette. The mechanism of type IA topoisomerases. *Proceedings of the National Academy of Sciences of the United States of America*, 99:12126–12131, 2002.

[7] M. R. Redinbo, L. Stewart, P. Kuhn, J. J. Champoux, and W. G. J. Hol. Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. *Science*, 279:1504–1513, 1998.

[8] J. J. Champoux and R. Dulbecco. Activity from mammalian-cells that untwists superhelical DNA - possible swivel for dna-replication. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 69:143–&, 1972.

[9] L. Stewart, G. C. Ireton, and J. J. Champoux. The domain organization of human topoisomerase i. *Journal Of Biological Chemistry*, 271:7602–7608, 1996.

[10] L. Stewart, M. R. Redinbo, X. Y. Qiu, W. G. J. Hol, and J. J. Champoux. A model for the mechanism of human topoisomerase i. *Science*, 279:1534–1541, 1998.

[11] L. Sari and I. Andricioaei. Human topoisomerase I relaxes positive and negative DNA supercoils differently. submitted, 2005.

[12] D. A. Koster, K. Palle, E. S. M. Bot, M. A. Bjornsti, and N. H. Dekker. Antitumour drugs impede DNA uncoiling by topoisomerase i. *Nature*, 448:213–217, 2007.

[13] D. A. Koster, V. Croquette, C. Dekker, S. Shuman, and N. H. Dekker. Friction and torque govern the relaxation of dna supercoils by eukaryotic topoisomerase ib. *Nature*, 434:671–674, 2005.

[14] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.

[15] G. J. Martyna, D. J. Tobias, and M. L. Klein. Constant-pressure molecular-dynamics algorithms. *Journal Of Chemical Physics*, 101:4177–4189, 1994.

[16] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with namd. *Journal Of Computational Chemistry*, 26:1781–1802, 2005.

[17] S. Kumar, D. Bouzida, RH. Swendsen, PA. Kollman, and JM. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. *J. Comput. Chem*, 13:1011–1021, 1992.

[18] A. Grossfield. Source code from http://dasher.wustl.edu/alan/.

[19] N Foloppe and AD Mackerell. All-atom empirical force field for nucleic acids: I. parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal Of Computational Chemistry*, 21:86–104, Jan 2000.

[20] A. D. MacKerell, N. Banavali, and N. Foloppe. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, 56:257–265, 2000.

[21] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. Charmm - a program for macromolecular energy, minimization, and dynamics calculations. *Journal Of Computational Chemistry*, 4:187–217, 1983.

[22] E. Carlstein. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Annals Of Statistics*, 14:1171–1179, 1986.

[23] B. L. Staker, K. Hjerrild, M. D. Feese, C. A. Behnke, A. B. Burgin, and L. Stewart. The mechanism of topoisomerase I poisoning by a camptothecin analog. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 99:15387–15392, 2002.

[24] R. F. Frohlich, C. Veigaard, F. F. Andersen, A. K. McClendon, A. C. Gentry, A. H. Andersen, N. Osheroff, T. Stevnsner, and B. R. Knudsen. Tryptophane-205 of human topoisomerase I is essential for camptothecin inhibition of negative but not positive supercoil removal. *Nucleic Acids Research*, 35:6170–6180, 2007.

[25] Robert Zwanzig. *Nonequilibrium Statistical Mechanics*. Oxford University Press, 2001.

[26] J. Chahine, R. J. Oliveira, V. B. P. Leite, and J. Wang. Configuration-dependent diffusion can shift the kinetic transition state and barrier height of protein folding. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 104:14646–14651, 2007.

[27] N. D. Socci, J. N. Onuchic, and P. G. Wolynes. Diffusive dynamics of the reaction coordinate for protein folding funnels. *Journal Of Chemical Physics*, 104:5860–5868, 1996.

[28] J. Newman, H. L. Swinney, and L. A. Day. Hydrodynamic properties and structure of fd virus. *Journal Of Molecular Biology*, 116:593–603, 1977.

[29] Y. J. Lu, B. Weers, and N. C. Stellwagen. Dna persistence length revisited. *Biopolymers*, 61:261–275, 2001.

[30] J. F. Marko. Torque and dynamics of linking number relaxation in stretched supercoiled dna. *Physical Review E*, 76, 2007.

[31] S. Forth, C. Deufel, M. Y. Sheinin, B. Daniels, J. P. Sethna, and M. D. Wang. Abrupt buckling transition observed during the plectoneme formation of individual DNA molecules. *Physical Review Letters*, 100, 2008.

[32] T. Strick, J. F. Allemand, D. Bensimon, R. Lavery, and V. Croquette. Twisting and stretching a DNA molecule leads to structural transitions. *Biological Physics*, 487:249–270, 1999.

[33] Z. Bryant, M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante. Structural transitions and elasticity from torque measurements on dna. *Nature*, 424:338–341, 2003.

[34] J. Wereszczynski and I. Andricioaei. On structural transitions, thermodynamic equilibrium, and the phase diagram of DNA and RNA duplexes under torque and tension. *Proceedings of the National Academy of Sciences*, 103(44):16200–16205, 2006.

[35] G. Chillemi, A. Bruselles, P. Fiorani, S. Bueno, and A. Desideri. The open state of human topoisomerase I as probed by molecular dynamics simulation. *Nucleic Acids Research*, 35:3032–3038, 2007.

CHAPTER V

Directionally Negative Friction: A Method for Enhanced Sampling of Rare Event Kinetics

5.1 Introduction

Substantial interest exists in the conception of computer methods that can accelerate the simulation of conformational kinetics in a variety of condensed matter fields, spanning from material science to biomolecular chemistry, for systems that undergo rare transitions between two or more important conformational states [1, 2]. For example, many important kinetic properties of proteins, nucleic acids or their complexes can be cast in terms of paths in multi-dimensional spaces, i.e., trajectories describing the time evolution of phase points between unfolded and folded geometries, or between active and inactive states. The renewed theoretical interest in rare event kinetics is fueled, on the experimental side, by recent single molecule experiments that have started to offer the unique possibility to distinguish sub-populations, detect intermediates or describe path heterogeneity, kinetic properties which otherwise would be hidden in ensemble measurements [3, 4]. The hope is that, on the theoretical side, atomistic simulations, as *virtual* single-molecule experiments, can follow the above properties with temporal and spatial resolutions not available to the experiments.

However, while the theoretical formalism that connects the statistics of trajecto-

ries to kinetic properties is rigorous [5], its numerical implementation for complex systems is marred by severe difficulties. The simulation time often falls short by several orders of magnitude in comparison to the natural relaxation time. Free energy barriers trap the simulated trajectories inside basins of attraction. As a result, one does not sample all relevant paths. This makes the calculation of any kinetic property inaccurate.

A similar problem had been recognized to exist for thermodynamics. Because molecular simulations are confined to sample only a limited region of conformation space, ergodicity is broken [6] and thermodynamic averages are not accurate. A large set of thermodynamics-rectifying methods (see Refs. [7, 8, 9] for a review) are based on strategies that enhance sampling by facilitating the scaling of barriers in some way (e.g., by raising the temperature, modifying the potential energy, etc.). Subsequently, if the sampling strategy has a known equilibrium distribution asymptotically, states can be reweighted (by dividing the estimator by the incorrect weight and multiplying it by the correct one) to recover exact thermodynamical averages [10]. However, this is done at the expense of rendering incorrect kinetics (because trajectories are non-physical, the sequencing of dynamical events is lost).

A key question is then on whether one can devise analogous strategies to enhance *kinetic* calculations. Can one accelerate dynamical trajectories in some non-physical but computationally expedient fashion, and then find a *functional* reweight to also recover the exact kinetics? In other words, can one dilate the short computation time to the longer time that is genuine to the evolution of the real, physical system, without altering its kinetics? The method presented herein addresses one such strategy based on the concept of negative friction.

The plan of this chapter is as follows. We continue this section with a review

on several methods for the calculation of conformational kinetics. In the following section, Section 5.2, we then present the main idea behind the use of a negative frictional coefficient. Section 5.3 introduces our directionally negative friction method, and Section 5.4 presents numerical tests on prototypical model systems of increasing dimensionality. We end with a concluding discussion section.

A number of theoretical methods attempt to address rare event problems in the calculation of kinetic properties. Transition state theory (TST) [11] defines a transition state of the system according to some order parameter and initiates trajectories from that state. However, in high dimensional systems the energy landscape is exponentially dense with saddle points and it is often difficult to identify a region of conformational space corresponding to the transition state. Other approaches rely on defining a reaction coordinate along which a free energy profile (a potential of mean force) is calculated. While kinetic rates can then be calculated from this profile, for complex systems with massively heterogeneous pathways it is difficult to know a priori what the reaction coordinate is. For such complex cases there has been considerable recent advances by using transition path sampling approaches that generate ensembles of dynamical trajectories to calculate kinetic properties of conformational transitions [12, 13]. While these approaches are expected to be useful in particular for large-dimensional systems (such as proteins or nucleic acids) because they need not calculate saddle points, substantial computational effort is still required to generate a representative ensemble of paths connecting an initial and final region of phase space. Thereby the need for methods that are able to generate a multitude of conformational transition trajectories still exists. Progress along these lines is being made. For example, the stochastic difference equation method of Elber and coworkers for long-time dynamics [14], uses a minimization of a discretized version of a stochas-

tic action functional with impressive speed-up of the propagation of the equations of motion for proteins, with time steps that are two orders of magnitude longer than regular dynamics usually employs. Such approaches also have a diffusive-limit method as particular case [15, 16].

A distinct strategy is to make intelligent use of the availability of parallel processing from a large collection of CPUs. Given such resources, one can run so many independent physical trajectories in parallel that a rare event can be seen by reasonable probability in a relatively short time. This strategy was introduced in parallel-replica dynamical methods of the type developed by Voter [17] and further refined by Pande and co-workers [18], who have extended dynamical simulations of protein folding in the microsecond range.

A host of other existing methods for accelerated sampling of conformational transitions are based on the paradigm of modifying the dynamics by adding external forces that drive the conformational change of interest. Examples are the "steered" [19], "targeted" [20], or "biased" [21] molecular dynamics methods. While such approaches are valuable for inducing the transition at atomistic detail faster, in instances when large magnitude forces are needed to observe in nanoseconds what otherwise takes milliseconds, the pathways may not be characteristic for the force-free real system; this leads to incorrect kinetics. A set of recent methods recognize the need for the correction [22, 12, 23, 24]. Accordingly, trajectories generated with the modified dynamics are reweighted using weights based on either a probability functional that depends on the entire trajectory, or depending on how the modification is made, just the equilibrium distribution probability of the initial state positions [25] or momenta [26]. Also in this category of expedient modifications, an important advance for condensed matter simulations of rare events is offered by the

hyperdynamics method of Voter [27]. By adding a bias potential devised such that it raises the energy in regions other than the transition states between potential basins, transitions in hyperdynamics simulations occur at an accelerated rate. The elapsed time becomes a statistical property of the system, and rates can be calculated in principle exactly by importance sampling manipulations.

The approach that we shall focus on in the present chapter involves a strategy for modifying the dynamics to calculate kinetics inspired by an intriguing novel idea to use negative friction Langevin dynamics suggested by Chen et al. [28, 29, 30] The theoretical background for the development of our method follows in Sect. 5.2.

Conformational transitions are often accurately simulated in the framework of a Langevin dynamics, which has extensive uses in multiple areas of science as an efficient method for simulating a system in which primary degrees of freedom are coupled to others whose behavior can be modeled statistically [31]. We write the Langevin equation as a pair of coupled differential equations:

$$\begin{aligned} m\dot{\mathbf{r}}(t) &= \mathbf{p}(t) \\ \dot{\mathbf{p}}(t) &= -\gamma\mathbf{p}(t) - \nabla V(\mathbf{r}(t)) + \xi(t), \end{aligned} \tag{5.1}$$

where $V(\mathbf{r}(t))$ is a potential energy function, γ is a friction constant, $\mathbf{r}(t)$ and $\mathbf{p}(t)$ specify position and momentum, respectively, and $\xi(t)$ is a random force modeling the effects of the bath and obeying a fluctuation-dissipation relation:

$$\begin{aligned} \langle \xi(t) \rangle &= 0, \\ \langle \xi_i(t)\xi_j(t') \rangle &= 2m\gamma k_B T \delta_{ij}\delta(t-t'). \end{aligned} \tag{5.2}$$

In modeling directly with Langevin dynamics realistic systems, however, one often encounters the rare event issue: if the energy required to effect an interesting transition is many times larger than the thermal energy provided by the bath (specified

by k_BT, where k_B is Boltzmann's constant and T is the temperature), the system must absorb a large amount of heat in order to surmount the energy barrier, an event which is exponentially unlikely (in energy). Moreover, since a system obeying Equation (5.1) tends to dissipate added energy back into the bath, transition events tend to involve a long period of dwelling around a stable minimum before a crossing of the barrier is observed. Therefore, events of interest often occur over time scales that are inaccessible to straightforward simulation.

For biomolecular processes such as folding or other major conformational rearrangements, of particular interest are computational studies of transition events in which the initial and final states can be specified by the forward reaction rate, $k_{\mathcal{R}\to\mathcal{P}}$, of transition from one region to another. $k_{\mathcal{R}\to\mathcal{P}}$ can be found from the time derivative of a trajectory average:

$$C(t) \equiv \frac{\langle h_{\mathcal{R}}(\mathbf{r}_0) h_{\mathcal{P}}(\mathbf{r}_t)\rangle_0}{\langle h_{\mathcal{R}}(\mathbf{r}_0)\rangle_0} \tag{5.3}$$

where $\langle \cdots \rangle$ indicates an average over an ensemble of trajectories initiated from an equilibrium distribution in phase space, and $h_{\mathcal{R},\mathcal{P}}$ are container functions given by:

$$h_{\mathcal{R},\mathcal{P}}(\mathbf{r}) = \begin{cases} 1 & \text{if } \mathbf{r} \in \mathcal{R}, \mathcal{P} \\ 0 & \text{otherwise,} \end{cases} \tag{5.4}$$

For times longer than an initial transient period but shorter than the equilibration time of the system, the derivative of $C(t)$ (i.e., the reactive flux) displays a plateau corresponding to $k_{\mathcal{R}\to\mathcal{P}}$ [32]. In transition path sampling, the time-correlation functions can be calculated by an advantageous factorization of $C(t)$ [13] that reduces considerably the computational cost compared to the original formulation, but it requires umbrella sampling in phase space that still requires investing significant numerical effort.

