Potentials of Mean Force as a Starting Point for Understanding Biomolecular Interactions

by

K. Maria Mills

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biophysics)
in The University of Michigan
2010

Doctoral Committee:

Associate Professor Ioan Andricioaei, Co-Chair, University of California, Irvine
Professor Mark M. Banaszak Holl, Co-Chair
Professor Gordon M. Crippen
Assistant Professor Jennifer P. Ogilvie

ACKNOWLEDGEMENTS

I would like to thank my advisor, Ioan Adricioaei, as well as the members of the Andricioaei research group, both past and present, for their support and guidance during my time as a graduate student. I would especially like to thank Jeff Wereszczynski, Aaron Frank, Gavin Bascom, Prasanth Jose, and Pu Tian. I would additionally like to thank my collaborators at the University of Michigan, Professors Mark Banaszak Holl and Brad Orr and their lab members for many productive discussions and suggestions, and for providing me with a desk for the months preceding my move to Irvine. I would also like to thank my collaborators from the University of California Irvine, Professor Zhibin Guan and his student Dora Guzman for sharing their AFM data. My acknowledgements also to my committee members, Professors Banaszak Holl, Crippen, and Ogilvie for their support and assistance. Thank you to the the Molecular Biophysics Training Grant, the NSF, and National Institute of Biomedical Imaging and Biomedical Engineering for financial support, and to the Pittsburgh Supercomputing Center and Department of Energy National Energy Research Scientific Computing Center for providing computing resources, without which the research presented here would not have been possible. I would also like to thank my family, especially my parents, who always went out of their way to make sure I had as many educational opportunities as possible and encouraged my strange pursuits, and every teacher I've ever had, from grade school to grad.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

CHAPTER I

Introduction

1.1 Background

1.1.1 Perspective of Biophysics

For most of its history, the study of biology has been a matter of qualitative analysis. Unlike physics, which seeks to derive the rules that define our universe, and chemistry, which is primarily concerned with the fundamental properties of atoms and molecules, biology involves systems of such complexity and variability that quantitative analysis is difficult. These complex systems, however, exist within the confines of the laws of physics, and at their most basic involve interactions between atoms and molecules. As scientific historian M. Vol'kenshtein. has put it, "it is not clear what is the center of all sciences. But physics lies in the depth of biology [1]." It follows that biology can be defined in terms of physics and chemistry, and therefore that biology can be described quantitatively. Advances in our understanding of biomolecular systems and in technology, as well as interdisciplinary research programs, have made such analysis possible for all levels of biology, from molecular biology to the organization of ecosystems. The questions being asked by biologists are no longer so general as; "What sequence of events leads to the folding of a polypeptide chain into a three-dimensional structure?" or, "What is the nature of the interaction between proteins and DNA?" but "What precise balance of forces

are responsible for the sequence of events in protein folding and the final structure obtained?" and "What are the energetic contributions to the molecular interactions between DNA and proteins?"

For molecular biology, this emphasis on physics and chemistry means, ultimately, studying the complex interplay between each atom within a system, which consists not only of the biological molecules of interest, but also water molecules and ions. Understanding not only the structure of biomolecules, but also their dynamic motions, and the effect on both these properties on their function is now central to modern studies of biology. Sophisticated biophysical techniques such as X-ray crystallography [2, 3, 4, 5] and Nuclear Magnetic Resonance spectroscopy [6, 7, 8, 9, 10] have made it possible to determine the atomic structures of biomolecules such as proteins and nucleic acids, starting with the famous "photograph 51," which led to the determination of the structure of DNA in 1953 [11]. Further, with NMR it is possible to determine the bulk dynamics of individual atoms [12, 13, 14, 15, 16]. Other techniques which similarly exploit the physical properties of molecules [17, 18, 19, 20] have made it possible to follow the motion of one or more degrees of freedom of molecules in bulk, and more recently of individual molecules [21, 22, 23, 24, 25, 26, 27, 28], allowing for the calculation of accurate probability distributions.

1.1.2 Potentials of Mean Force

One of the most commonly used quantitative descriptions in biophysics is the potential of mean force (PMF) [29]. The potential of mean force, is the potential which gives the average force over all coordinates. It is also, essentially, the free energy profile along a coordinate. Potentials of mean force are therefore a useful way to describe the behavior of complicated systems in a straightforward and statistically

rigorous manner. The PMF is defined as

$$\mathrm{G}(\xi) = -k_B T \ln\langle p(\xi)\rangle \quad (1.1)$$

where $\langle p(\xi)\rangle$ is the Boltzmann probability distribution with respect to a reaction coordinate of interest, ξ, defined as

$$\langle p(\xi)\rangle = \int dR\delta\xi'(R) - \xi)e^{-U(R)/k_B T} \times [\int dR e^{-U(R)/k_B T}]^{-1} (1.2)$$

[29]. The reaction coordinate can be any coordinate of interest, such as a distance or angle or the potential energy of the system. Potentials of mean force can be calculated from both simulations and experiments run under equilibrium conditions by histogramming the values of the chosen reaction coordinate to obtain an approximation of the probability distribution [29, 30, 24]. Similarly, equilibrium PMFs can be calculated from non-equilibrium systems by calculating the distribution function for the work put into the system [31]. Alternatively, it can also be used to describe non-equilibrium processes from the solution of a reduced Fokker-Planck equation for the time-dependence of the probability distribution of the ξ variable diffusing in the potential of mean force $G(\xi)$ (describing, e.g., the transfer of an ion or a conformational transition of a biomolecule). Therefore, if the chosen coordinate is a good reaction coordinate (i.e., if it is much slower than any other degree of freedom), a dynamical propagation on the PMF can simulate the kinetics of the reaction of interest.

PMFs can be used to represent the energetics of a range of biological systems of interest, such as protein folding and unfolding, interactions between molecules, and conformational changes within a molecule. Potentials of mean force are commonly

used in computational studies of processes ranging from hydrophobic interactions [32] or organic reactions in water [33], to proton transfer [34] or ionic permeation through membrane channels [35], to peptide [36] and protein [37] equilibria, to nucleic acid base flipping [38] or more complex conformational changes [39] in DNA and RNA. Recently, advances in experimental techniques have made it possible to determine such profiles for a real system, providing a useful basis for direct comparison between simulations and experiments. PMFs have been calculated from experiments for the extension of the mechanical protein domain titin I27 under force [40] and the membrane protein bacteriorhodopsin [41], as well as for the distance between the N-terminus and loop of chymotrypsin inhibitor 2 [42]. Potentials of mean force can be used to clearly represent the equilibrium distribution of one or a few relevant conformational variables with increasing accuracy as longer sampling and improved force fields and more sophisticated experimental techniques become available.

1.1.3 Molecular Dynamics Simulations

Although potentials of mean force represent biomolecular reactions in a clear, quantitative manner, they do so by reducing complex processes to an easily understandable one or two-dimensional free energy function. Biomolecular systems, however, contain many degrees of freedom which contribute to their behavior. The nuances of these individual contributions are necessarily lost when the system is reduced to a single reaction coordinate. Furthermore, though in some cases biological processes can be represented accurately with a single reaction coordinate, it is more common that not just nuanced, but vital, information is lost in such simplifications. In order to fully understand the behavior of biological molecules, it is best to keep track not just of a single coordinate, but ideally of every atom within the system.

The goal of observing the motions of a single macromolecule in atomistic detail

is one that experimental techniques have thus far failed to achieve. It is, however,theoretically possible to predict how molecules will behave based on the fundamental principles of physics. Such calculations are the underlying idea behind molecular dynamics simulations [43, 44, 45, 46, 47]. Molecular Dynamics simulations have been in use since the 1950s, when simulations of simple liquids were run using hard sphere models [48, 49, 50, 51]. By the 1960s, more sophisticated models of molecular liquids were in use [52, 53, 54], and by the 1970s molecular dynamics simulations had been expanded to include proteins [55]. At present, atomistic simulations are being used to study increasingly complex systems, including proteins [47], nucleic acids [56], membranes [57], and nanoparticles [58, 59]. The great advantage of MD is that, unlike experiments, information can be obtained for all degrees of freedom of the system. Since the physical equations of motion and interaction are calculated for all atoms in the system, dynamic, structural, and energetic information can be obtained for not only the system as a whole, but also for each of its components.

Molecular dynamics simulations have been used extensively in the past few decades to study biomolecular systems, as the atomistic details provided by simulations can be used to explain and clarify experimental results. They have proven a great resource for aiding in the interpretation of a wide variety of experiments, including spectroscopic data [60], NMR parameters [61], and X-ray crystallography [62, 63]. Simulations allow for the determination of structural stability and dynamics of biolomolecules, the free energy differences and reaction pathways between states. Additionally, because the potential energy function in simulations is under the control of the user, they can be used to create systems which are not possible in nature, for example by removing a single term from the energy function [64], or by running different parts of a system at different temperatures [65, 66].

Molecular dynamics simulations of biological molecules have been used to determine the pathway between known states of a system, such as the transition from the open to closed state of the chaperonin protein GroEL [67]. Simulations have also revealed the importance of dynamics to the function of many biological systems, leading to insight into the role of gate-fluctuations in the activity of acetylcholinesterase [68], for example, and the effect of internal motions on the electron transfer mechanism of photosynthetic proteins [69, 70]. In some cases these findings have been validated by later experiments [71]. The mechanism of protein folding, a central question in molecular biology, has also been studied extensively with computational methods. In most cases the use of indirect methods such as coarse-grained models [72], has been necessary, since the timescales of protein folding are far beyond the reach of what is usually computationally feasible. However, several recent studies have taken advantage of shared computing to directly simulate the folding of small, fast-folding protein domains such as the villin headpiece [73].

Improvements in computing power and the development of scalable programs such as NAMD [74] and Desmond [75], which allow simulations to be run efficiently over multiple processors, have made it possible to study increasingly complex systems over longer timescales. Whereas a decade ago the simulation limit was in the order of hundreds of picoseconds to a few nanoseconds for relatively small systems, such as a solvated protein domain, it is now possible to run simulations on the order of microseconds to miliseconds, of much larger systems such as proteins embedded in membranes[76, 77]. Alternatively, modern technology has also made it possible to run a large ensemble of short timescale simulations for the calculation of statistical variables without the need for special techniques to enhance sampling. It is to be expected that ongoing advances in both computer power and computational techniques

will continue to increase the utility of MD simulations.

Although MD simulations have improved dramatically in both efficiency and accuracy over the past 50 years, there is still a need to be cautious when designing and interpreting simulations [78]. Although the parameters for molecular dynamics force fields have been designed to match experimental data, there are still may be instances in which simulations fail to replicate experimental results. It is wise, therefore, to design simulations with experimental techniques in mind, so that the results of simulations can be compared directly to experiment. It is also possible, in some cases, to use experimental results to anticipate possible problems with simulations, in which case simulations can be designed with experimental input to minimize errors.

1.1.4 Single Molecule Experiments

Just as advances in computing power and the development of statistical techniques have increased the utility of molecular dynamics simulations, so have technological advances and innovative methods added significant improvements to the types of information that can be detected through experiments. While molecular dynamics simulations have always been "single molecule" experiments, until recently, physical experiments have only been able to access information for bulk systems. Advances in techniques and optical equipment have made it possible to finally look at individual molecules. Although it is not possible to follow the behavior of all the individual atoms of large molecules such as proteins and nucleic acids, it is possible to measure broad conformational properties of individual *molecules* [79].

Scanning tunnel microscopes and atomic force microscopes allow for the detection of structural features of large molecules by bringing a very fine probe tip into contact with a molecule and measuring the tunnel current [80, 81, 82] or deflection of the tip [83], respectively. Optical methods have also advanced to the point where it

is possible to detect emission or absorption spectra of individual molecules, thanks to improvements in both light sources and detectors [21, 22, 23, 24, 25, 26, 27]. In particular, single molecule fluorescence energy resonance transfer (FRET), and single molecule force manipulation studies have advanced the study of biomolecules by detecting the structural, dynamic, and elastic properties of molecules.

In FRET experiments, two complementary fluorophores, a donor and acceptor, are conjugated to the systems of interest. The donor fluorophore is excited by an external light source. If the two fluorophores are in close proximity to each other, then the energy emitted by the donor will be transferred to the acceptor rather than released as a photon. The amount of energy transferred is determined as a function of the proportion of acceptor emissions versus donor emissions. The distance between the fluorophores can then be calculated, as it is fundamentally related to the amount of energy transferred [84]. With advanced optical methods, it is possible to measure the distance between a single pair of fluorophores on the nm scale [85]. Single pair FRET can be used to follow the dynamics of a single molecule over time by trapping molecules spatially [86, 87, 88, 89] or to determine the distribution of states for individual molecules freely diffusing in solution [90, 91, 92, 93]. Such experiments have been used to study both intramolecular dynamics and intermolecular interactions, including protein folding and unfolding [94, 91, 95, 96, 97, 98, 99, 100, 101, 94], protein and nucleic acid dynamics [102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112], receptor-ligand interactions [108], and enzyme-ligand interactions [113, 114].

Force pulling studies [115, 28, 116], on the other hand, involve the physical manipulation of a single molecule using forces on the sub-pN scale. Such experiments involve the use of atomic force microscopy [117], optical tweezers [118, 119, 120, 121], magnetic tweezers [122, 123], or flow induced stretching [20, 124] to control the behav-

ior of molecules. In most cases, a force is applied to one end of a molecule, which has its other end fixed to a surface, via the probe, and the displacement of the molecule as a function of force is measured [115, 28]. This type of experiment has been used to study protein unfolding and folding [125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137], DNA condensation by cations [138], enzyme kinetics [139], and nucleic acid unfolding [140]. In terms of intermolecular interactions, they have been used to study ligand-receptor interactions [141, 142] , protein-protein interactions [143, 144], and protein-nucleic acid interactions [145, 118], as well as the step-size, stall force, and processivity of molecular motors [146, 147]. Force manipulation experiments can be used to measure the strength of (and distinguish between) attractive van der Waal forces, repulsive electrostatic forces, repulsive hydration forces, solvation forces, adhesion, and hydrophobic forces [116]. It is also possible with these techniques to measure the strength of individual hydrogen bonds and some covalent bonds [148].

Unlike bulk experiments , single molecule fluorescence and manipulation experiments allow for the calculation of accurate probability or work distributions, which in turn can be used to determine potentials of mean force for the measured coordinate (either the distance between fluorophores, in the case of FRET, or the extension of the molecule in force spectroscopy). They are therefore uniquely suited to comparison with MD simulations. This complementation has advantages for both sides. Experiments are able to access time scales unavailable to simulations, and can therefore yield more accurate data about the equilibrium state of a system, whereas simulations can provide a level of detail far beyond what is possible to observe experimentally.

1.2 Computational Techniques

The work presented here aims to use molecular dynamics simulations and theoretical tools to study the free energies of biomolecular systems. In particular, this dissertation focuses on the overlap between experiment and simulation, and the development of methods which use experimental information to inform the design of simulations to improve their accuracy, as well as to show the ways in which simulations can be used to supplement experimental discoveries. A variety of computational techniques were used in this research. Primarily, molecular dynamics simulations using the CHARMM force field were used to obtain free energy data for peptide dynamics, protein unfolding, and DNA condensation by poly(amidoamine) dendrimers. The computational methods used in the studies presented in this work are described below.

1.2.1 The CHARMM Force Field

One of the oldest and most widely used programs for MD simulations is the CHARMM program developed by the Karplus lab in the 70s [149, 150]. As a compromise between computational feasibility and physical accuracy, CHARMM uses classical approximations of potential energy functions, although it is also capable of Quantum Mechanical calculations. Atoms are modeled as spheres with van der Waals radii and point charges. The current CHARMM potential energy function is a combination of internal terms plus nonbonded interactions. It is defined as:

$$\begin{aligned} V(r) = & \sum_{binds} k_b(b-b_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} k_\chi(1+cos(n\chi-\delta)) \\ & + \sum_{impropers} k_{\psi)(\psi-\psi_0)^2+} \sum_{Urey-Bradley} k_{UB}(S-S_0)^2 \\ & + \sum_{nonbonded} (\epsilon_i j[(\frac{R_m in,ij}{r_i j})^1 2 - (\frac{R_m in,ij}{r_i j})^6] + \frac{q_i q_j}{\epsilon_d r_i j}) \end{aligned} \tag{1.3}$$

All internal terms are treated as harmonic, with the exception of dihedral angles. The nonbonded terms consist of electrostatic interactions between point charges and van der Waals repulsive and attractive interactions, which are approximated with a Leonard-Jones potential [56, 43, 44, 45, 47]. Because of the expense of the calculations, long-range nonbonded interactions are often cut off at a user-specified distance or, in the case of electrostatics, treated explicitly within a certain distance and calculated based on points on a cubic lattice beyond that distance [151, 152].

The initial CHARMM potential for proteins was an extended atom model which lacked explicit hydrogen atoms and required an additional term for hydrogen bonds. In the 1980s, the parameter set 19 was developed [153]. In this version of the CHARMM potential, polar hydrogens were added and the hydrogen bonding term was dropped. This parameter set is well suited for simulations in implicit solvent, in which water is modeled by an additional energy term rather than added explicitly to the system [154]. The current version of CHARMM uses an all-atom model. The parameters of this model have been developed to match experimental data on the structural and thermodynamics properties. All-atom parameter sets have been developed for proteins [155, 156], nucleic acids [157, 158], lipids [159], and some carbohydrates [160, 161]. More advanced polarizable models have also been developed, although they are not in wide use due to the high computational cost [162].

1.2.2 Umbrella Sampling

At present, it is still too computationally expensive to simulate a statistical ensemble of large systems on biologically relevant timescales. Molecular dynamics simulations are limited to the nanosecond time scale, and the presence of significant free energy barriers in many biological systems of interest makes sampling of the full conformational space difficult. Numerous techniques have been developed to circum-

vent this fundamental problem of computational studies. Among the oldest and most widely used is the umbrella sampling method. Umbrella sampling relies on the idea that restraining a system to a specific region of its conformational space will result in the system sampling the the full potential energy surface of that region. Harmonic potentials are added to the Hamiltonian of the system to constrain the reaction coordinate to a specific region of its conformational space. By running a series of such restrained simulations in which each window explores a different region, the entire conformational space will be efficiently sampled and the free energy profile of the restrained coordinate can be reconstructed by mathematically removing the effect of the restraining potential, usually using the weighted histogram analysis method [163]. The method is generally used to improve sampling along one or two specific reaction coordinates. The restrained coordinate can be nearly any feature of the system, such as a distance, dihedral, or even the potential energy. The usefulness of the umbrella sampling technique requires an appropriate choice of restraining coordinate, one that correlates to the function being studied and which is ideally the slowest degree of freedom. In complex systems, perfect sampling over finite timescales is unlikely even with the advantages of this method. Nevertheless, the technique is a powerful one for the improvement of statistical sampling and simulations efficiency.