While transition path sampling is providing a powerful strategy and a useful theoretical paradigm, it would still be useful to be able to generate directly and effectively a multitude of dynamical trajectories representative of the conformational transition of interest. In principle, $C(t)$ could be computed directly from an ensemble of trajectories initiated from the reactant region, but for low-temperature systems and/or rugged energy landscapes this approach is inefficient, as very few trajectories will reach the product region.

A few methods have been devised to overcome this problem and to allow direct sampling of trajectories. Tully and coworkers [25] have proposed a convenient bias function of general applicability for systems with large enthalpic barriers. They apply a "puddle" potential (used previously by the same group to enhance thermodynamic averaging [33]) that raises the potential energy well from which the trajectories are initiated, so that escape over the barrier is more frequent. Proper reweighting of the initial Boltzmann distribution of points allows for an exact expression for time-correlation functions and hence for kinetic observables. However, while the probability to leave the reactant basin is enhanced by puddling, the beneficial effect of puddling is washed-out in high dimensions by the fact that motion is accentuated in the vastitude of all directions. In previous work, we have conceived a related method that skews the Maxwellian momenta distribution only along preferential directions that are relevant for the conformational transition [26] in such a way that the important transitions are sampled more frequently; proper reweighting by the relative probability functional between the skewed and physical trajectories leads to the exact recovery of kinetic properties.

Recently, Chen et al. [34, 28, 29] have presented an intriguing new approach to the rare event problem in which the Brownian dynamics are altered by the introduc-

tion of a *negative* frictional value γ. In contrast to positive, physical friction which dissipates energy, negative friction has the effect of pumping energy into the system to increase the probability of transitions over barriers and decreasing the amount of time in stable minima. By reweighting these trajectories it is possible to extract accurate transition probabilities with much greater efficiency than with using a physical, positive frictional Langevin equation. While the method has been applied with success to one- or two-dimensional systems, a major limitation of this approach (just like in the case of "puddling") is that by adding energy in all dimensions, transitions of interest in high-dimensional systems are rarely sampled. In fact in this chapter we show that a system with as few as 10 dimensions may spend excessive amounts of time exploring degrees of freedom orthogonal to the reaction coordinate, creating an inefficient approach to most systems of interests. We propose a hybrid approach, in which negative frictional dynamics are propagated along a reaction coordinate of interest and positive frictional dynamics are calculated along perpendicular degrees of freedom. With a proper trajectory reweighting scheme that we propose, we may obtain accurate trajectory averages which correspond to their physical positive frictional counterparts. In principal this method may be extended to systems of arbitrary number of dimensions, making it a useful tool for studying complex systems of interest, such as biomolecules, in which there may be thousands of degrees of freedom and a transition of interest occurs along a low-dimensional reaction coordinate manifold.

5.2 Background

We consider a stochastic process (such as Langevin dynamics) that evolves according to a delta-correlated Gaussian random force, $\xi(t)$. The probability of observing

a given realization of the random force, i.e., a particular noise history $P[\xi(t)]$ along the trajectory, is specified by:

$$P[\xi(t)] = C \exp[\frac{-\beta}{4m\gamma} \int_0^t dt' \xi^2(t')] \tag{5.5}$$

where C is a normalization constant and the path integral extends over the total time of the trajectory.

We can therefore write the ensemble average of any functional observable $A[\xi(t)]$ (i.e., of an observable that depends on the entirety of the trajectory) as a weighted sum over all realizations of the noise history $\xi(t)$:

$$\langle A[\xi(t)] \rangle = \int \mathcal{D}\xi A[\xi] P[\xi]. \tag{5.6}$$

Utilizing the dynamical connection between the random force, $\xi(t)$ and the trajectory $\mathbf{r}(t)$, this average can be rewritten as a sum over trajectories. In the case of Langevin dynamics as specified by Equation (5.1), we obtain

$$\langle A[\mathbf{r}(t)] \rangle = \frac{\int \mathcal{D}\mathbf{r} J[\mathbf{r}] A[\mathbf{r}] \exp(\frac{-\beta S[\mathbf{r}(t)]}{4m\gamma})}{\int \mathcal{D}\mathbf{r} J[\mathbf{r}] \exp(\frac{-\beta S[\mathbf{r}(t)]}{4m\gamma})} \tag{5.7}$$

where $J[\mathbf{r}] = \left|\frac{\partial \xi}{\partial \mathbf{r}}\right|$ is the Jacobian connecting the measures $\mathcal{D}\xi$ and $\mathcal{D}\mathbf{r}$ and $S[\mathbf{r}(t)] = \int_0^t dt' [m\ddot{\mathbf{r}}(t') + \gamma \dot{\mathbf{r}}(t') + \nabla V(\mathbf{r}(t'))]^2$ is the "action" functional [35].

We note that the average in Equation (5.7) can be written, without approximation, as

$$\begin{aligned} \langle A[\mathbf{r}(t)] \rangle &= \frac{\int \mathcal{D}\mathbf{r} J'[\mathbf{r}] \exp(\frac{-\beta S'}{4m\gamma}) J_w[\mathbf{r}] A[\mathbf{r}] \exp(\frac{-\beta(S-S')}{4m\gamma})}{\int \mathcal{D}\mathbf{r} J'[\mathbf{r}] \exp(\frac{-\beta S'}{4m\gamma}) J_w[\mathbf{r}] \exp(\frac{\beta(S-S')}{4m\gamma})} \\ &= \frac{\langle J_w[\mathbf{r}] A[\mathbf{r}] \exp(\frac{-\beta(S-S')}{4m\gamma}) \rangle'}{\langle J_w[\mathbf{r}] \exp(\frac{-\beta(S-S')}{4m\gamma}) \rangle'}, \end{aligned} \tag{5.8}$$

where $J_w[\mathbf{r}] \equiv \frac{J[\mathbf{r}]}{J'[\mathbf{r}]}$, S' and $J'[\mathbf{r}]$ are the action and Jacobian functionals, respectively, for some altered dynamical scheme, $\langle \cdots \rangle'$ indicates a trajectory average using the altered dynamics, and we have omitted the arguments to S and S' for convenience.

Equation (5.8) is in principle exact, but in practice we will achieve an accurate average only if trajectories found from realizations of the altered dynamics overlap sufficiently with trajectories from the unbiased dynamics. It would be infeasible, for instance, to calculate an average in a system obeying conventional Langevin dynamics by generating an ensemble of negative-friction trajectories and then rescaling as in Equation (5.8): such trajectories are fantastically unlikely according to positive-friction dynamics. Ideally, we wish to preferentially sample the paths which are most significant in the average of A.

With this in mind, we note that if the evolution of our system is Markovian for all t and for all dynamical schemes considered, the probability of a full trajectory is simply the product of the probabilities of any series of trajectory subsections. Moreover, in general these subsections can evolve according to different dynamical schemes. Therefore, following Chen, we consider trajectories of length t that evolve up to an intermediate time $t_m < t$ according to negative-friction dynamics and then switch to conventional Langevin dynamics for the remaining time $t - t_m$.

Such trajectories correspond to significant paths: transition events in systems with large barriers are typically characterized by a rapid increase in the energy of the system as the barrier is surmounted, followed by a decay back to energies on the order of $k_B T$.

In Chen's implementation, the altered dynamics are propagated according to the negative-friction Langevin equation,

$$m\ddot{\mathbf{r}}(t) = m\gamma\dot{\mathbf{r}}(t) - \nabla V(\mathbf{r}(t)) + \xi(t) \tag{5.9}$$

and the difference in action functionals between the negative and positive-friction

dynamics is path-independent [28]:

$$S' - S = 4m\gamma(H_2 - H_1) \tag{5.10}$$

where H_1 and H_2 are the initial and final Hamiltonians of the system, respectively.

The Jacobian term can be solved analytically [36] and comes to:

$$\begin{aligned} J_w[\mathbf{r}] \equiv |\frac{\mathcal{D}\xi}{\mathcal{D}\mathbf{r}}| \cdot |\frac{\mathcal{D}\xi'}{\mathcal{D}\mathbf{r}}|^{-1} &= |\frac{\mathcal{D}(\ddot{\mathbf{r}} + m\gamma\dot{\mathbf{r}} + \nabla V(\mathbf{r}(t))}{\mathcal{D}\mathbf{r}}| \\ &\times |\frac{\ddot{\mathbf{r}} - m\gamma\dot{\mathbf{r}} + \nabla V(\mathbf{r}(t))}{\mathcal{D}\mathbf{r}}|^{-1} = e^{n\gamma t_m} \end{aligned} \tag{5.11}$$

with n being the dimensionality of the system. In straightforward applications when t_m is constant, this term may be excluded since it is constant and does not contribute to the average in Equation (5.8) (but see below).

The all-negative friction method of Chen et al. is effective in surmounting barriers in low-dimensional systems, but it is not efficient in simulations containing many degrees of freedom. Transition events in large systems tend to involve the motion of collective variables that are composed primarily of only a subset of the degrees of freedom in the system; trajectories generated as a result of pumping energy into all degrees of freedom, like those generated by Chen's method, are extremely unlikely. It is therefore of interest to consider schemes that can introduce energy in specific subsets of the overall conformational space of the system.

5.3 Directionally Negative Friction: The Method

5.3.1 Fundamental idea

If we consider a Langevin equation with a generalized friction matrix,

$$m\ddot{\mathbf{r}}(t) = -m\Gamma\dot{\mathbf{r}}(t) - \nabla V(\mathbf{r}(t)) + \xi(t) \tag{5.12}$$

where Γ is a diagonal matrix whose entries determine the friction coefficients for each degree of freedom, it is a simple matter to selectively introduce energy into the system

by choosing certain diagonal entries in Γ to be negative. An exact average can then be recovered using Equation (5.8). However, this is still a clumsy approach, as in most systems the potential $V(\mathbf{r})$ will be defined by variables (say, atomic positions in Cartesian space) that do not in themselves effectively define the transitions of interest. If the motion of one or more of these variables is governed by negative-friction dynamics, we are still unlikely to sample interesting pathways, as energy will be introduced along directions that may have little to do with our desired reaction coordinate.

A more productive approach would be to identify a normalized guiding vector $\hat{\mathbf{e}}_0$ in configuration space that points along the direction of desired motion, and apply negative friction dynamics only along that direction - that is, to transform Equation (5.12) to a coordinate system in which $\hat{\mathbf{e}}_0$ is a basis vector, and then set the corresponding diagonal entry in Γ to $-\gamma$. We would thereby avoid introducing energy into directions perpendicular to our desired path and, as such, we would generate more probable trajectories. However, such a transformation is computationally too demanding to be efficient because it would require coordinate transformations (rotations) at each step. Instead, we introduce an additional term into the Langevin equation:

$$m\ddot{\mathbf{r}}(t) = -m\gamma\dot{\mathbf{r}}(t) + 2m\gamma(\dot{\mathbf{r}}(t)\cdot\hat{\mathbf{e}}_0)\hat{\mathbf{e}}_0 - \nabla V(\mathbf{r}(t)) + \xi(t) \tag{5.13}$$

The dot product in the additional term renders friction negative (i.e., switches it from γ to $-\gamma$) in the direction parallel to $\hat{\mathbf{e}}_0$, but leaves it positive in all the other directions. We note that applying a similarity transformation to Equation (5.13) for which $\hat{\mathbf{e}}_0$ is a basis vector in the transformed system does indeed produce an equation in the form of Equation (5.12), which was our objective.

In contrast to the reweighting scheme by Chen et al. [28] using Equation (5.10),

the difference in action between Equation (5.13) and the conventional Langevin equation is path-dependent. Specifically,

$$\begin{aligned} S' - S &= \int_0^t d\tau([m\ddot{\mathbf{r}}(\tau) + m\gamma\dot{\mathbf{r}}(\tau) - 2m\gamma(\dot{\mathbf{r}}(\tau)\cdot\hat{\mathbf{e}}_0)\hat{\mathbf{e}}_0 \\ &\quad +\nabla V(\mathbf{r}(\tau))]^2 - [m\ddot{\mathbf{r}}(\tau) + m\gamma\dot{\mathbf{r}}(\tau) + \nabla V(\mathbf{r}(\tau))]^2) \end{aligned} \tag{5.14}$$

and after some algebra, we find

$$S' - S = -4m\gamma \int_0^t d\tau(\dot{\mathbf{r}}(\tau)\cdot\hat{\mathbf{e}}_0)(\nabla V(\mathbf{r}(\tau))\cdot\hat{\mathbf{e}}_0 + m\ddot{\mathbf{r}}(\tau)\cdot\hat{\mathbf{e}}_0), \tag{5.15}$$

where the expression on the right hand side must be evaluated numerically along each trajectory. We can then determine the average of an observable $A[\mathbf{r}(t)]$ by substituting this expression into Equation (5.8). The Jacobian may still be derived as in Equation (5.11) with d this time being set to the number of degrees of freedom defined by $\hat{\mathbf{e}}_0$ that get assigned negative friction (in this work $d = 1$, but in more complex cases several directions describing large-scale cooperative motions might be included).

5.3.2 Identifying $\hat{\mathbf{e}}_0$

Complex molecules (such as proteins or nucleic acids) have a dynamical evolution in which fast oscillatory modes are coupled to slowly varying ones. While the *actual* time spent by an *ensemble* of molecules in the fast manifold is the same as that spent in the slow manifold, the *computer* time needed for convergence of properties in the slow manifold (when simulating a single molecule) is much larger than that spent in the fast manifold. Therefore, identifying the slow manifold and accentuating the motion along it is a good strategy for enhanced sampling of conformational transitions. This is because in most cases motion along the slow manifold includes the largest conformational changes (reactions), which are often of primary interest. [37, 38]

In our previous work aimed at using slow manifold dynamics to enhance the calculation of thermodynamic [39] or kinetic properties [26], we showed that we can identify the slow manifold by calculating $\mathbf{p}_s$, the average of the momentum, $\mathbf{p}$, over t time units,

$$\hat{\mathbf{e}}_0 = \frac{\mathbf{p}_s}{\|\mathbf{p}_s\|}, \text{with } \mathbf{p}_s = \frac{1}{t}\int_0^t \mathbf{p}(\tau)d\tau \tag{5.16}$$

In order for $\hat{\mathbf{e}}_0$ to point along the slow manifold (i.e., for the components of momentum in the fast manifold to average to zero), one has to choose the averaging time t to be several times larger than the period of the fast modes but smaller than those of the slow modes.