1.2.3 Steered Molecular Dynamics

Free energy profiles can also be obtained from force manipulation simulations, commonly referred to as steered molecular dynamics or SMD, in which one end of the system is held fixed while the other is pulled at constant velocity or with constant force. These simulations mimic experimental studies in which an optical or magnetic trap, or atomic force microscope cantilever is used to apply a unidirectional force on a molecule, although much higher forces are required in simulations than in

experiments, in order to observe most processes in a computationally accessible time scale. In 1997, Jarzynski derived an equality for calculating the equilibrium free energy profile along the pulling coordinate from such experiments using the work put into the system [31]. This method is not unlike umbrella sampling, in that an external force is used to manipulate the system and force it to explore conformations that would not normally be seen in straight forward simulations. As in umbrella sampling, the removal of this bias allows for the determination of the equilibrium free energy.

1.2.4 Monte Carlo Calculations

Monte Carlo methods are a class of stochastic computational calculations that have been used to investigate a wide range of processes [164, 165, 166]. The input for Monte Carlo calculations is generated from random numbers pulled from a probability distribution, with the goal of simulating sampling of a real population. The probability of an event occurring at each time step is calculated, and then a random number is used to "choose" whether said event occurs or not. Monte Carlo techniques have been used in a variety of mathematical studies, including economics and physical systems, as well as in complex calculations of the behavior of biomolecular systems.

1.3 Specific Aims

This dissertation aims to demonstrate how molecular dynamics simulations can be used to improve understanding of biomolecular systems using experimental data as a means of both improving the accuracy of simulation results and of testing the validity of simulations. Potentials of mean force are employed as an ideal measure for bridging the gap between experiment and simulations.

1. In Chapter 2, a simple method for utilizing experimental data to improve the efficiency of numerical calculations of free energy profiles from molecular dynamics simulations is presented. It is shown that using umbrella sampling simulations with restraining potentials based on a known approximate estimate of the free energy profile derived solely from experimental data to "guide" the simulations results in an optimal initial restraining potentials and decreased overall computational time. In demonstration of the method, guided and unguided (regular) umbrella sampling simulations, as well as exhaustive sampling simulations are comparatively used to calculate the free energy profile for the distance between the ends of a pentapeptide. To showcase the method, guiding umbrella sampling potentials were based on a simulated "experimental" free energy profile of the end-to-end distance that would measured by FRET (Fluorescence Energy resonance Transfer, obtained from an exhaustive sampling of the conformations of the model peptide). Statistical analysis shows a dramatic improvement in efficiency for a 5-window guided umbrella sampling over 5- and 17-window unguided umbrella sampling simulations. Furthermore, the form of the potential of mean force for the guided umbrella sampling simulations evolves, as one approaches convergence, in the same way as the extensive simulations, but exponentially faster. Comparison of the dihedral angles of the peptide from the different simulations shows modest differences in their potentials of mean force for all methods. The method was further validated by replicating the forced unfolding pathway of the titin I27 domain using a guiding PMF determined from single molecule force spectroscopy. Comparison with unguided umbrella sampling reveals that the use of a guiding potential encourages the unfolding simulations to converge to a forced unfolding pathway that agrees with previous

results and produces a more accurate potential of mean force.

2. Force-pulling experiments on the unfolding of mechanical and non-mechanical protein domains have greatly increased our understanding of the structural stability of proteins. Because these experiments are done on the single molecule level, they also enable experimentalists to observe differences in the unfolding behavior of individual molecules. However, it is difficult to determine the source of any unfolding heterogeneity through experiments alone. In Chapter 3, evidence is presented from experiments and simulations that the β domain of Streptokinase, a non-mechanical protein, unfolds under force via three distinct pathways. High temperature SMD simulations were used to determine the source of the velocity-dependent heterogeneity observed in AFM force pulling experiments. It is demonstrated through the simulations that hydrophobic interactions in the core of the protein underlie the differences observed in experiments and contribute significantly to the structural stability of the protein under force. Using an expansion of the Jarzynski equality [167], multidimensional free energy surfaces are calculated to describe the energetics of the different pathways.

3. Polyamidoamine (PAMAM) dendrimers are biomedical nanoparticles with substantial promise as delivery vehicles for therapeutic nucleic acids. A description of the fundamental interactions between these molecules and DNA is essential to understanding their potential for such purposes. An atomistic study of the structural changes, energetics and driving forces of the interaction between DNA and dendrimers is detailed in Chapter 4. Using molecular dynamics simulations and free energy profile calculations, we have characterized interactions between DNA and generation 3 PAMAM dendrimers with 32 amine terminations and

dendrimers with a mixture of amine and acetamide terminations. A novel insight into dendrimer-DNA condensation revealed by the simulation shows that, in addition to the experimentally shown contraction of DNA on scales comparable to its persistence length, dendrimers may induce local bending in DNA on scales much smaller than the persistence length. Additionally, the dendrimer experiences a local, microscopic contraction on scales comparable to its size, driven by a change in the local electrostatic environment. Analysis of the energetic contribution to DNA-dendrimer complexation as well as water ordering between DNA and a dendrimer with positively charged terminations indicates that not only does complexation of the dendrimer with DNA affect the local water structure, but ordered water molecules facilitate long range interactions between the molecules and contribute significantly to the interaction free energy. The results of these simulations are compared to experimental results for DNA complexation with dendrimers, as well as experimental calculations of the contribution of attractive hydration forces to DNA condensation by polyamine cations. The simulations are shown to be in good agreement with the available experimental data.

4. The mechanism of DNA compaction by dendrimers is key to the design of nanotechnologies that can deliver genetic material into cells. In Chapter 5, mesoscopic modeling and single-molecule pulling experiments are used to further describe DNA dendrimer interactions in the context of large-scale DNA condensation. The energy, force and geometry parameters computed at the atomic level in the molecular dynamics simulations described in Chapter 4 are used in conjunction with a Monte Carlo model, yielding mesoscopic force-extension curves. Actual experimental single-molecule curves obtained with optical tweez-

ers are also presented and show remarkable agreement with the virtual curves from our model. The calculations reveal the microscopic origin of the hysteresis observed in the phase transition underlying compaction. A broad range of ionic and pulling parameters are sampled, and suggestions for windows of conditions to probe new single-molecule behavior are made.

BIBLIOGRAPHY

[1] M. V. Vol'kenshtein. Complementarity, physics, and biology. *Soviet Physics Usp.*, 2:140–150, 1988.

[2] J. Barraud. Principles of X-ray crystallography. [Principes de Radiocristallographie], pages xi–236, 1960.

[3] B. C. Giessen and G. E. Gordon. X-ray diffraction - New high-speed technique based on X-ray spectrography. *Science*, 159:973, 1968.

[4] L. Bragg. X-ray crystallography. *Scientific American*, 219:58–70, 1968.

[5] B. Strandberg, R. E. Dickerson, and M. G. Rossmann. 50 years of protein structure analysis. *Journal of Molecular Biology*, 392:2–32, 2009.

[6] J. S. Cohen and Jardetzk.O. Nuclear magnetic resonance studies of structure and binding sites of enzymes .2. spectral assignments and inhibitor binding in hen egg-white lysozyme. *Proceedings of The National Academy of Sciences of The United States of America*, 60:92–99, 1968.

[7] G. Wagner. NMR investigations of protein-structure. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 22:101–139, 1990.

[8] S. F. Ohandley, D. G. Sanford, R. Xu, C. C. Lester, B. E. Hingerty, S. Broyde, and T. R. Krugh. Structural characterization of an N-acetyl-2-aminofluorene (aaf) modified DNA oligomer by NMR, energy minimization, and molecular-dynamics. *Biochemistry*, 32:2481–2497, 1993.

[9] D. S. Wishart and B. D. Sykes. Chemical-shifts as a tool for structure determination. *Nuclear Magnetic Resonance, Pt C*, 239:363–392, 1994.

[10] G. Varani, F. Aboulela, and F. H. T. Allain. NMR investigation of RNA structure. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 29:51–127, 1996.

[11] J. D. Watson and F. H. C. Crick. The structure of DNA. *Cold Spring Harbor Symposia on Quantitative Biology*, 18:123–131, 1953.

[12] A. G. Palmer, J. Williams, and A. McDermott. Nuclear magnetic resonance studies of biopolymer dynamics. *Journal of Physical Chemistry*, 100:13293–13310, 1996.

[13] S. S. Wijmenga and B. N. M. van Buuren. The use of NMR methods for conformational studies of nucleic acids. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 32:287–387, 1998.

[14] R. Ishima and D. A. Torchia. Protein dynamics from NMR. *Nature Structural Biology*, 7:740–743, 2000.

[15] A. G. Palmer. NMR probes of molecular dynamics: Overview and comparison with other techniques. *Annual Review of Biophysics and Biomolecular Structure*, 30:129–155, 2001.

[16] A. Mittermaier and L. E. Kay. Review - new tools provide new insights in NMR studies of protein dynamics. *Science*, 312:224–228, 2006.

[17] B. Drake, C. B. Prater, A. L. Weisenhorn, S. A. C. Gould, T. R. Albrecht, C. F. Quate, D. S. Cannell, H. G. Hansma, and P. K. Hansma. Imaging crystals, polymers, and processes in water with the atomic force microscope. *Science*, 243:1586–1589, 1989.

[18] E. Betzig and J. K. Trautman. Near-field optics - microscopy, spectroscopy, and surface modification beyond the diffraction limit. *Science*, 257:189–195, 1992.

[19] R. Garcia and R. Perez. Dynamic atomic force microscopy methods. *Surface Science Reports*, 47:197–301, 2002.

[20] S. K. Sia and G. M. Whitesides. Microfluidic devices fabricated in poly(dimethylsiloxane) for biological studies. *Electrophoresis*, 24:3563–3576, 2003.

[21] S. M. Nie and R. N. Zare. Optical detection of single molecules. *Annual Review of Biophysics and Biomolecular Structure*, 26:567–596, 1997.

[22] S. Weiss. Fluorescence spectroscopy of single biomolecules. *Science*, 283:1676–1683, 1999.

[23] J. K. Gimzewski and C. Joachim. Nanoscale science of single molecules using local probes. *Science*, 283:1683–1688, 1999.

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

[25] T. A. Byassee, W. C. W. Chan, and S. Nie. Single molecule detection in single living cells. *Proceedings of the SPIE - The International Society for Optical Engineering*, vol.3922:2–10, 2000.

[26] F. Kulzer and M. Orrit. Single-molecule optics. *Annual Review of Physical Chemistry*, 55:585–611, 2004.

[27] E. Haustein and P. Schwille. Single-molecule spectroscopic methods. *Current Opinion in Structural Biology*, 14:531–540, 2004.

[28] Y. E. Pak, M. Marimuthu, and S. Kim. Nanomechanics of biomolecules: A review. *Biochip Journal*, 2:235–241, 2008.

[29] J.G. Kirkwood. Statistical mechanics of fluid mixtures. *Journal of Chemical Physics*, 3:300–313, 1935.

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

[31] C. Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78:2690–2693, 1997.

[32] L. R. Pratt and D. Chandler. Theory of hydrophobic effect. *Journal of Chemical Physics*, 67:3683–3704, 1977.

[33] J. L. Gao. A priori computation of a solvent-enhanced SN2 reaction profile in water - the Menshutkin reaction. *Journal of The American Chemical Society*, 113:7796–7797, 1991.

[34] K. Hinsen and B. Roux. Potential of mean force and reaction rates for proton transfer in acetylacetone. *Journal of Chemical Physics*, 106:3567–3577, 1997.

[35] T. W. Allen, O. Andersen, and B. Roux. Energetics of ion conduction through the gramicidin channel. *Biophysical Journal*, 86:351A–351A, 2004.

[36] B. M. Pettitt and M. Karplus. The potential of mean force surface for the alanine dipeptide in aqueous solution: A theoretical approach. *Chemical Physics Letters*, 121:194–201, 1985.

[37] E. M. Boczko and C. L. Brooks. Free-energy surface for the unfolding of a 48 residue 3 helix bundle protein. *Abstracts of Papers of The American Chemical Society*, 208:324–PHYS, 1994.

[38] N. Huang, N. K. Banavali, and A. D. MacKerell. Protein-facilitated base flipping in DNA by cytosine-5-methyltransferase. *Proceedings of The National Academy of Sciences of The United States of America*, 100:68–73, 2003.

[39] 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.

[40] N. C. Harris, Y. Song, and C. H. Kiang. Experimental free energy surface reconstruction from single-molecule force spectroscopy using Jarzynski's equality. *Physical Review Letters*, 99:068101–068104, 2007.

[41] J. Preiner, H. Janovjak, C. Rankl, H. Knaus, D. A. Cisneros, A. Kedrov, F. Kienberger, D. J. Muller, and P. Hinterdorfer. Free energy of membrane protein unfolding derived from single-molecule force measurements. *Biophysical Journal*, 93:930–937, 2007.

[42] N. G. Walter, D. A. Harris, M. J. B. Pereira, and D. Rueda. In the fluorescent spotlight: Global and local conformational changes of small catalytic RNAs. *Biopolymers*, 61:224–242, 2001.

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

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

[45] M. Karplus. Molecular dynamics simulations of biomolecules. *Accounts of Chemical Research*, 35:321–323, 2002.

[46] P. Entel, W. A. Adeagbo, M. Sugihara, G. Rollmann, A. T. Zavak, T. Kreth, and K. Kadau. Molecular dynamics simulations in biology, chemistry and physics. *Computational materials science. From basic principles to material properties (Lecture Notes in Phys. Vol.642)*, pages 177–206, 2004.

[47] M. Karplus and J. Kuriyan. Molecular dynamics and protein function. *Proceedings of The National Academy of Sciences of The United States of America*, 102:6679–6685, 2005.

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

[49] B. J. Alder and T. E. Wainwright. Studies in molecular dynamics .2. behavior of a small number of elastic spheres. *Journal of Chemical Physics*, 33:1439–1451, 1960.

[50] E. B. Smith and B. J. Alder. Perturbation calculations in equilibrium statistical mechanics .1. hard sphere basis potential. *Journal of Chemical Physics*, 30:1190–1199, 1959.

[51] M. P. Allen, D. Frenkel, and J. Talbot. Molecular-dynamics simulation using hard particles. *Computer Physics Reports*, 9:301–353, 1989.

[52] A. Rahman. Correlations in motion of atoms in liquid argon. *Physical Review A-General Physics*, 136:A405–A411, 1964.

[53] A. Rahman and F. Stilling. Molecular dynamics study of liquid water. *Journal of Chemical Physics*, 55:3336–3359, 1971.

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

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

[56] McCammon J A and Harvey S C. Dynamics of proteins and nucleic acids. *Cambridge University Press: Cambridge, England, Uk; New York, New York, USA. Illus*, pages XII–234, 1987.

[57] R. W. Benz, F. Castro-Roman, D. J. Tobias, and S. H. White. Experimental validation of molecular dynamics simulations of lipid bilayers: A new approach. *Biophysical Journal*, 88:805–817, 2005.

[58] P. M. R. Paulo, J. N. C. Lopes, and S. M. B. Costa. Molecular dynamics simulations of charged dendrimers: Low-to-intermediate half-generation pamams. *Journal of Physical Chemistry B*, 111:10651–10664, 2007.

[59] C. V. Kelly, P. R. Leroueil, E. K. Nett, J. M. Wereszczynski, J. R. Baker, B. G. Orr, M. M. B. Holl, and I. Andricioaei. Poly(amidoamine) dendrimers on lipid bilayers I: Free energy and conformation of binding. *Journal of Physical Chemistry B*, 112:9337–9345, 2008.

[60] T. Ichiye and M. Karplus. Fluorescence depolarization of tryptophan residues in proteins - a molecular-dynamics study. *Biochemistry*, 22:2884–2893, 1983.

[61] J. O. Wrabl, D. Shortle, and T. B. Woolf. Correlation between changes in nuclear magnetic resonance order parameters and conformational entropy: Molecular dynamics simulations of native and denatured staphylococcal nuclease. *Proteins-Structure Function and Genetics*, 38:123–133, 2000.

[62] P. J. Artymiuk, C. C. F. Blake, D. E. P. Grace, S. J. Oatley, D. C. Phillips, and M. J. E. Sternberg. Crystallographic studies of the dynamic properties of lysozyme. *Nature*, 280:563–568, 1979.

[63] L. Nilsson, G. M. Clore, A. M. Gronenborn, A. T. Brunger, and M. Karplus. Structure refinement of oligonucleotides by molecular-dynamics with nuclear overhauser effect interproton distance restraints - application to 5' d(CGTACG)2. *Journal of Molecular Biology*, 188:455–475, 1986.

[64] T. Simonson, G. Archontis, and M. Karplus. Free energy simulations come of age: Protein-ligand recognition. *Accounts of Chemical Research*, 35:430–437, 2002.

[65] D. Vitkup, D. Ringe, G. A. Petsko, and M. Karplus. Solvent mobility and the protein 'glass' transition. *Nature Structural Biology*, 7:34–38, 2000.

[66] A. L. Lee and A. J. Wand. Microscopic origins of entropy, heat capacity and the glass transition in proteins. *Nature*, 411:501–504, 2001.

[67] J. P. Ma, P. B. Sigler, Z. H. Xu, and M. Karplus. A dynamic model for the allosteric mechanism of GroEL. *Journal of Molecular Biology*, 302:303–313, 2000.

[68] K. Tai, T. Y. Shen, U. Borjesson, M. Philippopoulos, and J. A. McCammon. Analysis of a 10-ns molecular dynamics simulation of mouse acetylcholinesterase. *Biophysical Journal*, 81:715–724, 2001.

[69] M. Ben-Nun, F. Molnar, H. Lu, J. C. Phillips, T. J. Martinez, and K. Schulten. Quantum dynamics of the femtosecond photoisomerization of retinal in bacteriorhodopsin. *Faraday Discussions*, 110:447–462, 1998.

[70] I. A. Balabin and J. N. Onuchic. Dynamically controlled protein tunneling paths in photosynthetic reaction centers. *Science*, 290:114–117, 2000.

[71] N. A. Ranson, G. W. Farr, A. M. Roseman, B. Gowen, W. A. Fenton, A. L. Horwich, and H. R. Saibil. Atp-bound states of GroEL captured by cryo-electron microscopy. *Cell*, 107:869–879, 2001.

[72] P. J. Bond, J. Holyoake, A. Ivetac, S. Khalid, and M. S. P. Sansom. Coarse-grained molecular dynamics simulations of membrane proteins and peptides. *Journal of Structural Biology*, 157:593–605, 2007.

[73] J. Kubelka, T. K. Chiu, D. R. Davies, W. A. Eaton, and J. Hofrichter. Sub-microsecond protein folding. *Journal of Molecular Biology*, 359:546–553, 2006.

[74] 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.

[75] K. J. Bowers, E. Chow, Huageng Xu, R. O. Dror, M. P. Eastwood, B. A. Gregersen, J. L. Klepeis, I. Kolossvary, M. A. Moraes, F. D. Sacerdoti, J. K. Salmon, Y. Shan, and D. E. Shaw. Scalable algorithms for molecular dynamics simulations on commodity clusters. *SC 2006 Proceedings Supercomputing 2006*, pages 13 pp.—CD–ROM, 2006.