Alternatively, $\hat{\mathbf{e}}_0$ can be found from either a normal-mode or a quasiharmonic-mode decomposition [40] by solving an eigenvalue-eigenvector problem,

$$\hat{\mathbf{e}}_0 = \min_{\lambda}\{\hat{\mathbf{e}} \mid F\hat{\mathbf{e}} = \lambda\hat{\mathbf{e}}\} \tag{5.17}$$

and choosing the lowest-eigenvalue eigenvector obtained from diagonalizing F, which is either the Hessian of the potential (in the case of the harmonic modes) or the inverse of the covariance matrix of atomic fluctuations (in the case of the quasiharmonic modes). The normal mode decomposition can be performed only once for a minimized structure in the reactant well, and $\hat{\mathbf{e}}_0$ aligned along a low-frequency mode (or a linear combination of several such slow modes). The quasiharmonic calculation can be performed, also only once, on the same trajectory that was used to generate the initial distribution of the starting conformations; again, a combination of slow quasiharmonic modes can be used.

Using the momentum-averaging scheme, we showed that slow-mode directions are of promise for enhanced sampling and the exploration of large conformational changes, and have provided a hybrid MC scheme to obtain exact thermal equilibrium

properties [39]. For the case of the alternative choice of $\hat{\mathbf{e}}_0$ in Equation (5.17), normal mode analyse have provided considerable insight into the nature of motion in many proteins [41, 42, 43, 44, 45]. They have demonstrated that, in systems where an initial and a final structure are available, the first few low frequency modes are often sufficient to describe the large-scale conformational changes involved in going from one structure to the other [46, 47], particularly when collective motions are involved [48]. This strategy has worked well both for protein-DNA complexes [49] and for systems as large as the ribosome [50].

In our directional negative-friction Langevin scheme, $\hat{\mathbf{e}}_0$ can be calculated around the initial minimum. Alternatively, if the system visits several stable minima, the trajectory can be constructed from a series of activation/dissipation events, with each successive $\hat{\mathbf{e}}_0$ found from momentum averaging during dynamics in each minimum. In this context, the trick in Equation (5.13) becomes reminiscent of the activation-relaxation technique [51], that inverts the component of the force vector parallel to the displacement from the current position to the local minimum, while leaving it unmodified in all other directions. However, a straightforward adaptation of negative friction along the displacement direction would still not work if the direction from the local minimum to the current position does not coincide with $\hat{\mathbf{e}}_0$. A connection is also worthwhile making here to the self-guided Langevin dynamical scheme proposed by Wu et al. [52], for which the added self-averaging guiding term is basically akin to negative friction, with the difference that the injected energy is then removed by positive friction, so the formal similarity to directionally negative friction becomes unclear.

Although the dynamics specified by Equation (5.13) are unphysical, we emphasize that actual systems tend to escape deep minima via a rapid increase in energy

along a few pertinent coordinates followed by a frictional dissipation. Therefore, the transition paths found from our directional negative friction method are likely to be among the probable paths connecting various minima. The bundle of such generated transition paths can be used as is, can be refined by the subsequent application of more sophisticated transition path sampling techniques [53], or can serve as the starting ensemble for techniques that identify optimal reaction pathways [54, 55, 56, 57, 58, 59, 60, 61].

5.3.3 Efficiency tuning by adaptive choice of t_m

As an additional enhancement, t_m need not be set to a specific value. Rather, if information of the transition state is known, then the frictional coefficient may be switched from negative to positive as the trajectory enters the product region, i.e., adaptively. Since the time that trajectories spend in negative-friction dynamics is not consistently the same for all trajectories, the Jacobian $J_w[\mathbf{r}]$, now trajectory-dependent and no longer constant over the entire trajectory bundle, may not be neglected in calculations. In practice, Equation (5.11) must now also be used when calculating the reweighting factors in Equation (5.8). This may prove useful in a variety of applications, for example if a rotatable bond is the reaction coordinate of interest, then friction may be changed from negative to positive when the dihedral transitions from *cis* to *trans*, or in a system such as alanine dipeptide where a reaction coordinate is the $N - H \cdots O = C$ distance, with stable minim at a short and a long distance, the dynamics may be altered when the reaction coordinate becomes far from the minimum in which it started.

5.4 Model System Examples

As a test of our method, we consider the motion of a Brownian particle on a potential surface composed of a double well in the y direction, with $n-1$ identical perpendicular "udder" [62, 63] potentials:

$$V_n(\mathbf{x}, y) = y^2(y-a)^2 + \sum_{i=1}^{n-1}(b(x_i - c)^2 - de^{-fx_i^2} + g) \tag{5.18}$$

We choose $a = 2$, $b = 0.1$, $c = 2$, $d = 0.75$, $f = 50$, and $g = 0.35$, so that motion along y, in the double well, is slow and the perpendicular degrees of freedom are tightly bound around $x_i = 0$, with a large, relatively flat basin at higher values of x_i. The contrast between low and high-dimensionality is showcased by setting either $n = 2$ or 10, i.e., having either one or nine perpendicular degrees of freedom, thereby modeling a two- or, respectively, ten-dimensional system. Figure 5.1 depicts a plot of the double well component of V_n and one of the udder potentials as a function of distance. The double well has a local maximum at $y = 1$, with a barrier height $V(\mathbf{x} = 0, y = 1) = 1$ and stable minima at $y = 0$ and $y = 2$. Each udder potential has a local maximum at $x_i = 0.288$, with a barrier height of 0.631. We consider the region for which $y < 1$ and $x_i < 0.288$ for all i to be our reactant region, and the region for which $y > 1$ and $x_i < 0.288$ for all i to be our product region.

Although this system does not correspond to any physical system of interest, it possesses qualitative properties that correspond to a higher-dimensional, realistic system: namely, the configuration space is composed largely of uninteresting regions in which transition to the product state is nearly impossible. In our system, these regions are the states in which $x_i > 0.288$ for one or more i; the long, flat portion of the udder potential serves as a crude approximation for the regions of a higher-dimensional potential that are not of interest. This test potential captures one of

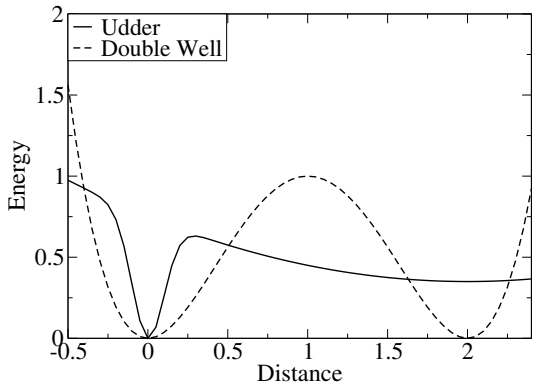


Figure 5.1: Potential energy surface for model system. The direction of interest, y, has a double well potential with a transition state at $y = 1$ of energy 1 kcal/mol. Orthogonal directions, x_i, are represented by the udder potential with a transition state at $x_i = .288$ of energy 0.631 kcal/mol. For a large number of dimensions equally pumping energy into each dimensional, as is done in the Chen negative-friction method, will results in trajectories predominantly sampling the uninteresting states x_i, while with a Direction negative-friction model energy may be added only to the direction of interest, in this case y.

the crucial problems of Chen's negative friction method when applied to higher-dimensional systems: the negative-friction dynamics will tend to drive the system randomly into regions of conformational space for which no transition is possible. In the case of V_n, the negative friction dynamics will drive the system out of the well around $\mathbf{x} = 0$, so that most trajectories will terminate in regions for which $h_P = 0$ and will not contribute to the average in Equation (5.3).

We suppose the dynamics of this system to be specified by the Langevin equation, Equation (5.1), with $\gamma = 1.0$, and we take a time step $\delta t = 0.01$ for our numerical integration. For the variable t_m method we choose the time t_m to switch the sign of γ for each trajectory (i.e., we chose the transition from negative to positive friction dynamics to occur) when the trajectory first enters the product region, that is when the coordinates become such that $y > 1$ and $x_i < 0.288$ for all i.

Trajectories are initiated from an equilibrium distribution in the reactant region and propagated according to either the Chen all-negative friction implementation, our directional negative-friction equation with $t_m = 1$ (Equation (5.13)), our direc-

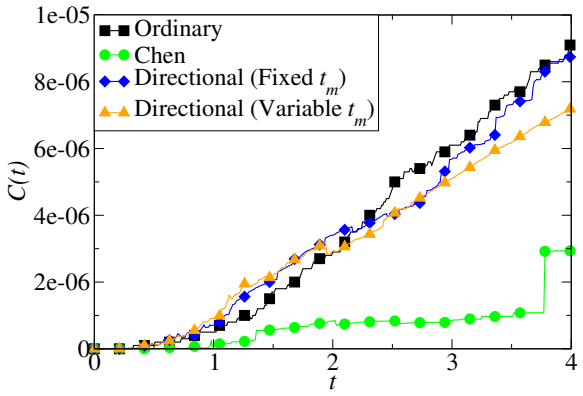
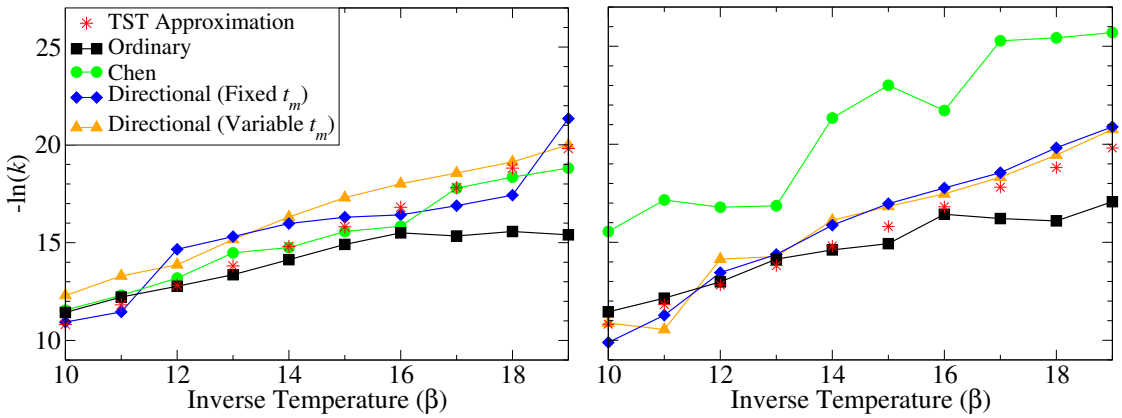


Figure 5.2: Correlation functions for creation of product as given by Equation (5.3). All calculations are for 10D potentials, $\beta = 12$, 10^7 trajectories, and the Chen and Directional methods both have $t_m = 1$. Reweighting of Directional methods accurately recreates the ordinary method which uses only positive frictional terms, while reweighting of the Chen implementation does not, as trajectories are lost to sampling of degrees of freedom orthogonal to that which is of interest.

tional negative-friction equation with variable t_m, or an unadorned Langevin equation (the ordinary approach). We then calculate the time correlation function $C(t)$ (see Figure 5.2), as specified in Equation (5.3), as a weighted average of $h_P[\mathbf{r}(t)]$ over all trajectories beginning in the product region using Equation (5.8). After discarding the portion of the correlation function corresponding to the transient time, the rate constant, k, is calculated from a least-squares regression of $C(t)$.

Figures 5.3(a) and 5.3(b) display $\ln k$ as a function of the inverse temperature β, i.e., the Arrhenius plot, for systems with $n = 2$ and $n = 10$ respectively. Simulations with a direct approach of calculating the ratio in Equation (5.3) using 10^7 trajectories are compared with calculations of 10^4 trajectories for the Chen implementation, our directional method with a fixed t_m, our directional method with a variable t_m, and "exact" values calculated using transition state theory.

While it is recognized that TST does not provide an exact rate but only an upper bound, and, as a result, that a systematic error can appear when comparing to it, TST is chosen as a reference state since the basic assumptions of the model (no



(a) (b)

Figure 5.3: Arrhenius plot for the two dimensional model system in Equation (5.18) (1 y and 1 x dimension) (a.). The ordinary approach predicts the rate constant accurately for higher temperatures but for lower ones begins to perform poorly as few trajectories make the transition, while the Chen and two directional implementations all perform well in reproducing the results from Transition State Theory. Arrhenius plot for the ten dimensional model system (1 y and 9 x dimensions) (b.). The ordinary and directional results perform similarly to their 2D counterparts, but the Chen method drastically underestimated the rate constant due to few trajectories moving to the product region because of the increase in energy that was introduced to the orthogonal degrees of freedom. For both the 2D and 10D data 10^7 trajectories were used for the ordinary approach while 10^4 were used for the Chen and directional approaches.

barrier re-crossings) should hold quite well especially in the low temperature case, which is the region we are most interested in. In both the case of $n = 2$ and $n = 10$ the direct method yields results for the barrier height that are higher than the TST approximation at high temperatures and lower than it at low temperature. For high temperatures, this is expected (due to the fact that the assumptions of TST break down) as TST gives an upper bound for the rate, therefore under-estimates the barrier; the choice of TST therefore results in a systematic error that is not reflective of the quality of the proposed method (which is exact in practice, but is designed to work for temperatures that are low compared to barrier heights). For them, i.e., for low temperatures, the error is statistical and is simply due to a lack of convergence in the numerical results because of insufficient sampling.

In other words, because the derivation of our method does not imply any approximation and is supposed to be exact in practice, insufficient sampling should give an unbiased error (variance error), and not a systematic error (i.e., a bias). Systematic deviations might be due to the fact that the comparison should not have been to TST, but to the exact, converged value. Systematic error is indeed expected because of the TST breakdown at high temperature. At low temperature, however, the TST approximation is much better so the significant error is the statistical error. It is important to emphasize that the method is designed (and works optimally) for rare event transitions that are difficult to simulate directly. Another method designed for this purpose and closely related is the puddle jumping method of Tully et al. [25], it is also employing TST as a reference to showcase its efficiency.

In the 2D case the Chen method performed the best, with an error of 3.39% relative to the TST results (compared to 6.29% and 7.78% for the fixed and variable t_m simulations). However in the higher-dimensional 10D case the Chen method performed poorly, by drastically underestimating the rate constant for all temperatures as simulations explored degrees of freedom perpendicular to the chosen reaction coordinate, giving an error of 36.77% relative to the TST result. In contrast, our new directional method performed quite well; while slightly underestimating rate constants relative to TST reference, the errors of 6.08% and 5.40% are on the same order as what is calculated for 1000 times as many trajectories with the direct approach (5.90%). The presence of a variable t_m allows more trajectories to sample the product region: in simulations with $\beta = 16$, 66.5% of trajectories went to product, while with a fixed t_m only 2% did; in stark contrast, in the direct implementation only $6 \cdot 10^{-5}$% only made it into the product region. While the error in the Chen implementation may be reduced by sampling additional trajectories, we found that

	TST	Ordinary	Chen	Directional	Directional/Variable t_m
$2D$	1.00	0.48	0.82	0.91	0.87
$10D$	1.00	0.62	1.26	1.19	1.11

Table 5.1: Energy barriers for the model system as calculated from the slope of the Arrhenius plots in Figure 5.3. For the ordinary case the calculated barriers are well off from their actual value due to the poor sampling of transitions at low temperature. The calculated barriers for the Chen implementation became less accurate as the number of dimensions increased while the accuracy of the directional implementations was not significantly affected by the increase in the number of dimensions.

even for 10^7 trajectories it did not perform as well as the direct method, which can be attributed to wandering off in the perpendicular manifold.

By fitting a linear regression to the Arrhenius plot the energy barrier of the transition state may be extrapolated. Since this has been defined to be 1.00 in our model (the units are such that $k_{\mathrm{B}} = 1$ is dimensionless), this serves as another comparison between methods, and the results are shown in Table 5.1. The Chen method worked decently in the 2D case but, not surprisingly, decreased in accuracy for the 10D case. Both directional methods performed well, with the directional/variable t_m giving an energy barrier of 1.11 in the 10D case, compared to 1.26 for the Chen method. The unadorned Langevin ordinary method provided quite poor estimates of the energy barrier due to the low number of transitions observed at low T.

Overall, the tests on the potential energy model of increasing dimensionality proved that, while negative friction is instrumental in injecting energy that can lead to enhanced sampling of conformational transitions, and that proper reweighting recovers physically-sound kinetics, the beneficial effects are exponentially diminished as the number of negatively frictioned degrees of freedom increases. Instead, if the direction of progress for the relevant conformational chance is identified, the directionally negative friction method can focus energy pumping so that relevant pathways are generated effectively; as in the original negative friction method, reweigthing done based on stochastic path integral probabilities recovers the exact value of time-dependent,

kinetic observables.

5.5 Concluding Discussion

The problem of sampling rare events trajectories, a problem of both significant interest and considerable difficulty, is caused, at its core, by free energy barriers that render the computer-generated search in conformational space non-ergodic. As a consequence, there are two problems that computer simulations face, and correspondingly, two tiers of methods designed to overcome them. The first problem is that of thermodynamical averaging, which has customarily been solved by importance sampling manipulations that produce samples from an unphysical but convenient ensemble, and then reweight them according to the proper relative Boltzmann weight. The second tier, addressed in the present chapter, has to do with kinetics, for which both trajectory generation and reweighting are substantially more complex. We have introduced a modification to the Langevin equation which may be used to introduce energy into specific degrees of freedom and increase the rate of observing rare events for a system. The trajectories created with this method may be reweighted to obtain dynamical information about the system as if it were obeying unbiased dynamics, thereby offering the possibility for exact calculations of kinetic properties for the system undergoing the physical, Langevin dynamics. Positive friction dynamics are propagated for an amount of time t_m which may be fixed or, if information about the transition state is known, may be varied so that a trajectory obeys positive friction dynamics while in a reactant state and negative friction dynamics while in a product state. Unlike previous negative friction dynamics implementations, this one does not introduce energy into all dimensions of the system, thus its performance does not degrade as the dimensionality of the system increases and it may therefore be used

for the study of complicated system such as biomolecules.

This work has been published in The Journal of Chemical Physics [64].

BIBLIOGRAPHY

[1] A. F. Voter, F. Montalenti, and T. C. Germann. Extending the time scale in atomistic simulation of materials. *Annual Review of Materials Research*, 32:321–346, 2002.

[2] R. Elber. Long-timescale simulation methods. *Current Opinion In Structural Biology*, 15:151–156, 2005.

[3] S. Weiss. Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nature Structural Biology*, 7:724–729, 2000.

[4] C. Bustamante, J. C. Macosko, and G. J. L. Wuite. Grabbing the cat by the tail: Manipulating molecules one by one. *Nature Reviews Molecular Cell Biology*, 1:130–136, 2000.

[5] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annual Review of Physical Chemistry*, 53:291–318, 2002.

[6] R.G. Palmer. Broken ergodicity. *Adv. Phys.*, 31:669–735, 1982.

[7] B.J. Berne and J.E. Straub. Novel methods of sampling phase space in the simulation of biological systems. *Curr. Opin. Struc. Biol.*, 7:181–189, 1997.

[8] K. Tai. Conformational sampling for the impatient. *Biophysical Chemistry*, 107(3):213–220, Feb 2004.

[9] S. A. Adcock and J. A. McCammon. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chemical Reviews*, 106(5):1589–1615, May 2006.

[10] G. M. Torrie and J. P. Valleau. Non-physical sampling distributions in Monte-Carlo free-energy estimation - umbrella sampling. *Journal of Computational Physics*, 23:187–199, 1977.

[11] DG Truhlar, BC Garrett, and SJ Klippenstein. Current status of transition-state theory. *Journal of Physical Chemistry*, 100:12771–12800, Aug 1996.

[12] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler. Transition path sampling and the calculation of rate constants. *Journal of Chemical Physics*, 108:1964–1977, 1998.

[13] C. Dellago, P. G. Bolhuis, and D. Chandler. On the calculation of reaction rate constants in the transition path ensemble. *Journal of Chemical Physics*, 110:6617–6625, 1999.

[14] R. Elber, J. Meller, and R. Olender. Stochastic path approach to compute atomically detailed trajectories: Application to the folding of C peptide. *Journal of Physical Chemistry B*, 103:899–911, 1999.

[15] A. K. Faradjian and R. Elber. Computing time scales from reaction coordinates by milestoning. *Journal of Chemical Physics*, 120:10880–10889, 2004.

[16] D. Moroni, P. G. Bolhuis, and T. S. van Erp. Rate constants for diffusive processes by partial path sampling. *Journal of Chemical Physics*, 120:4055–4065, 2004.

[17] A. F. Voter. Parallel replica method for dynamics of infrequent events. *Physical Review B*, 57:R13985–R13988, 1998.

[18] M. R. Shirts and V. S. Pande. Mathematical analysis of coupled parallel simulations. *Physical Review Letters*, 86:4983–4987, 2001.

[19] B. Isralewitz, J. Baudry, J. Gullingsrud, D. Kosztin, and K. Schulten. Steered molecular dynamics investigations of protein function. *Journal of Molecular Graphics & Modelling*, 19:13–25, 2001.

[20] J. Schlitter, M. Engels, and P. Kruger. Targeted molecular-dynamics - a new approach for searching pathways of conformational transitions. *Journal of Molecular Graphics*, 12:84–89, 1994.

[21] E. Paci and M. Karplus. Forced unfolding of fibronectin type 3 modules: An analysis by biased molecular dynamics simulations. *Journal of Molecular Biology*, 288:441–459, 1999.

[22] O. Mazonka, C. Jarzynski, and J. Blocki. Computing probabilities of very rare events for langevin processes: a new method based on importance sampling. *Nuclear Physics A*, 641:335–354, 1998.

[23] D. M. Zuckerman and T. B. Woolf. Dynamic reaction paths and rates through importance-sampled stochastic dynamics. *Journal of Chemical Physics*, 111:9475–9484, 1999.

[24] C. Xing and I. Andricioaei. On the calculation of time-correlation functions by potential scaling. *J. Chem. Phys.*, 124:034110, 2006.

[25] S. A. Corcelli, J. A. Rahman, and J. C. Tully. Efficient thermal rate constant calculation for rare event systems. *Journal of Chemical Physics*, 118:1085–1088, 2003.

[26] J. MacFadyen and I. Andricioaei. A skewed-momenta method to efficiently generate conformational-transition trajectories. *Journal of Chemical Physics*, 123:074107, 2005.

[27] A. F. Voter. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Physical Review Letters*, 78:3908–3911, 1997.

[28] L. Y. Chen. Brownian motion in a two-dimensional potential: A new numerical solution method. *International Journal of Modern Physics B*, 16(24):3643–3654, Sep 2002.

[29] L. Y. Chen, S. C. Ying, and T. Ala-Nissila. Finding transition paths and rate coefficients through accelerated langevin dynamics. *Phys. Rev. E*, 65(4):042101, Apr 2002.

[30] L. Y. Chen and P. L. Nash. Transition path sampling with a one-point boundary scheme. *Journal of Chemical Physics*, 119:12749–12752, 2003.

[31] N.G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland, 1992.

[32] D. Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, 1987.

[33] J. A. Rahman and J. C. Tully. Puddle-jumping: a flexible sampling algorithm for rare event systems. *Chemical Physics*, 285:277–287, 2002.

[34] L. Y. Chen and S. C. Ying. Solution of the langevin equation for rare event rates using a path-integral formalism. *Phys. Rev. B*, 60(24):16965–16971, Dec 1999.

[35] L. Onsager and S. Machlup. Fluctuations and irreversible processes. *Physical Review*, 91:1505–1512, 1953.

[36] H. Kleinert. *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics and Financial Markets*. World Scientific, 3rd edition, 2004.

[37] A. Amadei, A.B.M. Linssen, and H.J.C. Berendsen. Essential dynamics of proteins. *Proteins*, 17:412–425, 1993.

[38] M.A. Balsera, W. Wriggers, Y. Oono, and K. Schulten. Principal component analysis and long time protein dynamics. *J. Phys. Chem.*, 100:2567–2572, 1996.

[39] I. Andricioaei, A. R. Dinner, and M. Karplus. Self-guided enhanced sampling methods for thermodynamic averages. *Journal of Chemical Physics*, 118:1074–1084, 2003.

[40] B. R. Brooks, D. Janezic, and M. Karplus. Harmonic-analysis of large systems .1. Methodology. *Journal of Computational Chemistry*, 16:1522–1542, 1995.

[41] N. Go, T. Noguti, and T. Nishikawa. Dynamics of a small globular protein in terms of low-frequency vibrational-modes. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, 80:3696–3700, 1983.

[42] M. Levitt, C. Sander, and P. S. Stern. Protein normal-mode dynamics - trypsin-inhibitor, crambin, ribonuclease and lysozyme. *Journal of Molecular Biology*, 181:423–447, 1985.

[43] B. Brooks and M. Karplus. Normal-modes for specific motions of macromolecules - application to the hinge-bending mode of lysozyme. *Proceedings of the National Academy of Sciences of the United States of America*, 82:4995–4999, 1985.

[44] J. P. Ma and M. Karplus. Ligand-induced conformational changes in ras p21: A normal mode and energy minimization analysis. *Journal of Molecular Biology*, 274:114–131, 1997.

[45] Q. Cui, G. H. Li, J. P. Ma, and M. Karplus. A normal mode analysis of structural plasticity in the biomolecular motor F-1-ATPase. *Journal of Molecular Biology*, 340:345–372, 2004.

[46] F. Tama and Y. H. Sanejouand. Conformational change of proteins arising from normal mode calculations. *Protein Engineering*, 14:1–6, 2001.

[47] W. G. Krebs, V. Alexandrov, C. A. Wilson, N. Echols, H. Y. Yu, and M. Gerstein. Normal mode analysis of macromolecular motions in a database framework: Developing mode concentration as a useful classifying statistic. *Proteins-Structure Function and Genetics*, 48:682–695, 2002.

[48] L. Yang, G. Song, and R.L. Jernigan. How well can we understand large-scale protein motions using normal modes of elastic network models. *Biophys. J.*, 93:920–929, 2007.

[49] M. Delarue and Y. H. Sanejouand. Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. *Journal of Molecular Biology*, 320:1011–1024, 2002.

[50] F. Tama, M. Valle, J. Frank, and C. L. Brooks. Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 100:9319–9323, 2003.

[51] G. T. Barkema and N. Mousseau. Event-based relaxation of continuous disordered systems. *Physical Review Letters*, 77:4358–4361, 1996.

[52] X. W. Wu and B. R. Brooks. Self-guided langevin dynamics simulation method. *Chemical Physics Letters*, 381:512–518, 2003.

[53] J. Hu, A. Ma, and A. R. Dinner. Bias annealing: A method for obtaining transition paths de novo. *Journal Of Chemical Physics*, 125, 2006.

[54] C. J. Cerjan and W. H. Miller. On finding transition-states. *J. Chem. Phys.*, 75:2800–2806, 1981.

[55] L. R. Pratt. A statistical-method for identifying transition-states in high dimensional problems. *Journal of Chemical Physics*, 85:5045–5048, 1986.

[56] R. Elber and M. Karplus. A method for determining reaction paths in large molecules - application to myoglobin. *Chem. Phys. Lett.*, 139:375–380, 1987.

[57] R. E. Gillilan and K. R. Wilson. Shadowing, rare events, and rubber bands - A variational Verlet algorithm for molecular-dynamics. *Journal of Chemical Physics*, 97:1757–1772, 1992.

[58] E. M. Sevick, A. T. Bell, and D. N. Theodorou. A chain of states method for investigating infrequent event processes occurring in multistate, multidimensional systems. *Journal of Chemical Physics*, 98:3196–3212, 1993.

[59] R. Olender and R. Elber. Calculation of classical trajectories with a very large time step: Formalism and numerical examples. *Journal of Chemical Physics*, 105:9299–9315, 1996.

[60] S. H. Huo and J. E. Straub. The maxflux algorithm for calculating variationally optimized reaction paths for conformational transitions in many body systems at finite temperature. *Journal of Chemical Physics*, 107:5000–5006, 1997.

[61] G. Henkelman, B. P. Uberuaga, and H. Jonsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *Journal of Chemical Physics*, 113:9901–9904, 2000.

[62] I. Andricioaei and J. E. Straub. On Monte Carlo and molecular dynamics methods inspired by Tsallis statistics: Methodology, optimization, and application to atomic clusters. *Journal of Chemical Physics*, 107:9117–9124, 1997.

[63] T. W. Whitfield, L. Bu, and J. E. Straub. Generalized parallel sampling. *Physica A-Statistical Mechanics And Its Applications*, 305:157–171, 2002.