[76] P. W. Lau, A. Grossfield, S. E. Feller, M. C. Pitman, and M. F. Brown. Retinal flexibility of rhodopsin illuminated through large-scale molecular dynamics simulations. *Biophysical Journal*, pages 314A–314A, 2007.

[77] A. Grossfield, S. E. Feller, and M. C. Pitman. Convergence of molecular dynamics simulations of membrane proteins. *Proteins-Structure Function and Bioinformatics*, 67:31–40, 2007.

[78] W. F. van Gunsteren and A. E. Mark. Validation of molecular dynamics simulation. *Journal of Chemical Physics*, 108:6109–6116, 1998.

[79] A. N. Kapanidis and T. Strick. Biology, one molecule at a time. *Trends in Biochemical Sciences*, 34:234–243, 2009.

[80] U. Durig, D. W. Pohl, and F. Rohner. Near-field optical-scanning microscopy. *Journal of Applied Physics*, 59:3318–3327, 1986.

[81] G. Binnig and H. Rohrer. Scanning tunneling microscopy. *IBM Journal of Research And Development*, 30:355–369, 1986.

[82] J. A. Stroscio and D. M. Eigler. Atomic and molecular manipulation with the scanning tunneling microscope. *Science*, 254:1319–1326, 1991.

[83] J. Madl, S. Rhode, H. Stangl, H. Stockinger, P. Hinterdorfer, G. J. Schutz, and G. Kada. A combined optical and atomic force microscope for live cell investigations. *Ultramicroscopy*, 106:645–651, 2006.

[84] E. A. Jares-Erijman and T. M. Jovin. FRET imaging. *Nature Biotechnology*, 21:1387–1395, 2003.

[85] E. Toprak and P. R. Selvin. New fluorescent tools for watching nanometer-scale conformational changes of single molecules. *Annual Review of Biophysics and Biomolecular Structure*, 36:349–369, 2007.

[86] T. Ha, T. Enderle, D. F. Ogletree, D. S. Chemla, P. R. Selvin, and S. Weiss. Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor. *Proceedings of The National Academy of Sciences of The United States of America*, 93:6264–6268, 1996.

[87] Y. Ishii, T. Yoshida, T. Funatsu, T. Wazawa, and T. Yanagida. Fluorescence resonance energy transfer between single fluorophores attached to a coiled-coil protein in aqueous solution. *Chemical Physics*, 247:163–173, 1999.

[88] T. Ha. Single-molecule fluorescence resonance energy transfer. *Methods*, 25:78–86, 2001.

[89] J. Y. Shang and E. Geva. Extracting the time scales of conformational dynamics from single-molecule single-photon fluorescence statistics. *Journal of Physical Chemistry B*, 111:4220–4226, 2007.

[90] A. A. Deniz, M. Dahan, J. R. Grunwell, T. J. Ha, A. E. Faulhaber, D. S. Chemla, S. Weiss, and P. G. Schultz. Single-pair fluorescence resonance energy transfer on freely diffusing molecules: Observation of Forster distance dependence and subpopulations. *Proceedings of The National Academy of Sciences of The United States of America*, 96:3670–3675, 1999.

[91] A. A. Deniz, T. A. Laurence, G. S. Beligere, M. Dahan, A. B. Martin, D. S. Chemla, P. E. Dawson, P. G. Schultz, and S. Weiss. Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proceedings of The National Academy of Sciences of The United States of America*, 97:5179–5184, 2000.

[92] A. A. Deniz, T. A. Laurence, M. Dahan, D. S. Chemla, P. G. Schultz, and S. Weiss. Ratiometric single-molecule studies of freely diffusing biomolecules. *Annual Review of Physical Chemistry*, 52:233–253, 2001

[93] N. K. Lee, A. N. Kapanidis, Y. Wang, X. Michalet, J. Mukhopadhyay, R. H. Ebright, and S. Weiss. Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophysical Journal*, 88:2939–2953, 2005.

[94] G. Nienhaus. Single-molecule fluorescence studies of protein folding. *Methods in Molecular Biology*, pages 311–337, 2009.

[95] X. W. Zhuang and M. Rief. Single-molecule folding. *Current Opinion in Structural Biology*, 13:88–97, 2003.

[96] E. A. Lipman, B. Schuler, O. Bakajin, and W. A. Eaton. Single-molecule measurement of protein folding kinetics. *Science*, 301:1233–1235, 2003.

[97] B. Schuler. Single-molecule fluorescence spectroscopy of protein folding. *Chemphyschem*, 6:1206–1220, 2005.

[98] E. V. Kuzmenkina, C. D. Heyes, and G. U. Nienhaus. Single-molecule Forster resonance energy transfer study of protein dynamics under denaturing conditions. *Proceedings of The National Academy of Sciences of The United States of America*, 102:15471–15476, 2005.

[99] K. A. Merchant, R. B. Best, J. M. Louis, I. V. Gopich, and W. A. Eaton. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proceedings of The National Academy of Sciences of The United States of America*, 104:1528–1533, 2007.

[100] A. Borgia, P. M. Williams, and J. Clarke. Single-molecule studies of protein folding. *Annual Review of Biochemistry*, 77:101–125, 2008.

[101] B. Schuler and W. A. Eaton. Protein folding studied by single-molecule FRET. *Current Opinion in Structural Biology*, 18:16–26, 2008.

[102] H. Noji, R. Yasuda, M. Yoshida, and K. Kinosita. Direct observation of the rotation of F-1-ATPase. *Nature*, 386:299–302, 1997.

[103] N. G. Walter. Structural dynamics of catalytic RNA highlighted by fluorescence resonance energy transfer. *Methods*, 25:19–30, 2001.

[104] D. Klostermeier and D. P. Millar. RNA conformation and folding studied with fluorescence resonance energy transfer. *Methods*, 23:240–254, 2001.

[105] T. Heyduk. Measuring protein conformational changes by FRET/LRET. *Current Opinion in Biotechnology*, 13:292–296, 2002.

[106] M. C. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, and T. J. Ha. Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophysical Journal*, 86:2530–2537, 2004.

[107] H. T. Li, X. J. Ren, L. M. Ying, S. Balasubramanian, and D. Klenerman. Measuring single-molecule nucleic acid dynamics in solution by two-color filtered ratiometric fluorescence correlation spectroscopy. *Proceedings of The National Academy of Sciences of The United States of America*, 101:14425–14430, 2004.

[108] J. Yin, A. J. Lin, P. D. Buckett, M. Wessling-Resnick, D. E. Golan, and C. T. Walsh. Single-cell fret imaging of transferrin receptor trafficking dynamics by sfp-catalyzed, site-specific protein labeling. *Chemistry & Biology*, 12:999–1006, 2005.

[109] J. Y. Lee, B. Okumus, D. S. Kim, and T. J. Ha. Extreme conformational diversity in human telomeric DNA. *Proceedings of The National Academy of Sciences of The United States of America*, 102:18938–18943, 2005.

[110] H. Wang, Y. S. Yeh, and P. F. Barbara. HIV-1 nucleocapsid protein bends double-stranded nucleic acids. *Journal of The American Chemical Society*, 131:15534–15543, 2009.

[111] R. Zhao and D. Rueda. RNA folding dynamics by single-molecule fluorescence resonance energy transfer. *Methods*, 49:112–117, 2009.

[112] G. Shan and W. Huang. Extracting the single-molecule fluorescence trajectories of folding protein in single-pair fluorescence resonance energy transfer experiment. *Journal of Nanoscience and Nanotechnology*, 9:1176–1180, 2009.

[113] M. R. Whorton, M. P. Bokoch, S. G. F. Rasmussen, B. Huang, R. N. Zare, B. Kobilka, and R. K. Sunahara. A monomeric g protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its g protein. *Proceedings of The National Academy of Sciences of The United States of America*, 104:7682–7687, 2007.

[114] C. Eggeling, P. Kask, D. Winkler, and S. Jager. Rapid analysis of forster resonance energy transfer by two-color global fluorescence correlation spectroscopy: Trypsin proteinase reaction. *Biophysical Journal*, 89:605–618, 2005.

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

[116] K. C. Neuman and A. Nagy. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nature Methods*, 5:491–505, 2008.

[117] J. Zlatanova, S. M. Lindsay, and S. H. Leuba. Single molecule force spectroscopy in biology using the atomic force microscope. *Progress in Biophysics & Molecular Biology*, 74:37–61, 2000.

[118] D. E. Smith, G. J. Gemmen, R. Millin, J. P. Rickgauer, A. L. Schweitzer, and D. N. Fuller. Using optical tweezers to study protein-DNA interactions. *Proceedings of the SPIE - The International Society for Optical Engineering*, 5930:593012–1–10, 2005.

[119] A. Ashkin, J. M. Dziedzic, and T. Yamane. Optical trapping and manipulation of single cells using infrared-laser beams. *Nature*, 330:769–771, 1987.

[120] K. Svoboda and S. M. Block. Biological applications of optical forces. *Annual Review of Biophysics and Biomolecular Structure*, 23:247–285, 1994.

[121] D. G. Grier. A revolution in optical manipulation. *Nature*, 424:810–816, 2003.

[122] C. Haber and D. Wirtz. Magnetic tweezers for DNA micromanipulation. *Review of Scientific Instruments*, 71:4561–4570, 2000.

[123] C. Gosse and V. Croquette. Magnetic tweezers: Micromanipulation and force measurement at the molecular level. *Biophysical Journal*, 82:3314–3329, 2002.

[124] C. Q. Yi, C. W. Li, S. L. Ji, and M. S. Yang. Microfluidics technology for manipulation and analysis of biological cells. *Analytica Chimica Acta*, 560:1–23, 2006.

[125] M. Sotomayor and K. Schulten. Single-molecule experiments in vitro and in silico. *Science*, 316:1144–1148, 2007.

[126] R. B. Best, D. J. Brockwell, J. L. Toca-Herrera, A. W. Blake, D. A. Smith, S. E. Radford, and J. Clarke. Force mode atomic force microscopy as a tool for protein folding studies. *Analytica Chimica Acta*, 479:87–105, 2003.

[127] M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez. Mechanical and chemical unfolding of a single protein: A comparison. *Proceedings of The National Academy of Sciences of The United States of America*, 96:3694–3699, 1999.

[128] P. E. Marszalek, H. Lu, H. B. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten, and J. M. Fernandez. Mechanical unfolding intermediates in titin modules. *Nature*, 402:100–103, 1999.

[129] M. S. Z. Kellermayer, S. B. Smith, C. Bustamante, and H. L. Granzier. Complete unfolding of the titin molecule under external force. *Journal of Structural Biology*, 122:197–205, 1998.

[130] W. A. Linke and A. Grutzner. Pulling single molecules of titin by AFM - recent advances and physiological implications. *Pflugers Archiv-European Journal of Physiology*, 456:101–115, 2008.

[131] M. Cieplak and P. E. Marszalek. Mechanical unfolding of ubiquitin molecules. *Journal of Chemical Physics*, 123:194903–194910, 2005.

[132] C. L. Chyan, F. C. Lin, H. B. Peng, J. M. Yuan, C. H. Chang, S. H. Lin, and G. L. Yangy. Reversible mechanical unfolding of single ubiquitin molecules. *Biophysical Journal*, 87:3995–4006, 2004.

[133] M. Carrion-Vazquez, H. B. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez. The mechanical stability of ubiquitin is linkage dependent. *Nature Structural Biology*, 10:738–743, 2003.

[134] Y. Cao, C. Lam, M. J. Wang, and H. B. Li. Nonmechanical protein can have significant mechanical stability. *Angewandte Chemie-International Edition*, 45:642–645, 2006.

[135] D. J. Brockwell, G. S. Beddard, E. Paci, D. K. West, P. D. Olmsted, D. A. Smith, and S. E. Radford. Mechanically unfolding the small, topologically simple protein l. *Biophysical Journal*, 89:506–519, 2005.

[136] D. J. Brockwell, E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham, and S. E. Radford. Pulling geometry defines the mechanical resistance of a beta-sheet protein. *Nature Structural Biology*, 10:731–737, 2003.

[137] R. B. Best, B. Li, A. Steward, V. Daggett, and J. Clarke. Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation. *Biophysical Journal*, 81:2344–2356, 2001.

[138] C. G. Baumann, V. A. Bloomfield, S. B. Smith, C. Bustamante, M. D. Wang, and S. M. Block. Stretching of single collapsed DNA molecules. *Biophysical Journal*, 78:1965–1978, 2000.

[139] H. Gumpp, E. M. Puchner, J. L. Zimmermann, U. Gerland, H. E. Gaub, and K. Blank. Triggering enzymatic activity with force. *Nano Letters*, 9:3290–3295, 2009.

[140] M. T. Woodside, C. Garcia-Garcia, and S. M. Block. Folding and unfolding single RNA molecules under tension. *Current Opinion in Chemical Biology*, 12:640–646, 2008.

[141] P. Hinterdorfer, W. Baumgartner, H. J. Gruber, K. Schilcher, and H. Schindler. Detection and localization of individual antibody-antigen recognition events by atomic force microscopy. *Proceedings of The National Academy of Sciences of The United States of America*, 93:3477–3481, 1996.

[142] H. Grubmuller, B. Heymann, and P. Tavan. Ligand binding: Molecular mechanics calculation of the streptavidin biotin rupture force. *Science*, 271:997–999, 1996.

[143] A. J. Maniotis, C. S. Chen, and D. E. Ingber. Demonstration of mechanical connections between integrins cytoskeletal filaments, and nucleoplasm that stabilize nuclear structure. *Proceedings of The National Academy of Sciences of The United States of America*, 94:849–854, 1997.

[144] P. P. Lehenkari and M. A. Horton. Single integrin molecule adhesion forces in intact cells measured by atomic force microscopy. *Biochemical and Biophysical Research Communications*, 259:645–650, 1999.

[145] B. D. Brower-Toland, C. L. Smith, R. C. Yeh, J. T. Lis, C. L. Peterson, and M. D. Wang. Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA. *Proceedings of The National Academy of Sciences of The United States of America*, 99:1960–1965, 2002.

[146] K. Svoboda, C. F. Schmidt, B. J. Schnapp, and S. M. Block. Direct observation of kinesin stepping by optical trapping interferometry. *Nature*, 365:721–727, 1993.

[147] T. Mori, R. D. Vale, and M. Tomishige. How kinesin waits between steps. *Nature*, 450:750–U15, 2007.

[148] G. Diezemann and A. Janshoff. Force-clamp spectroscopy of reversible bond breakage. *Journal of Chemical Physics*, 130:041101–041104, 2009.

[149] 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. *Journal of Computational Chemistry*, 4:187–217, 1983.

[150] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. Charmm: The biomolecular simulation program. *Journal of Computational Chemistry*, 30:1545–1614, 2009.

[151] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald - an n.log(n) method for Ewald sums in large systems. *Journal of Chemical Physics*, 98:10089–10092, 1993.

[152] 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.

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

[154] B. Roux and T. Simonson. Implicit solvent models. *Biophysical Chemistry*, 78:1–20, 1999.

[155] 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.

[156] 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.

[157] 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.

[158] 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.

[159] M. Schlenkrich, J. Brickmann, A. D. Mackerell., and M. Karplus. An empirical potential energy function for phospholipids: Criteria for parameter optimization and applications. *Biological membranes: A molecular perspective from computation and experiment*, pages 31–81, 1996.

[160] G. Kamath, O. Guvench, and A. D. MacKerell. CHARMM additive all-atom force field for acyclic carbohydrates and inositol. *Journal of Chemical Theory and Computation*, 4:765–778, 2008.

[161] E. R. Hatcher, O. Guvench, and A. D. MacKerell. CHARMM additive all-atom force field for acyclic polyalcohols, acyclic carbohydrates, and inositol. *Journal of Chemical Theory and Computation*, 5:1315–1327, 2009.

[162] S. Patel, A. D. Mackerell, and C. L. Brooks. CHARMM fluctuating charge force field for proteins: II - protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *Journal of Computational Chemistry*, 25:1504–1514, 2004.

[163] 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 .1. the method. *Journal of Computational Chemistry*, 13:1011–1021, 1992.

[164] J. G. Laski. On time structure in (Monte-Carlo) simulations. *Operational Research Quarterly*, 16:329–339, 1965.

[165] T. S. Logsdon and R. C. Africano. A modified Monte Carlo procedure. *AIAA Journal*, 6:1111–1117, 1968.

[166] M. Wehrli. On sampling time reduction in Monte Carlo simulations. *Unternehmensforschung*, vol.14:97–108, 1970.

[167] D. D. L. Minh. Multidimensional potentials of mean force from biased experiments along a single coordinate. *Journal of Physical Chemistry B*, 111:4137–4140, 2007.

CHAPTER II

Using Experimental Potentials of Mean Force to Guide Umbrella Sampling Simulations

2.1 Introduction

The potential of mean force (PMF) [1], i.e., the free energy $G(\xi) = -k_B T \ln\langle p(\xi)\rangle$ with respect to a chosen reaction coordinate ξ, is fundamentally related to that coordinate's distribution function, averaged over all the other degrees of freedom $\langle p(\xi)\rangle$. As such, it is a central concept in the statistical mechanical representation of molecular systems, and has been employed in a number of computational applications [2, 3]. Calculating accurate PMFs for large systems on biologically relevant timescales is computationally expensive. However, advances in computing power and the development of techniques to improve sampling have made it possible to calculate PMFs for increasingly complex systems.

With the increase in the accuracy of calculating equilibrium distributions by computer simulations, in the experimental realm, advances in molecular optical spectroscopy [4] have also made it possible to determine more accurate distribution functions of measurable coordinates. Such measurements can be done either in the bulk or at the single molecule-level, thereby resolving possible differences between the behavior of a number of individual molecules versus measuring ensemble averages, which in turn can reflect, for skewed probability distribution, differences between observing

the most probable vs. the average observables [5]. These experimental distribution functions can be used to obtain free energy landscapes (PMFs) for conformational dynamics of proteins and other biomolecules [6, 7]. For example, a study by Weiss and coworkers [6] using single molecule fluorescence resonance energy transfer (FRET) between sites that are fluorescently labeled in a biomolecule has enabled the characterization of protein folding distributions by measuring the probability density of the distance between the N-terminus and the loop of chymotrypsin inhibitor 2 at various concentrations of denaturant. FRET can also probe conformational changes for RNA molecules, which have been shown to require large equilibrium dynamical transitions as a prerequisite for their function [8]. Single-molecule pulling techniques have also been employed to perform unfolding manipulations of biomolecules by using atomic force microscopy [9], optical [10] or magnetic trapping [11], thereby generating extension trajectories that, when binned and processed appropriately [12, 13], can produce a mapping of the potential of mean force and possibly of its entropic and enthalpic components [14] along the pulling coordinate. PMFs from these experiments can be calculated both in the absence of the pulling force and for various values of the force [15]. Such experiments have been used to calculate potentials of mean force for the unfolding of the I27 domain of titin [16] and for the membrane protein bacteriorhodopsin [17]. Constant-force reversible folding-unfolding measurements can also be used to map the folding energy landscape, as was showcased in recent work on DNA hairpins [18].