[64] J. MacFadyen, J. Wereszczynski, and I. Andricioaei. Directionally negative friction: A method for enhanced sampling of rare event kinetics. *Journal Of Chemical Physics*, 128, 2008.

CHAPTER VI

Conclusions

Work presented in this dissertation has focused primarily on understanding the kinetic and equilibrium properties of DNA and RNA duplexes in solution, under external forces, and in complex with a topoisomerase protein. Molecular Dynamics (MD) simulations have primarily been employed with the aid of computational algorithms such as umbrella sampling and quasiharmonic analysis to elucidate energetic contributions to experimental observables and to offer predictions for future experiments. Additionally, a method is presented which may be applied to a variety of systems, including nucleic acids, in which a negative frictional term is used in a Langevin dynamics scheme to enhance sampling along a slow manifold.

In Chapter 2 simulations of DNA and RNA under tensile and torsional loads were presented. In the first part forces one to two orders of magnitude larger then those used in experiments were applied to observe transitions on the nanosecond timescale, and results indicated transitions to a recently observed structure termed P- form, a supercoiled analog (scP), and denatured states. Transitions to P- form were further examined with the aid of umbrella sampling and a reversible isothermal pathway was plotted onto a previously experimentally derived DNA phase diagram, showing excellent agreement. By following the same protocol with RNA it was possible to

provide a hypothesis for the location of a P-,scP-,A- triple point on a similar RNA phase diagram. Energetic and entropic contributions to P- form were also discussed.

A more comprehensive analysis of the entropic contribution to the stabilization of nucleic acid helices was presented in Chapter 3. This work was inspired by fluorescence experiments which demonstrated that the thermoresponsive nature of RNA and DNA helices could be used in the construction of "nanothermometers." Specifically it was observed, by the use of a fluorescent base analogue sensitive to the helical handedness, that RNA transitioned from right to left handed (A- to Z- form) while DNA transitioned from left to right handed helices (Z- to B- form) as the temperature increased for a specific sequence in high salt concentrations. Configurational, solvent, and ionic entropies were calculated and it was shown that a complex interplay between the three terms accounts for these transitions. A recently developed method of permutation reduction was employed for solvent and ionic calculations, allowing for a set of entropy calculations which would otherwise of not been attainable.

While the work in Chapter 2 focused on forces directly applied in laboratory conditions to nucleic acid molecules, Chapter 4 changed the focus to forces applied *in vivo* as a result of supercoiling and protein interaction. Specifically, human topoisomerase I in complex with a 22 base pair DNA sequence was simulated in conjunction with umbrella sampling to further our understanding of the mechanism and driving forces in supercoil relaxation. It was demonstrated that positive and negative supercoils are relieved through distinct pathways and evidence for a "semi-open" state was presented. Additional simulations were performed in which the inhibitor topotecan (TPT) was present in the topoisomerase/DNA complex, and its effect on the mechanism and free energy of supercoil relaxation showed a selective inhibition of positive supercoil relaxation, in accord with experimental results.

Recent advances in computational power and theoretical techniques have extended the scope of what may be studied with MD simulations, however there continues to exist substantial interest in the development of new methods for enhanced sampling of rare events. In Chapter 5 a new method in which Langevin dynamics are altered by the introduction of a negative frictional coefficient is introduced as one of these new methods. It has the advantage over similar techniques that energy is introduced solely in the direction of interest, allowing for its use on systems of high-dimensionality (such as MD on nucleic acids). Through proper trajectory-dependent reweighting it was demonstrated that time-correlation functions describing the physical, all-positive friction systems may be recovered from the directionally negative frictional dynamics.

Results presented in this dissertation aim at advancing our understanding of nucleic acid structure and energetics under native and strained conditions, and their interactions with proteins. Although a powerful theoretical device, MD can not provide a complete understanding of nucleic acid systems, thus whenever possible comparisons to experimental results have been made and have been shown to be mostly favorable. As computational power increases so do the systems sizes and simulation lengths that are attainable by methods utilized and developed herein, and with advances in methodology and force fields one would expect that results will become even more accurate when compared to experiments. The continued evolution of molecular dynamics from theoretical exercises to a rigorous method providing pertinent details relating to nanotechnology, drug discovery, HIV, and cancer research (to name only a few) has significant impact as multidisciplinary approaches become essential to making scientific progress.