However, experiments such as the ones described above or others (including bulk experiments) which can resolve general conformational changes along a small set of degrees of freedom (for regular FRET one, or –with the emergence of three color FRET [19, 20]– up to three spatial separations) usually give only on-off signals. Sim-

ilarly in single-molecule pulling, the extension that is measurable is poorly resolved at the point of transition (e.g., when a protein unfolds). Molecular dynamics simulations [21], on the other hand, are useful for computing dynamic and equilibrium processes of large molecules in atomic detail for any degree of freedom and with high, femtosecond time resolution, but processes which involve large free energy barriers or long timescales are too computationally demanding to be simulated directly.

For this reason, it is often difficult to compute the averages required for the calculation of a free energy profile from a direct molecular dynamics simulations, i.e., by directly counting the frequency of occurrence of the values of the reaction coordinate during the simulation run. The most used method for generating a PMF efficiently from molecular dynamics or Monte Carlo simulations is the venerable umbrella sampling method [22]. In this method, a reaction coordinate is chosen and restraining potentials acting on it are used over a series of windows to sample the range of the reaction coordinate. By applying the restraining potentials, the system is encouraged to sample regions of conformational space that would not otherwise be accessible during the direct sampling. The result is a series of histograms which contain the biased distribution of the reaction coordinate from each window. These histograms are then unbiased and combined [23, 24], usually using the weighted histogram analysis method (WHAM) [25]. Typically, the restraining potentials are chosen through trial and error, which results in a great deal of preliminary data being discarded as the potentials are optimized. We propose here a simple method which would use umbrella sampling potentials based on an *experimentally* determined (but not necessarily accurate) potential of mean force from measured equilibrium distributions to improve molecular dynamics simulation efficiency.

For a uniform sampling of the reaction coordinate, the ideal "restraining" po-

tential would be the *negative* of the exact potential of mean force (that is, if the PMF would be known a priori). Using the negative of the PMF flattens the energy landscape, circumventing the problem of trapping by large energy barriers [25, 3]. Most techniques for improving the efficiency of umbrella sampling, such as adaptive umbrella sampling methods [26, 27, 28, 29, 30, 31, 32] and other techniques that generate flat distributions [33, 34, 35, 36], focus on this approach. In adaptive umbrella sampling, a continuously updated umbrella potential which is a function of one or a few important degrees of freedom is added iteratively to the unperturbed Hamiltonian of the system. The umbrella potential for each iteration of the simulation is chosen to be an estimate of the negative of the PMF determined from the previous iteration. As a result, one obtains a uniform sampling distribution along the important degree(s) of freedom. In the multicanonical sampling method [33, 34] or its twin sister, the entropy sampling method [37], and in Wang-Landau sampling [36] and related methods [38, 39, 40], a uniform *energy* probability distribution is obtained by assigning weights that are inversely proportional to a pre-calculated (or a dynamically updated) density of states.

Flattening the conformation or energy distributions to promote uniform sampling comes however at the substantial computational cost of pre-determining either the adaptive umbrella potentials or the density of states, respectively. Moreover, as most biological molecules do not uniformly sample all conformation space, there would be no need to flatten energy barriers if the location of the barriers would be known beforehand. Instead, we propose here to add to the unperturbed Hamiltonian a set of experimentally-derived harmonic potentials of different curvatures, fitted to the *positive* of an experimental PMF. This will serve as a "guide" for the reaction coordinate along the free energy path. It has been suggested previously that adding

harmonic umbrella potentials minus the PMF would improve convergence for multi-window umbrella sampling techniques (as opposed to adaptive umbrella sampling, in which the negative of the PMF alone is used in a single window that is constantly updated) [41, 3]. Subtracting the PMF from individual harmonic potentials gives similar potentials to our guiding technique, as long as the initial harmonic potentials are well chosen, as is depicted in the cartoon in Fig. 2.1 (see also discussion in Sect. 2.3). In other words, our experimentally-guided umbrella sampling should produce similar potentials to the best case of this convergence method. Using a known experimental PMF should allow one to choose optimal initial restraining potentials, eliminating guesswork and wasted data.

In both the straightforward harmonic-wells-minus-PMF method and the adaptive umbrella sampling approaches, a good guess of the approximate negative of the PMF is crucial to their rapid convergence. Here we suggest how this initial guess can be provided not by preliminary simulations but directly from experimental input. This approach is similar in spirit to recent approaches that use experimental input in the form of, say, order parameter nuclear magnetic resonance (NMR) data [42] or hydrogen exchange [43] to improve the description of conformational equilibrium of proteins.

To showcase our method, we devised an initial feasibility study using an unstructured pentapeptide with sequence Cys-Ala-Gly-Gln-Trp. In order to model loop formation dynamics, we chose the end-to-end distance as the reaction coordinate. This peptide was chosen as a test system because it has a size small enough to allow for exhaustive sampling by direct methods, and because both experimental [44] and computational [45] work have used this pentapeptide as a model for contact formation in protein folding. Because no experimental potential of mean force for

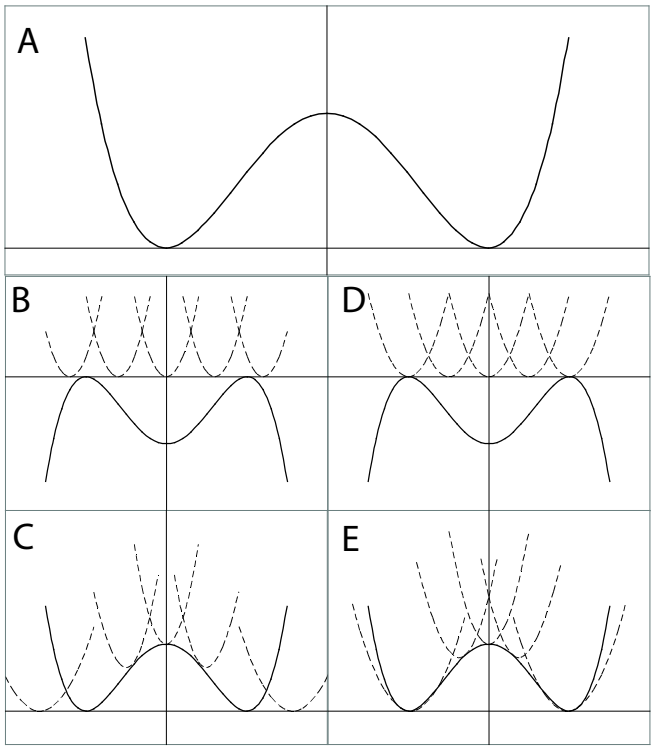

Figure 2.1: Schematic examples of restraining potentials for a double well energy potential using the PMF subtracted from harmonic potentials. (A) The original double well potential. (B) A set of uneducated-guess (poor) starting potentials (dashed lines) with the negative of the PMF (solid line) and (C) the results of subtracting the PMF from these harmonic potentials (dashed lines) overlaid with the original PMF. (D) A nearly-optimal set of starting potentials and (E) the results of subtracting the PMF from these optimal harmonic potentials, overlaid with the original PMF. The potentials in E resulting from the subtraction (dotted lines) produce biasing restraints similar to the type we suggest to use in our guided umbrella sampling protocol.

our chosen reaction coordinate was available, we have instead generated the PMF from extensive sampling simulations. This PMF was subsequently used as the "experimental" guiding PMF. Guided umbrella sampling was run using five windows with restraining potentials of varying curvature based on the guiding PMF. Regular, unguided umbrella sampling experiments were also run in five and seventeen windows for comparison with guided umbrella sampling simulations. The method was then tested using a more complex system for which an experimental PMF has been determined; the forced unfolding of the I27 domain of the muscle protein titin [16] The systems are described in more detail in the following sections, followed by

a discussion of the theory behind umbrella sampling and the experimentally guided umbrella sampling method.

Pentapetide Dynamics

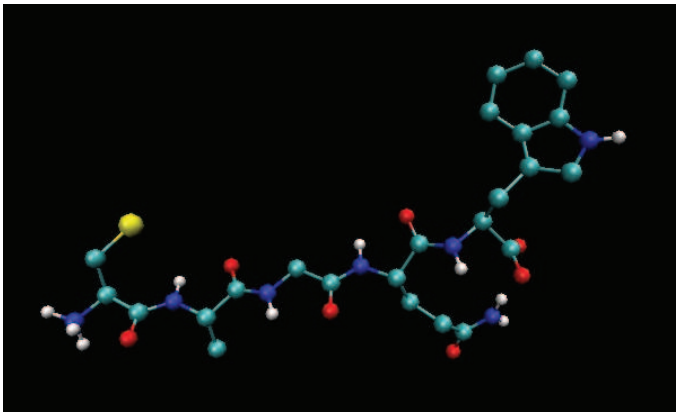

Figure 2.2: Structure of pentapeptide CAGQW

The pentapeptide with sequence CAGQW (Figure 2.2) has proven a useful system for the study of peptide interactions due to its relative flexibility and the unique properties of its terminal amino acids. Tryptophan phosphoresces with a triplet state lifetime of $40\mu s$ [46, 47, 48]. Cysteine, on the other hand, is an efficient enough quencher that the rates for contact formation between a cysteine and tryptophan residue are expected to be on the same order as the measured decay rate of the tryptophan triplet state[49]. Interactions between the ends of the pentapeptide can therefore be measured directly by looking at the quenching of tryptophan's triplet state by cysteine [44] making it possible to experimentally determine the kinetics of contact formation. Experiments by Lapidus et al. [44] have found the contact rate for this peptide to be approximately $1.3 \times 10^8 s^{-1}$. Molecular dynamics studies in both CHARMM and AMBER have calculated similar kinetic rates, confirming that the dynamics of this system are fast enough to be sampled adequately in MD [45, 50, 51].