APPENDICES

APPENDIX A

CHARMM Source Code for Torque Generation

For non-equilibrium calculations reported in Chapter 2 it was necessary to introduce a subroutine to calculate the forces on individual atoms due to an applied torque. This routine (TORQ) was based on the PULL routine and resulted in the editing of several files in the CHARMM source code, so a general LINUX patch is given below to add this function to CHARMM c31b1.

```
diff -Naur c31b1/source/charmm/charmm_main.src source/charmm/charmm_main.src
--- c31b1/source/charmm/charmm_main.src 2004-08-15 02:05:02.000000000 -0400
+++ source/charmm/charmm_main.src 2005-05-11 15:37:06.000000000 -0400
@@ -759,6 +759,8 @@
       ELSE IF (WRD.EQ.'PULL') THEN
 C---- 23-Jul-96, LN
         CALL PULL(COMLYN,COMLEN)
+      ELSE IF (WRD.EQ.'TORQ') THEN
+        CALL TORQ(COMLYN,COMLEN)
       ELSE IF (WRD.EQ.'QUAN') THEN
         CALL QMDEFN(COMLYN,COMLEN)
       ELSE IF (WRD.EQ.'READ') THEN
diff -Naur c31b1/source/charmm/iniall.src source/charmm/iniall.src
--- c31b1/source/charmm/iniall.src 2004-08-15 02:05:03.000000000 -0400
+++ source/charmm/iniall.src 2005-05-11 14:24:14.000000000 -0400
@@ -844,6 +844,7 @@
 C=====================================================================
 C PULL.FCM
       NPULL=0
+      NTORQ=0
 C
 C=====================================================================
 C QUANTM.FCM
diff -Naur c31b1/source/energy/energy.src source/energy/energy.src
--- c31b1/source/energy/energy.src 2004-08-15 02:05:04.000000000 -0400
+++ source/energy/energy.src 2005-05-11 14:32:07.000000000 -0400
@@ -1794,6 +1794,18 @@
       ENDIF                                                     !##MTS
 C
 C---------------------------------------------------------------------
+C . TORQ, MAY 2005 LN
+C
+      IF(ENE2) THEN                                             !##MTS
+        IF(MYNOD.EQ.INODE(12)) THEN  !##PARALLEL
+          IF(QETERM(TORQUE)) THEN
+            CALL ETORQ(ETERM(TORQUE),DX,DY,DZ,X,Y,Z)
```

```
+            IF (TIMER.GT.1) CALL WRTTIM('PULL energy times:')
+          ENDIF
+        ENDIF                !##PARALLEL
+      ENDIF                                                  !##MTS
+C
+C-----------------------------------------------------------------
 C . Stochastic boundary terms.
       IF(ENE3) THEN                                          !##MTS
         IF(MYNOD.EQ.INODE(13)) THEN  !##PARALLEL
diff -Naur c31b1/source/energy/epull.src source/energy/epull.src
--- c31b1/source/energy/epull.src 2003-08-13 18:02:13.000000000 -0400
+++ source/energy/epull.src 2005-05-11 14:44:05.000000000 -0400
@@ -313,3 +313,295 @@
       WRITE(OUTU,'(//)')
       RETURN
       END
+
+
+      SUBROUTINE ETORQ(EP,DX,DY,DZ,X,Y,Z)
+C-----------------------------------------------------------------
+C This routine calculates the externally applied forces (and energies)
+C as specified in the TORQ command.
+C Lennart Nilsson, Karolinska institutet, March 1996
+C
+##INCLUDE '~/charmm_fcm/impnon.fcm'
+##INCLUDE '~/charmm_fcm/dimens.fcm'
+      REAL*8 EP,DX(*),DY(*),DZ(*),X(*),Y(*),Z(*)
+##INCLUDE '~/charmm_fcm/pull.fcm'
+##INCLUDE '~/charmm_fcm/heap.fcm'
+##INCLUDE '~/charmm_fcm/psf.fcm'
+C
+      IF(NTORQ .GT. 0) CALL ETORQ2(EP,NATOM,DX,DY,DZ,
+     &                 X,Y,Z,NTORQ,
+     &                 HEAP(TTO), HEAP(TAMPL), HEAP(TPER), TTIME)
+      RETURN
+      END
+
+      SUBROUTINE ETORQ2(EP,NATOM,DX,DY,DZ,X,Y,Z,NTORQ,
+     &                  TTO,AMPL,PERIOD,TIME)
+C-----------------------------------------------------------------
+##INCLUDE '~/charmm_fcm/impnon.fcm'
+##INCLUDE '~/charmm_fcm/number.fcm'
+##INCLUDE '~/charmm_fcm/consta.fcm'
+##INCLUDE '~/charmm_fcm/corman.fcm'
+      INTEGER NATOM,NTORQ,TTO(NTORQ)
+      REAL*8  EP,DX(NATOM),DY(NATOM),DZ(NATOM)
+      REAL*8  X(NATOM),Y(NATOM),Z(NATOM)
+      REAL*8  PERIOD(NTORQ),AMPL(NTORQ)
+      REAL*8  YD,ZD,TIME,RADIUS,XD
+      REAL*8  DIFX,DIFY,DIFZ,LENG
+      REAL*8  CX,CY,CZ,DOTX,DOTY,DOTZ
+c      REAL*8  AXISCX, AXISCY,AXISCZ
+c      REAL*8  AXISX,AXISY,AXISZ,AXISR
+      REAL*8  ADOTX,ADOTY,ADOTZ,VX,VY,VZ
+      INTEGER APPLY
+C
+      INTEGER I,J
+      REAL*8  T,XF,YF,ZF
+C
```

```
+      EP=ZERO
+      T=TIME*TWOPI
+      DO I=1,NTORQ
+         APPLY=1
+         J=TTO(I)
+         IF(J.GT.NATOM)THEN
+            CALL WRNDIE(-4,'<ETORQ>','Torque on non existing atom')
+         ENDIF
+         DIFX=AXISCX-X(J)
+         DIFY=AXISCY-Y(J)
+         DIFZ=AXISCZ-Z(J)
+         ADOTX=-1*DIFX*AXISX/(AXISR**2)
+         ADOTY=-1*DIFY*AXISY/(AXISR**2)
+         ADOTZ=-1*DIFZ*AXISZ/(AXISR**2)
+         CX=AXISCX+ADOTX*AXISX
+         CY=AXISCY+ADOTY*AXISY
+         CZ=AXISCZ+ADOTZ*AXISZ
+         VX=X(J)-CX
+         VY=Y(J)-CY
+         VZ=Z(J)-CZ
+         RADIUS=SQRT(VX**2+VY**2+VZ**2)
+c         IF(RADIUS.LT.1) THEN
+c            RADIUS=1
+c         END
+         XD=-VY*AXISZ+VZ*AXISY
+         YD=-VZ*AXISX+VX*AXISZ
+         ZD=-VX*AXISY+VY*AXISX
+         LENG=SQRT(XD**2+YD**2+ZD**2)
+         XD=XD/LENG
+         YD=YD/LENG
+         ZD=ZD/LENG
+         IF(PERIOD(I) .GT. ZERO)THEN
+            XF=XD*AMPL(I)*COS(T/PERIOD(I))/RADIUS
+            YF=YD*AMPL(I)*COS(T/PERIOD(I))/RADIUS
+            ZF=ZD*AMPL(I)*COS(T/PERIOD(I))/RADIUS
+         ELSE
+            XF=XD*AMPL(I)/RADIUS
+            YF=YD*AMPL(I)/RADIUS
+            ZF=ZD*AMPL(I)/RADIUS
+         ENDIF
+         DX(J)=DX(J)-XF
+         DY(J)=DY(J)-YF
+         DZ(J)=DZ(J)-ZF
+         EP=EP-(XF*X(J)+YF*Y(J)+ZF*Z(J))
+      ENDDO
+      RETURN
+      END
+
+      SUBROUTINE TORQ(COMLYN,COMLEN)
+C----------------------------------------------------------------------
+C Parse the TORQ command:
+C
+C TORQ { TORque  <real> } [PERIod <real>]
+C      { OFF       }
+C      { LIST      }
+C                     [WEIGht] [atom-selection]
+C
+C
+C
```

```
+##INCLUDE '~/charmm_fcm/impnon.fcm'
+      INTEGER COMLEN
+      CHARACTER*(*) COMLYN
+C
+##INCLUDE '~/charmm_fcm/dimens.fcm'
+##INCLUDE '~/charmm_fcm/exfunc.fcm'
+##INCLUDE '~/charmm_fcm/coord.fcm'
+##INCLUDE '~/charmm_fcm/heap.fcm'
+##INCLUDE '~/charmm_fcm/number.fcm'
+##INCLUDE '~/charmm_fcm/pull.fcm'
+##INCLUDE '~/charmm_fcm/psf.fcm'
+##INCLUDE '~/charmm_fcm/stack.fcm'
+##INCLUDE '~/charmm_fcm/stream.fcm'
+##INCLUDE '~/charmm_fcm/corman.fcm'
+C
+      LOGICAL QWEIG
+      INTEGER I,N,OLDUSD,ISLCT,IP1,TPER1,TAMP1,IX1,IY1,IZ1
+      REAL*8 TOR,PER,EFI
+      INTEGER WDMAX,WDLEN
+      PARAMETER (WDMAX=4)
+      CHARACTER*(WDMAX) WD
+C
+C
+C What to do?
+      IF(.NOT. QAXISC)THEN
+            CALL WRNDIE(-3,'<TORQ>','No Axis defined')
+      ENDIF
+      WD=CURRA4(COMLYN,COMLEN)
+      IF(WD .EQ. 'OFF ')THEN
+C Reset torques - if there are any defined
+         IF(NTORQ .GT. 0)THEN
+            CALL FREHP(TTO,INTEG4(NTORQ))
+            CALL FREHP(TAMPL,IREAL8(NTORQ))
+            CALL FREHP(TPER,IREAL8(NTORQ))
+         ENDIF
+         NTORQ=0
+         COMLEN=0
+         RETURN
+      ELSEIF (WD .EQ. 'LIST')THEN
+C Produce listing
+         CALL TORQLIST(NTORQ,HEAP(TTO),HEAP(TAMPL),HEAP(TPER))
+         COMLEN=0
+         RETURN
+      ENDIF
+C What atoms are we using
+      OLDUSD=LSTUSD
+      ISLCT=ALLSTK(INTEG4(NATOM))
+      CALL SELCTA(COMLYN,COMLEN,STACK(ISLCT),X,Y,Z,WMAIN,.TRUE.)
+      N=NSELCT(NATOM,STACK(ISLCT))
+      IF(N .LE. 0)THEN
+C Nothing to be done now, clean up and return
+         IF(PRNLEV .GT. 2)THEN
+            WRITE(OUTU,90) NTORQ
+90          FORMAT(/' TORQ> No torque added. Total now',I8/)
+         ENDIF
+         LSTUSD=OLDUSD
+         RETURN
+      ENDIF
+C
```

```
+C Set up torques
+      QWEIG= (INDXA(COMLYN,COMLEN,'WEIG').GT.0)
+      TOR=GTRMF(COMLYN,COMLEN,'TOR',ZERO)
+      EFI=GTRMF(COMLYN,COMLEN,'EFIE',ZERO)
+      PER=GTRMF(COMLYN,COMLEN,'PERI',MINONE)
+C
+      IF(NTORQ .GT. 0)THEN
+C
+C Have to make room for new data
+C
+         IP1=ALLHP(INTEG4(N+NTORQ))
+         TAMP1=ALLHP(IREAL8(N+NTORQ))
+         TPER1=ALLHP(IREAL8(N+NTORQ))
+C now copy existing data
+         CALL COPYI4(HEAP(TTO),HEAP(IP1),NTORQ)
+         CALL COPYR8(HEAP(TAMPL),HEAP(TAMP1),NTORQ)
+         CALL COPYR8(HEAP(TPER),HEAP(TPER1),NTORQ)
+C and release no longer used space
+         CALL FREHP(TTO,INTEG4(NTORQ))
+         CALL FREHP(TAMPL,IREAL8(NTORQ))
+         CALL FREHP(TPER,IREAL8(NTORQ))
+C
+         TTO=IP1
+         TPER=TPER1
+         TAMPL=TAMP1
+      ELSE
+         TTO=ALLHP(INTEG4(N))
+         TAMPL=ALLHP(IREAL8(N))
+         TPER=ALLHP(IREAL8(N))
+      ENDIF
+C
+      CALL TORQ2(N,NTORQ,NATOM,TOR,PER,EFI,QWEIG,
+     &           STACK(ISLCT),WMAIN,CG,
+     &           HEAP(TTO),HEAP(TAMPL),HEAP(TPER))
+
+C
+      IF(PRNLEV .GT. 2)THEN
+        WRITE(OUTU,100) N, N+NTORQ
+100     FORMAT(/' TORQ>',I8,' Torque(s) added. Total now',I8/)
+      ENDIF
+      NTORQ=N+NTORQ
+      LSTUSD=OLDUSD
+      RETURN
+      END
+
+      SUBROUTINE TORQ2(N,NTORQ,NATOM,TOR,PER,EFI,QWEIG,
+     &           ISLCT,WMAIN,CG,
+     &           TTO,AMPL,PERIOD)
+C-----------------------------------------------------------------------
+##INCLUDE '~/charmm_fcm/impnon.fcm'
+##INCLUDE '~/charmm_fcm/number.fcm'
+C
+      INTEGER N,NATOM,ISLCT(NATOM),TTO(N),NTORQ
+      LOGICAL QWEIG
+      REAL*8 TOR,PER,EFI,PERIOD(N)
+      REAL*8 WMAIN(NATOM),AMPL(N),CG(NATOM)
+C
+      INTEGER I,J,JS,JS1
+      REAL*8 XX
```

```
+      REAL*8 AKMPN,AKMEVM,EQ,NAVG,ANGST,KCAL,PICO
+      PARAMETER (EQ= 1.6021892D-19,NAVG= 6.022045D23,KCAL=4184D0)
+      PARAMETER (ANGST=1D-10,PICO=1D-12)
+C Conversion between AKMA force units and pN or electron charge * Volts/m
+      PARAMETER (AKMPN=1D0/(KCAL/NAVG/ANGST/PICO),AKMEVM=AKMPN*EQ/PICO)
+C
+      IF(TOR.NE.ZERO)THEN
+C Use this value
+         XX=TOR*AKMPN
+      ELSE IF(EFI .NE. ZERO)THEN
+         XX=EFI*AKMEVM
+      ELSE
+         CALL WRNDIE(-3,'<TORQ>','No torque specified')
+         XX=ZERO
+      ENDIF
+      JS=0
+      DO I=1+NTORQ,N+NTORQ
+         PERIOD(I)=PER
+         AMPL(I)=XX
+C Find atom index of first selected atom
+         JS1=JS
+         DO J=JS1,NATOM
+            IF(ISLCT(J).EQ.1)THEN
+               JS=J
+               ISLCT(J)=0
+               GOTO 99
+            ENDIF
+         ENDDO
+99       CONTINUE
+         TTO(I)=JS
+         IF(QWEIG) AMPL(I)=AMPL(I)*WMAIN(JS)
+         IF(EFI .NE. ZERO) AMPL(I)=AMPL(I)*CG(JS)
+      ENDDO
+      RETURN
+      END
+
+      SUBROUTINE TORQLIST(NTORQ,TTO,AMPL,PER)
+C----------------------------------------------------------------------
+C List currently effective torques to OUTU
+C
+      INTEGER NTORQ,TTO(NTORQ)
+      REAL*8 AMPL(NTORQ),PER(NTORQ)
+
+##INCLUDE '~/charmm_fcm/stream.fcm'
+      INTEGER I,J
+      CHARACTER*4 ATNAM,RID,SID,REN
+      REAL*8 AKMPN,AKMEVM,EQ,NAVG,ANGST,KCAL,PICO
+      PARAMETER (EQ= 1.6021892D-19,NAVG= 6.022045D23,KCAL=4184D0)
+      PARAMETER (ANGST=1D-10,PICO=1D-12)
+C Conversion between AKMA force units and pN or electron charge * Volts/m
+      PARAMETER (AKMPN=1D0/(KCAL/NAVG/ANGST/PICO),AKMEVM=AKMPN*EQ/PICO)
+C
+      IF(IOLEV .LE. 0 .OR. PRNLEV .LE. 1)RETURN
+      IF(NTORQ.LE.0)THEN
+         WRITE(OUTU,'(/,A,/)') ' TORQ> No external forces defined.'
+         RETURN
+      ENDIF
+      WRITE(OUTU,801) NTORQ
+801   FORMAT(' TORQ> Currently these',I6,' torques are defined',
```

```
+     & /16X,'Atom      Ampl [AKMA]       [pN]   Per [ps]',
+     & '  Direction (X,Y,Z)')
+      DO I=1,NTORQ
+         J=TTO(I)
+         CALL ATOMID(J,SID,RID,REN,ATNAM)
+         WRITE(OUTU,901) J,SID,RID,REN,ATNAM,AMPL(I),AMPL(I)/AKMPN,
+     &                   PER(I)
+901      FORMAT(1X,I5,4(1X,A4),1X,F10.4,1X,F10.3,1X,F10.3,3(1X,F5.2))
+      ENDDO
+      WRITE(OUTU,'(//)')
+      RETURN
+      END
+
diff -Naur c31b1/source/energy/eutil.src source/energy/eutil.src
--- c31b1/source/energy/eutil.src 2004-02-13 13:55:09.000000000 -0500
+++ source/energy/eutil.src 2005-05-11 14:56:23.000000000 -0400
@@ -502,6 +502,7 @@
       CETERM(SHAP)   = 'SHAP'  ! Shape restraint energy
       CETERM(STRB)   = 'STRB'  ! Strech-Bend coupling energy (MMFF and CFF)
       CETERM(OOPL)   = 'OOPL'  ! Out-off-plane energy (MMFF and CFF)
+      CETERM(TORQUE) = 'TORQ'
       CETERM(PULL)   = 'PULL'  ! Pulling force energy
       CETERM(POLAR)  = 'POLA'  ! Polarizable water energy
       CETERM(DMC )   = 'DMC '  ! Distance map restraint energy
diff -Naur c31b1/source/fcm/energy.fcm source/fcm/energy.fcm
--- c31b1/source/fcm/energy.fcm 2003-08-13 18:02:13.000000000 -0400
+++ source/fcm/energy.fcm 2005-05-11 14:59:52.000000000 -0400
@@ -136,7 +136,7 @@
      &         STRB, OOPL, PULL, POLAR, DMC, RGY, EWEXCL, EWQCOR,
      &         EWUTIL, PBELEC, PBNP, PINT, MbDefrm, MbElec, STRSTR,
      &         BNDBND, BNDTW, EBST, MBST, BBT, SST, GBEnr, GSBP
-     &       , HMCM             !## HMCM
+     &       , HMCM, TORQUE     !## HMCM
      &       , ADUMB            !## ADUMB
      &       , HYDR             !## ACE
      &       , GrvdW, GrElec    !## GRID
@@ -161,7 +161,7 @@
      &           PBELEC = 49, PBNP   = 50, MbDefrm= 51, MbElec = 52,
      &           STRSTR = 53, BNDBND = 54, BNDTW  = 55, EBST   = 56,
      &           MBST   = 57, BBT    = 58, SST    = 59, GBEnr  = 60,
-     &           GSBP   = 64
+     &           GSBP   = 64, TORQUE = 72
      &         , HMCM   = 61             !## HMCM
      &         , ADUMB  = 62             !## ADUMB
      &         , HYDR   = 63             !## ACE
diff -Naur c31b1/source/fcm/pull.fcm source/fcm/pull.fcm
--- c31b1/source/fcm/pull.fcm 1996-08-07 00:34:24.000000000 -0400
+++ source/fcm/pull.fcm 2005-05-11 15:02:38.000000000 -0400
@@ -7,9 +7,17 @@
       REAL*8  PTIME
       COMMON /PULLI/ NPULL,IPU,IAMPL,IPER,IXDIR,IYDIR,IZDIR
       COMMON /PULLR/ PTIME
+
+
+      INTEGER NTORQ,TTO,TAMPL,TPER
+      REAL*8  TTIME
+      COMMON /TORQI/ NTORQ,TTO,TAMPL,TPER
+      COMMON /TORQR/ TTIME
 C
```

```
 ##IF SAVEFCM
       SAVE /PULLI/
       SAVE /PULLR/
+      SAVE /TORQI/
+      SAVE /TORQR/
 ##ENDIF
 C
```

APPENDIX B

Code for Dihedral Umbrella Sampling in NAMD

The topoisomerase project in Chapter 4 required the coding of an externally applied force to apply the dihedral constraint necessary for umbrella sampling. This was done via the TclForces subroutine in NAMD. To implement the code a section must be added to the NAMD input file such as:

```
tclForces               on
set tref                $window
set FOR                 $force
set targetAtomPdb       AtomsToMove.pdb
tclforcesscript         dihedral_force.tcl
```

With the variables *window* and *force* being previously defined as the minimum (in degrees) and force constant (in kcal/mol/radians^2) corresponding to the window of choice. The file AtomsToMove.pdb being a pdb file containing all the atoms of the systems, with those that are to be constrained having beta and occupancy values of 1.00, while all others have a value of 0.00, and the file dihedral_force.tcl the file to follow.

dihedral_force.tcl

```
set targetMark    "1.00"
set forcesRecalcFreq 2

set targets {}
set masses {}
set inStream [open $targetAtomPdb r]
foreach line [split [read $inStream] \n] {
    set string1 [string range $line 0 3]
    set string2 [string range $line 6 10]
    set string3 [string range $line 17 20]
    set string4 [string range $line 12 15]
    set string5 [string range $line 46 53]
    set string6 [string range $line 72 75]
    set string673 [string range $line 73 75]
    set string7 [string range $line 22 25]
    set string8 [string range $line 62 65]
    set string9 [string range $line 55 60]

    if {[string equal $targetMark $string8] } {
  lappend targets "[string trim $string6] \\
[string trim $string7] [string trim $string4]"
lappend masses "[string trim $string9]"
```

```
    }
}
close $inStream

set atoms {}
foreach target $targets {
    foreach {segname resid atom} $target { break }
    set atomindex [atomid $segname $resid $atom]
    lappend atoms $atomindex
    addatom $atomindex
}

set numatoms [llength $atoms]

set forces {}

foreach index $atoms {
lappend forces "0.0 0.0 0.0"
}

set forcecount  $forcesRecalcFreq
set printcount  0

set hbatoms {}
set hbforces {}
set dihatoms {}
set dihforces {}
proc calcforces { } {
global atoms numatoms masses tref FOR
global forcesRecalcFreq forcecount printcount \\
forces hbatoms hbforces
foreach atom $atoms force $forces {
addforce $atom $force
}
foreach atom $hbatoms force $hbforces {
addforce $atom $force
}
if { $forcecount == [expr $forcesRecalcFreq -1] } {
foreach atom $atoms {
addatom $atom
}
}
if { $forcecount == $forcesRecalcFreq } {

loadcoords cords
set comsum "0 0 0"
set totalmass 0
foreach atom $atoms mass $masses {
         set comsum [vecadd $comsum [vecscale $mass $cords($atom)]]
    set totalmass [expr $totalmass + $mass]
}
set com [vecscale [expr 1.0/$totalmass] $comsum]
               set phi [getdihedral $com \\
{-21.6920   -1.5580    7.3960} {-32.8790  -28.1320  -15.0160}\\
 {-4.988  10.286  16.868}]
set t [getstep]
if { $phi > 360 } { set phi [expr $phi - 360]}
if { $phi < 0 } { set phi [expr $phi + 360]}
```

```
print "DIHEDRAL $t $phi "

if { $tref < 0 } { set tref [expr $tref + 360] }
set phi [expr $phi - $tref]
if {$phi > 180} { set phi [expr $phi - 360] }
if {$phi < -180} { set phi [expr $phi + 360] }
set phi [expr $phi * .0174533]
set E [expr .5 * $FOR * $phi * $phi]
set DE [expr -1 * $FOR * $phi]
               foreach {g1 g2 g3 g4} [dihedralgrad $com \\
{-21.6920   -1.5580    7.3960} {-32.8790  -28.1320  -15.0160} \\
{-4.988  10.286  16.868}] {}
set DF [vecscale $DE $g1]
       set forces {}
addenergy $E
   foreach atom $atoms mass $masses {
            set force [vecscale [expr $mass/$totalmass] $DF]
             lappend forces $force
}
set hbatoms {}
set hbforces {}
set hb1 10524
set hb2 9475
set hvect [vecadd $cords($hb1) [vecscale -1 $cords($hb2)] ]
set hbond [getbond $cords($hb1) $cords($hb2)]
set fhb2 [vecscale [expr 50 * ($hbond - 3)] $hvect]
set fhb1 [vecscale -1  $fhb2]
lappend hbforces $fhb1
lappend hbforces $fhb2
lappend hbatoms $hb1
lappend hbatoms $hb2
set hb1 10521
set hb2 9469
set hvect [vecadd $cords($hb1) [vecscale -1 $cords($hb2)] ]
set hbond [getbond $cords($hb1) $cords($hb2)]
set fhb2 [vecscale [expr 50 * ($hbond - 3)] $hvect]
set fhb1 [vecscale -1  $fhb2]
lappend hbforces $fhb1
lappend hbforces $fhb2
lappend hbatoms $hb1
lappend hbatoms $hb2
set forcecount 0
clearconfig
}
incr forcecount
incr printcount

return
}
```

APPENDIX C

Parameters for Topotecan in the CHARMM Force Field

Topotecan was built in the program Insight (Accelyrs) and transfered to CHARMM by additions to the topology and parameter files. A LINUX patch to modify the files top_all27_prot_na.rtf and par_all27_prot_na.prm to include this residue (TTC) and to include an additional residue which comprises the tyrosine/DNA bond.

```
diff -Naur oldtopology/par_all27_prot_na.prm newtopology/par_all27_prot_na.prm
--- oldtopology/par_all27_prot_na.prm 2008-07-28 17:48:16.000000000 -0400
+++ newtopology/par_all27_prot_na.prm 2008-07-28 17:43:48.000000000 -0400
@@ -84,6 +84,38 @@
 !
 !atom type Kb          b0
 !
+C6R  C6R   440.0       1.383
+C6R  C6RP  450.0       1.370
+C6R  OT2   450.0       1.38
+HA   C6R   370.0       1.08
+C6RP C6RP  460.0       1.485
+C6RP N6R   390.0       1.36
+C6R  CT    345.0       1.500
+CR56 N6R   390.0       1.342
+CR56 CR56  450.0       1.410
+CR56 CUA1  325.0       1.49
+C6R  CR56  450.0       1.360
+HA   CT    340.0       1.09
+CUA1 CUA1  590.0       1.343
+CT   CUA1  385.0       1.502
+CR56 CT    345.0       1.47
+CUA1 NX    390.0       1.47
+C    CT    242.0       1.530
+CT   NX    310.0       1.458
+CT   NT    340.0       1.458
+HO   OT2   505.0       0.948
+C    NX    350.0       1.350
+HA   CUA1  365.0       1.074
+C    CUA1  282.0       1.476
+CT   OT2   375.0       1.405
+CT   CT    268.0       1.529
+C    OE    375.0       1.354
+CT   OE    390.0       1.407
+
+
+
+
```

```
+CA   ON2   310.0       1.433   ! from CN7  ON2
 C    C     600.000     1.3350 ! ALLOW ARO HEM
                ! Heme vinyl substituent (KK, from propene (JCS))
 CA   CA    305.000     1.3750 ! ALLOW   ARO
@@ -531,6 +563,71 @@
 !
 !atom types     Ktheta    Theta0   Kub     S0
 !
+C6R  C6R  C6R     70.0     120.0
+C6R  C6R  OT2     65.0     120.0
+HA   C6R  C6R     31.0     120.0
+HA   C6R  C6RP    31.0     120.0
+C6R  C6R  C6RP    70.0     120.0
+C6R  C6RP C6RP    70.0     121.0
+N6R  C6RP C6RP    70.0     120.0
+C6R  C6RP N6R     70.0     120.0
+C6R  C6R  CT      70.0     122.0
+CT   C6R  C6RP    60.0     120.0
+C6R  C6RP C6R     70.0     118.0
+C6RP N6R  CR56    70.0     120.0
+CR56 CR56 N6R     70.0     127.5
+CR56 CR56 C6R     70.0     122.5
+C    CUA1 CUA1    72.0     119.5
+CT   CUA1 CUA1    40.0     122.9
+CUA1 CUA1 CUA1    60.0     180.0
+HA   CUA1 CUA1    35.0     119.4
+CT   NX   CUA1    60.0     120.0
+CUA1 C    O       40.0     120.0
+CR56 C6R  CT      70.0     120.0
+HA   CT   NX      50.0     109.47
+CT   CT   OT      45.0     110.50
+C    CT   CT      52.0     109.90
+CUA1 CT   OT      60.0     109.47
+HA   CT   CR56    40.0     109.47
+CUA1 C    NX      55.0     119.0
+C    CT   CUA1    60.0     109.47
+HA   CT   CUA1    50.0     110.00
+C    OE   CT      70.0     120.0
+C    CUA1 CT      60.0     120.0
+CT   C    O       98.0     121.9
+CT   C    OE      75.0     114.0
+HA   CT   HA      33.0     107.80
+HA   CT   C6R     40.0     109.47
+HA   CT   NT      55.0     107.80
+CT   NT   CT      35.0     110.5
+CT   CT   CT      58.35    112.70
+HA   CT   CT      37.5     110.70
+HO   OT2  CT      59.0     106.7
+HO   OT2  C6R     50.0     109.47
+HA   CT   OE      55.5     109.47
+C    CT   OT2     80.0     109.47
+NX   C    O       98.0     124.5
+C    NX   CUA1    50.0     126.0
+C    NX   CT      65.0     117.6
+CUA1 CUA1 NX      50.0     122.0
+CUA1 CT   OT2     60.0     109.47
+CUA1 CT   OE      60.0     109.47
+C6R  CT   NT      70.0     111.6
+CT   CT   OT2  45.0     110.50
```

```
+CT   CT   CUA1 45.0     112.90
+O    C    OE   81.0     124.4
+HA   C6R  CR56 31.0     120.0
+C6RP C6R  CR56 70.0     120.0
+CR56 CUA1 CUA1 55.0     120.0
+CR56 CT   NX   70.0     111.6
+CR56 CR56 CUA1 65.0     133.0
+CT   CR56 CR56 70.0     107.0
+CT   CR56 C6R  70.0     133.0
+CR56 CUA1 NX   65.0     112.0
+N6R  CR56 CUA1 70.0     120.0
+
+CA   CA   ON2  45.200   120.0000 ! from OH1  CA   CA
+CA   ON2  P    20.0     120.0   35.00   2.33 ! from CN7  ON2  P
 CA   CA   CA    40.000    120.00   35.00   2.41620 ! ALLOW   ARO
                 ! JES 8/25/89
 CE1  CE1  CT3    48.00    123.50   !
@@ -1572,6 +1669,73 @@
 !
 !atom types             Kchi    n   delta
!
+X    C6RP C6RP X        0.8  2  180.0
+X    C6RP C6R  X        0.8  2  180.0
+O    C    C6R  C6R      1.3  2  180.0
+C6R  C6R  C6R  C6R      2.8  2  180.0
+C6R  C6R  C6R  HA       3.0  2  180.0
+HA   C6R  C6R  HA       2.5  2  180.0
+X    C6R  C6R  X        3.1  2  180.0
+X    CR56 CR56 X        3.1  2  180.0
+X    C6RP N6R  X        3.1  2  180.0
+X    CT   C6R  X        0.01 6    0.0
+X    C6R  OT2  X        0.5  2  180.0
+X    C6R  CR56 X        3.1  2  180.0
+O    C    CUA1 CUA1     1.5  1    0.0
+CT   CT   CUA1 CUA1     0.12 1    0.0
+CT   CT   CUA1 CUA1     4.4  3  180.0
+HA   CT   CUA1 CUA1     2.0  3  180.0
+CT   CUA1 CUA1 CT       0.36 1    0.0
+CT   CUA1 CUA1 CT       3.55 2  180.0
+CT   CUA1 CUA1 NP       3.45 2  180.0
+CUA1 CUA1 OE   CT       1.18 1  180.0
+CUA1 CUA1 OE   CT       1.80 2  180.0
+HA   CUA1 CUA1 CT       3.45 2  180.0
+HA   CUA1 CUA1 HA       3.45 2  180.0
+HA   CUA1 CUA1 NP       3.45 2  180.0
+X    CUA1 CUA1 X        3.55 2  180.0
+X    CT   CR56 X        3.1  2  180.0
+X    CUA1 CR56 X        3.1  2  180.0
+X    CUA1 NX   X        0.25 2  180.0
+C    CT   CT   OE       0.85 3    0.0
+CT   CT   CT   CT       0.70 1    0.0
+CT   CT   CT   CT       0.13 3    0.0
+CT   CT   CT   NT       0.82 1    0.0
+CT   CT   CT   OE       0.55 1    0.0
+CT   CT   CT   OT       0.65 1    0.0
+CUA1 CT   CT   NP       0.85 3    0.0
+NP   CT   CT   OE       0.85 3    0.0
+OT   CT   CT   OT       3.43 1    0.0
+OT   CT   CT   OT       1.70 3    0.0
```

```
+X    CT   CT   X        0.15 3    0.0
+X    C    NX   X        2.6  2  180.0
+X    CR56 N6R  X        3.1  2  180.0
+CT   CT   NT   CT       1.15 1    0.0
+CT   CT   NT   H        0.07 1    0.0
+X    CT   NT   X        0.16 3    0.0
+X    CT   NX   X        0.01 3    0.0
+X    C    CUA1 X        0.9  2  180.0
+X    CT   CUA1 X        1.2  3  180.0
+X    C    CT   X        0.05 3  180.0
+X    CT   OE   X        0.27 3    0.0
+CT   CT   OT2  HO       0.35 3    0.0
+X    CT   OT2  X        0.25 3    0.0
+X    C    OE   X        2.0  2  180.0
+
+
+
+CA   CA   CA   ON2      3.1000  2   180.00 ! from OH1  CA   CA   CA
+HP   CA   CA   ON2      4.2000  2   180.00 ! from OH1  CA   CA   HP
+CA   ON2  P    ON3      0.10    3     0.0  ! from ON3  P    ON2  CN7
+CA   ON2  P    ON2      1.20    1   180.0  ! from ON2  P    ON2  CN7
+CA   ON2  P    ON2      0.10    2   180.0  !
+CA   ON2  P    ON2      0.10    3   180.0  !
+CA   ON2  P    ON2      0.00    6     0.0  !
+CA   CA   ON2  P        0.6     5     0.0  ! from CN7  CN7  ON2  P
+CA   CA   ON2  P        0.2     4     0.0 !
+CA   CA   ON2  P        0.0     3   180.0 !
+CA   CA   ON2  P        0.4     2     0.0 !
+CA   CA   ON2  P        1.9     1   180.0 !
 C    CT1  NH1  C        0.2000  1   180.00 ! ALLOW PEP
                 ! ala dipeptide update for new C VDW Rmin, adm jr., 3/3/93c
 C    CT2  NH1  C        0.2000  1   180.00 ! ALLOW PEP
@@ -3407,6 +3571,21 @@
 !
 !atom  ignored    epsilon      Rmin/2   ignored   eps,1-4       Rmin/2,1-4
 !
+CT     0.000000  -0.110000     2.000000 !
+CUA1   0.000000  -0.110000     2.000000 !
+C6R    0.000000  -0.110000     2.000000 !
+C5R    0.000000  -0.110000     2.000000 !
+CR56   0.000000  -0.110000     2.000000 !
+C6RP   0.000000  -0.110000     2.000000 !
+NX     0.000000  -0.200000     1.850000
+N6R    0.000000  -0.200000     1.850000
+NT     0.000000  -0.200000     1.850000
+OT2    0.000000  -0.152100     1.768200 !
+OE     0.000000  -0.152100     1.768200 !
+HO     0.000000  -0.046000     0.224500 !
+
+
+
 C      0.000000  -0.110000     2.000000 ! ALLOW   PEP POL ARO
                 ! NMA pure solvent, adm jr., 3/3/93
 CA     0.000000  -0.070000     1.992400 ! ALLOW   ARO
diff -Naur oldtopology/top_all27_prot_na.rtf newtopology/top_all27_prot_na.rtf
--- oldtopology/top_all27_prot_na.rtf 2008-07-28 17:48:07.000000000 -0400
+++ newtopology/top_all27_prot_na.rtf 2008-07-28 17:43:48.000000000 -0400
@@ -220,6 +220,19 @@
 !MASS 105  HR1   1.008000 H ! For imidazole model compound (NF)
```

```
 !MASS 106  NR1  14.007000 H ! For nitrogen in imidazol (NF)
 !MASS 107  NR2  14.007000 H ! For nitrogen in imidazol (NF)
+!TTC SECTION
+MASS    229 C6RP 12.01100 C  ! for Aryl-Aryl bond between C6R rings
+MASS    235 N6R  14.00670 N  ! Nitrogen in a six membered aromatic ring
+MASS    226 CR56 12.01100 C  ! Aromatic carbon-merged five/six membered rings
+MASS    216 CUA1 12.01100 C  ! Carbon in double bond, first pair
+MASS    210 CT   12.01100 C  ! Aliphatic carbon (tetrahedral)
+MASS    250 OE   15.99940 O  ! Ether oxygen / Acetal oxygen
+MASS    245 OT2  15.99940 O  ! Hydroxyl oxygen (tetrahedral) or Ionizable acid
+MASS    236 NT   14.00670 N  ! Nitrogen (tetrahedral), i.e. Amine, etc.
+MASS    233 NX   14.00670 N  ! Proline nitrogen, or similar
+MASS    208 HO   1.00800 H  ! Hydrogen on an alcohol oxygen
+MASS    222 C6R  12.01100 C  ! Aromatic carbon in a six  membered ring
+MASS    223 C5R  12.01100 C  ! Aromatic carbon in a six  membered ring

 DECL -CA
 DECL -C
@@ -238,6 +251,100 @@
 DEFA FIRS NTER LAST CTER
 AUTO ANGLES DIHE

+RESI TTC   0.00000
+GROUP
+ATOM C1   C6R    0.25000
+ATOM C2   C6R   -0.13000
+ATOM C3   C6R   -0.13000
+ATOM C4   C6RP   0.10000
+ATOM C5   C6RP   0.00000
+ATOM C6   C6R    0.00000
+ATOM H9   HO     0.40000
+ATOM H10  HA     0.13000
+ATOM H11  HA     0.13000
+ATOM O2   OT2   -0.65000
+GROUP
+ATOM C19  CT    -0.10000
+ATOM C20  CT    -0.15000
+ATOM C21  CT    -0.15000
+ATOM H1   HA     0.05000
+ATOM H2   HA     0.05000
+ATOM H3   HA     0.05000
+ATOM H4   HA     0.05000
+ATOM H5   HA     0.05000
+ATOM H6   HA     0.05000
+ATOM H7   HA     0.05000
+ATOM H8   HA     0.05000
+ATOM N3   NT     0.00000
+GROUP
+ATOM N1   N6R   -0.45000
+ATOM C7   CR56   0.35000
+ATOM C8   CR56   0.00000
+ATOM C9   C6R   -0.13000
+ATOM H12  HA     0.13000
+GROUP
+ATOM C10  CUA1   0.05000
+ATOM N2   NX    -0.20000
+ATOM C11  CT     0.00000
+ATOM H13  HA     0.05000
+GROUP
```

```
+ATOM C12  CUA1  -0.10000
+ATOM C13  CUA1   0.00000
+ATOM C14  CUA1   0.00000
+ATOM C15  C      0.60000
+ATOM O3   O     -0.55000
+ATOM H14  HA     0.10000
+GROUP
+ATOM C16  CT     0.25000
+ATOM C17  C      0.70000
+ATOM O1   OE    -0.33000
+ATOM C18  CT     0.08000
+ATOM O4   O     -0.55000
+ATOM O5   OT2   -0.65000
+ATOM H15  HA     0.05000
+ATOM H16  HA     0.05000
+ATOM H17  HO     0.40000
+GROUP
+ATOM C22  CT    -0.10000
+ATOM H18  HA     0.05000
+ATOM H19  HA     0.05000
+GROUP
+ATOM C23  CT    -0.15000
+ATOM H20  HA     0.05000
+ATOM H21  HA     0.05000
+ATOM H22  HA     0.05000
+ATOM H23  HA     0.05000
+BOND  C1   C2   C1   C6   C1   O2
+BOND  C2   C3   C2   H10
+BOND  C3   C4   C3   H11
+BOND  C4   C5   C4   N1
+BOND  C5   C6   C5   C9
+BOND  C6   C19
+BOND  N1   C7
+BOND  C7   C8   C7   C10
+BOND  C8   C9   C8   C11
+BOND  C9   H12
+BOND  C10  N2   C10  C12
+BOND  N2   C11  N2   C15
+BOND  C11  H13  C11  H23
+BOND  C12  C13  C12  H14
+BOND  C13  C14  C13  C16
+BOND  C14  C15  C14  C18
+BOND  C15  O3
+BOND  C16  C17  C16  O5   C16  C22
+BOND  C17  O1   C17  O4
+BOND  O1   C18
+BOND  C18  H15  C18  H16
+BOND  O2   H9
+BOND  C19  N3   C19  H7   C19  H8
+BOND  N3   C20  N3   C21
+BOND  C20  H1   C20  H2   C20  H3
+BOND  C21  H4   C21  H5   C21  H6
+BOND  O5   H17
+BOND  C22  C23  C22  H18  C22  H19
+BOND  C23  H20  C23  H21  C23  H22
+
+
 RESI ALA          0.00
 GROUP
```

```
 ATOM N    NH1    -0.47  !     |
@@ -1220,6 +1327,69 @@
 IC CD1  CE2  *NE1 HE1   1.3752 108.8100  177.7800 124.6800  0.9767
 IC CG   NE1  *CD1 HD1   1.3679 110.1000  178.1000 125.4300  1.0820

+
+RESI NOH          0.00
+GROUP
+ATOM N    NH1    -0.47  !     |        HD1  HE1
+ATOM HN   H       0.31  !  HN-N         |    |
+ATOM CA   CT1     0.07  !     |   HB1  CD1--CE1
+ATOM HA   HB      0.09  !     |   |   //      \\
+GROUP                   !  HA-CA--CB--CG      CZ--OH
+ATOM CB   CT2    -0.18  !     |   |    \  __  /
+ATOM HB1  HA      0.09  !     |   HB2  CD2--CE2
+ATOM HB2  HA      0.09  !   O=C          |    |
+GROUP                   !     |        HD2  HE2
+ATOM CG   CA      0.00
+GROUP
+ATOM CD1  CA     -0.115
+ATOM HD1  HP      0.115
+GROUP
+ATOM CE1  CA     -0.115
+ATOM HE1  HP      0.115
+GROUP
+ATOM CZ   CA      0.11
+ATOM OH   ON2    -0.58
+GROUP
+ATOM CD2  CA     -0.115
+ATOM HD2  HP      0.115
+GROUP
+ATOM CE2  CA     -0.115
+ATOM HE2  HP      0.115
+GROUP
+ATOM C    C       0.51
+ATOM O    O      -0.51
+BOND CB  CA   CG  CB   CD2 CG   CE1 CD1
+BOND CZ  CE2  OH  CZ
+BOND N   HN   N   CA    C   CA   C   +N
+BOND CA  HA   CB  HB1  CB  HB2  CD1 HD1  CD2 HD2
+BOND CE1 HE1  CE2 HE2
+DOUBLE   O   C   CD1 CG  CE1  CZ  CE2 CD2
+IMPR N -C CA HN  C CA +N O
+DONOR HN N
+ACCEPTOR OH
+ACCEPTOR O C
+IC -C   CA   *N   HN    1.3476 123.8100  180.0000 114.5400  0.9986
+IC -C   N    CA   C     1.3476 123.8100  180.0000 106.5200  1.5232
+IC N    CA   C    +N    1.4501 106.5200  180.0000 117.3300  1.3484
+IC +N   CA   *C   O     1.3484 117.3300  180.0000 120.6700  1.2287
+IC CA   C    +N   +CA   1.5232 117.3300  180.0000 124.3100  1.4513
+IC N    C    *CA  CB    1.4501 106.5200  122.2700 112.3400  1.5606
+IC N    C    *CA  HA    1.4501 106.5200 -116.0400 107.1500  1.0833
+IC N    CA   CB   CG    1.4501 111.4300  180.0000 112.9400  1.5113
+IC CG   CA   *CB  HB1   1.5113 112.9400  118.8900 109.1200  1.1119
+IC CG   CA   *CB  HB2   1.5113 112.9400 -123.3600 110.7000  1.1115
+IC CA   CB   CG   CD1   1.5606 112.9400   90.0000 120.4900  1.4064
+IC CD1  CB   *CG  CD2   1.4064 120.4900 -176.4600 120.4600  1.4068
+IC CB   CG   CD1  CE1   1.5113 120.4900 -175.4900 120.4000  1.4026
```

```
+IC CE1  CG   *CD1 HD1   1.4026 120.4000  178.9400 119.8000  1.0814
+IC CB   CG   CD2  CE2   1.5113 120.4600  175.3200 120.5600  1.4022
+IC CE2  CG   *CD2 HD2   1.4022 120.5600 -177.5700 119.9800  1.0813
+IC CG   CD1  CE1  CZ    1.4064 120.4000   -0.1900 120.0900  1.3978
+IC CZ   CD1  *CE1 HE1   1.3978 120.0900  179.6400 120.5800  1.0799
+IC CZ   CD2  *CE2 HE2   1.3979 119.9200 -178.6900 119.7600  1.0798
+IC CE1  CE2  *CZ  OH    1.3978 120.0500 -178.9800 120.2500  1.4063
+
+
 RESI TYR          0.00
 GROUP
 ATOM N    NH1    -0.47  !     |        HD1  HE1
@@ -2263,7 +2433,104 @@
 BILD C4’  O5’  *C5’ H5’    0.0       0.0  -115.0     0.0    0.0
 BILD C4’  O5’  *C5’ H5’’   0.0       0.0   115.0     0.0    0.0

-
+RESI ADA         -1.50  !             H61  H62!
+ATOM P    P       1.00  !               \  /
+ATOM O1P  ON3    -0.68  !                N6
+ATOM O2P  ON3    -0.68  !                |
+ATOM O5’  ON2    -0.57  !                C6
+ATOM C5’  CN8B   -0.08  !               //  \
+ATOM H5’  HN8     0.09  !               N1   C5--N7\\
+ATOM H5’’ HN8     0.09  !               |    ||     C8-H8
+GROUP                   !               C2   C4--N9/
+ATOM C4’  CN7     0.16  !              / \\ /     \
+ATOM H4’  HN7     0.09  !            H2   N3       \
+ATOM O4’  ON6B   -0.50  !                           \
+ATOM C1’  CN7B    0.16  !                            \
+ATOM H1’  HN7     0.09  !                             \
+GROUP                   !         O1P    H5’ H4’  O4’  \
+ATOM N9   NN2    -0.05  !          |      |    \ /   \  \
+ATOM C5   CN5     0.28  !         -P-O5’-C5’---C4’    C1’
+ATOM N7   NN4    -0.71  !          |      |     \     / \
+ATOM C8   CN4     0.34  !         O2P    H5’’   C3’--C2’ H1’
+ATOM H8   HN3     0.12  !                      / \   / \
+ATOM N1   NN3A   -0.74  !                  O3’ H3’ O2’ H2’’
+ATOM C2   CN4     0.50  !                   |       |
+ATOM H2   HN3     0.13  !                          H2’
+ATOM N3   NN3A   -0.75
+ATOM C4   CN5     0.43
+ATOM C6   CN2     0.46
+ATOM N6   NN1    -0.77
+ATOM H61  HN1     0.38
+ATOM H62  HN1     0.38
+GROUP
+ATOM C2’  CN7B    0.14
+ATOM H2’’ HN7     0.09
+ATOM O2’  ON5    -0.66
+ATOM H2’  HN5     0.43
+GROUP
+ATOM C3’  CN7     0.01
+ATOM H3’  HN7     0.09
+ATOM O3’  ON2    -0.57
+BOND P    O1P       P    O2P       P     O5’
+BOND O5’  C5’       C5’  C4’       C4’   O4’       C4’  C3’       O4’  C1’
+BOND C1’  N9        C1’  C2’       N9    C4        N9   C8        C4   N3
+BOND C2   N1        C6   N6
```

```
+BOND N6   H61       N6   H62       C6   C5        C5   N7
+BOND C2'  C3'       C2'  O2'       O2'  H2'       C3'  O3'       O3'  +P
+BOND C1'  H1'       C2'  H2''      C3'  H3'       C4'  H4'       C5'  H5'
+BOND C5'  H5''      C8   H8        C2   H2
+DOUBLE  N1  C6      C2   N3        C4   C5        N7   C8
+IMPR N6   C6   H61  H62        C6   N1   C5   N6
+DONO H61  N6
+DONO H62  N6
+DONO H2'  O2'
+ACCE N3
+ACCE N7
+ACCE N1
+ACCE O1P  P
+ACCE O2P  P
+ACCE O2'
+ACCE O3'
+ACCE O4'
+ACCE O5'
+BILD -O3' P    O5'  C5'    1.6001  101.45  -46.90  119.00   1.4401 !alpha
+BILD -O3' O5'  *P   O1P    1.6001  101.45 -115.82  109.74   1.4802
+BILD -O3' O5'  *P   O2P    1.6001  101.45  115.90  109.80   1.4801
+BILD  P   O5'  C5'  C4'    1.5996  119.00 -146.00  110.04   1.5160 !beta
+BILD O5'  C5'  C4'  C3'    1.4401  108.83   60.00  116.10   1.5284 !gamma
+BILD C5'  C4'  C3'  O3'    1.5160  116.10  140.00  115.12   1.4212 !delta
+BILD C4'  C3'  O3'  +P     1.5284  111.92  155.00  119.05   1.6001 !epsilon
+BILD C3'  O3'  +P   +O5'   1.4212  119.05  -95.20  101.45   1.5996 !zeta
+BILD O4'  C3'  *C4' C5'    1.4572  104.06 -120.04  116.10   1.5160
+BILD C2'  C4'  *C3' O3'    1.5284  100.16 -124.08  115.12   1.4212
+BILD C4'  C3'  C2'  C1'    1.5284  100.16  -30.00  102.04   1.5251 !puck
+BILD C3'  C2'  C1'  N9     1.5284  101.97  147.80  113.71   1.4896
+BILD O4'  C1'  N9   C4     1.5251  113.71  -97.2   125.59   1.3783 !chi
+BILD C1'  C4   *N9  C8     1.4896  125.97 -179.94  106.0    1.367
+BILD C4   N9   C8   N7     1.376   106.0     0.0   113.6    1.312
+BILD C8   N9   C4   C5     1.367   106.0     0.0   105.6    1.382
+BILD C8   N7   C5   C6     0.0       0.0   180.0     0.0    0.0
+BILD N7   C5   C6   N1     0.0       0.0   180.0     0.0    0.0
+BILD C5   C6   N1   C2     0.0       0.0     0.0     0.0    0.0
+BILD N9   C5   *C4  N3     1.376   105.6  -180.0   126.9    1.342
+BILD C5   N1   *C6  N6     1.409   117.6  -180.0   121.2    1.337
+BILD N1   C6   N6   H61    1.337   121.2     0.0   119.0    1.01
+BILD H61  C6   *N6  H62    1.01    119.0   180.0   119.00   1.01
+BILD C5   N1   *C6  N6     1.409   117.6  -180.0   119.0    1.337
+BILD N1   C6   N6   H61    1.337   119.0     0.0   119.0    1.01
+BILD H61  C6   *N6  H62    1.01    119.0   180.0   121.00   1.01
+BILD N9   N7   *C8   H8    0.0       0.0   180.0     0.0    0.0
+BILD N1   N3   *C2   H2    0.0       0.0   180.0     0.0    0.0
+BILD C1'  C3'  *C2' O2'    1.5284  102.04 -114.67  110.81   1.4212
+BILD H2'  O2'  C2'  C3'    0.9600  114.97  148.63  111.92   1.5284
+BILD O4'  C2'  *C1' H1'    0.0       0.0  -115.0     0.0    0.0
+BILD C1'  C3'  *C2' H2''   0.0       0.0   115.0     0.0    0.0
+BILD C2'  C4'  *C3' H3'    0.0       0.0   115.0     0.0    0.0
+BILD C3'  O4'  *C4' H4'    0.0       0.0  -115.0     0.0    0.0
+BILD C4'  O5'  *C5' H5'    0.0       0.0  -115.0     0.0    0.0
+BILD C4'  O5'  *C5' H5''   0.0       0.0   115.0     0.0    0.0
+
+
 RESI ADE         -1.00  !             H61  H62!
 ATOM P    P       1.50  !               \  /
 ATOM O1P  ON3    -0.78  !                N6
```