Forced Unfolding of Titin I27 Domain

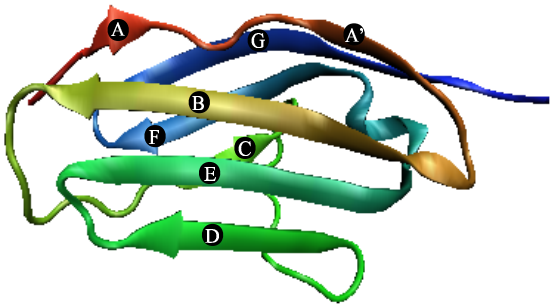

Figure 2.3: Structure of titin I27 domain.

The I27 domain of the elastomeric muscle protein titin is one of the most well characterized proteins in terms of mechanical properties. As such, it represents the perfect system on which to test our method. Titin's mechanically active region is made up of tandem repeats of immunoglobulin-like (IG) domains, which have a characteristic β fold structure [52, 53]. This I-band region acts as a molecular spring [54, 55]. Numerous studies on titin immunoglobulin domains in general, and domain I27 in particular, have been carried out using AFM force pulling [56, 100, 58, 59, 57], optical tweezers [60, 61, 62], theoretical calculations [63, 64, 65, 66], and simulations [67, 68, 69, 70, 71, 72, 73, 74, 75, **?**, 76]. These studies, taken together, represent a comprehensive picture of the forced unfolding pathway of titin I27, which can be used to validate the results of our simulations. Most importantly for our purposes,

the potential of mean force for the forced unfolding of I27 has been calculated from AFM pulling experiments [16] using the Jarzynski equality [77]. This calculation gives us an experimental PMF on which to base our biasing potentials.

2.2 Background on Umbrella Sampling

In umbrella sampling [22], a biasing potential is added to the Hamiltonian to direct the simulations toward a certain goal. The biasing potential is usually in the form of a harmonic potential restraint that keeps the value of a relevant reaction coordinate fluctuating around successive positions along that coordinate. The result of this "stratification" is a series of biased histograms. In order to obtain accurate information about the free energy of the system, the raw data from individual simulations must be unbiased and recombined. The addition of the harmonic restraining potential, w_i, is equivalent to a multiplicative weight in the Boltzmann factor of $e^{-w_i(\xi)/k_BT}$. Thus, the biased distribution function for each window is

$$\langle p(\xi)\rangle_i^{\text{biased}} = e^{-w_i(\xi)/k_BT}\langle p(\xi)\rangle\langle e^{-w_i(\xi)/k_BT}\rangle^{-1} \tag{2.1}$$

and the unbiased PMF in the ith window is

$$G_i(\xi) = -k_BT\ln\langle p(\xi)\rangle - w_i + f_i \tag{2.2}$$

where the unknown free energy constant f_i is defined by the equation

$$e^{f_i/k_BT} = \langle e^{w_i/k_BT}\rangle \tag{2.3}$$

In the weighted histogram analysis method (WHAM) [25], one uses an iterative process to determine the free energy constants, f_i. An initial guess set of f_i is used to estimate the unbiased probability distribution, where

$$\langle p(\xi)\rangle = \sum_i n_i\langle p(\xi)\rangle_i \times [\sum_j n_j e^{-(w_j(\xi)-f_j)/k_bT}]^{-1} \tag{2.4}$$

The resulting probability is then used in the equation

$$e^{-f_i/k_BT} = \int d\xi e^{-w_i(\xi)/k_BT}\langle p(\xi)\rangle \tag{2.5}$$

to determine a new set of f_i values. The last two equations are solved iteratively until they are self consistent.

These equations can be further extended to calculate the PMF along a secondary reaction coordinates. Assuming that the secondary coordinate is not independent of the biased coordinate, then the biasing function applied to the primary reaction coordinate will also affect the probability distribution of the secondary coordinate. It is therefore necessary to calculate the probability distribution of this coordinate as a conditional probability dependent on the primary coordinate and unbias it in the same was as the one-dimensional case. The PMF of the secondary reaction coordinates is unbiased using the equation,

$$\langle p(\phi|\xi)\rangle = \sum_{\xi}\sum_{i} n_i\langle p(\phi|\xi)\rangle_i \times [\sum_{j} n_j e^{-(w_j(\xi)-f_j)/k_bT}]^{-1} \tag{2.6}$$

where ξ is the restrained reaction coordinate and ϕ is the unrestrained coordinate. The resulting two-dimensional probability function is then collapsed along the biased reaction coordinate to yield a probability distribution and PMF of the secondary coordinate. Iterative calculations of $\langle p(\phi|\xi)\ \rangle and f_i$ are then done done using Eqs. (2.4) and (2.5) as in the 1D case [28, 78, 79].

For regular implementations of WHAM using multiple-window umbrella sampling techniques, the strength of the harmonic biasing potentials and the number of simulation windows must be chosen to allow sufficient overlap for the data to be recombined while at the same time encouraging more rapid sampling of the reaction coordinate. Simply increasing the number of windows will not necessarily improve convergence [41, 3]. Furthermore, in order to obtain accurate information about other degrees of

freedom, the chosen reaction coordinate must be the slowest one [41]. Determining optimal factors by trial and error leads to a great deal of wasted CPU time. Adaptive umbrella sampling overcomes this to a certain extent by updating the estimate of the PMF and using its negative to flatten barriers, but for regions which have not been sampled in previous simulations, the shape of the restraining potential must be set arbitrarily or extrapolated [26, 27, 28, 29, 30]. Another issue is the error in the PMF calculated by the WHAM equations, which is sensitive to the size of the bins used in constructing the histograms. Insufficient sampling leads to a large number of bins with too few configurations per bin, resulting in statistical error, while using too few bins can lead to errors in the probability density [80].

2.3 The Experimentally-Guided Umbrella Sampling Method

Knowing the PMF from experiment makes the choice of optimal restraining potentials possible, but it is important that the simulations are carried out in a way that yields information about the system of interest as a whole, rather than just the chosen reaction coordinate. This means that, in order for the experimentally-guided biasing method to be practical, it is important to use experimental guiding potentials that are reported along the slow degrees of freedom of the system, so that the other degrees of freedom are sufficiently sampled so that they converge in the simulation.

As presented in the Introduction section, we have used both a simulated and an experimental PMF as input in the calculation by fitting with harmonic restraining potentials of varying curvature. Assume that an experimental guiding PMF, G_{exp} is available, say from the probability distribution $p_{\text{exp}}(\xi)$ of a distance ξ measured from a FRET or micromanipulation experiment, i.e., up to a constant,

$$G_{\text{exp}}(\xi) = -k_B T \ln\langle p_{\text{exp}}(\xi)\rangle. \tag{2.7}$$

This guiding potential could be used in one of the following two ways (see panels B-C and D-E, respectively, in Fig. 2.1). The first possibility would be to choose a number of restraining potentials w_i centered each around a corresponding ξ_i value, and to run the corresponding ith simulation using the effective potential energy

$$V(\mathbf{r}) + w_i(\xi(\mathbf{r})) - G_{\text{exp}}(\xi(\mathbf{r})), \tag{2.8}$$

where V is the inter-atomic potential energy function employed in the simulation and $\mathbf{r}$ the $3N$-dimensional conformational vector. The subtraction of G_{exp} contributes to a flattening of the energy landscape along ξ, and the restraints focus the simulation in a particular domain of ξ. However, the distribution of the w_i's can be optimized (see Fig. 1D,E) such that the guiding $w_i^{(g)}$ restraining potentials replace $w_i(\xi(\mathbf{r})) - G_{\text{exp}}(\xi(\mathbf{r}))$, i.e., the combination of negative PMF plus equal force-constant restraints. In that case, this is equivalent to running simulations on an effective potential

$$V(\mathbf{r}) + w_i^{(g)}(\xi(\mathbf{r})), \tag{2.9}$$

where this time the guiding potentials of the ith window,

$$w_i^{(g)}(\xi) = \frac{1}{2}k_i^{(g)}(\xi - \xi_i^{(g)})^2 \tag{2.10}$$

have optimal $\xi_i^{(g)}$ *and* $k_i^{(g)}$ (force-constant) values found from piece-wise fitting of the PMF. This second way to perform umbrella sampling using experimental guiding data is the essence of the method we propose.

For the pentapetide test case we showcase herein, the guiding potentials were chosen such that the center of the harmonic well was located at either a minimum of a well or at the central position of a shoulder. In other words, the piece-wise fit of the above guiding potentials to the experimental PMF was done based on the plot of the second derivative $d^2G_{\text{exp}}/d\xi^2$, i.e., based on the long-scale change of the

curvature of the "experimentally" available PMF (see Fig. 2.4 for details). The value of each force constant, $k_i^{(g)}$, was chosen such that the harmonic well approximated the curvature of the PMF around the minimum. The choice of a harmonic potential that approximates the shape of the well is an obvious one, since there would be no point in using a potential wider than the well, and using a narrower potential would require more windows in order to sample over the excluded space. In situations where this does not provide sufficient overlap, another potential may be added with its center between those of the surrounding potentials and its $k_i^{(g)}$ value also chosen to approximate the curvature of the PMF. For titin I27 potentials were similarly fit to the curvature of the experimental PMF and spaced to provide sufficient overlap.

The PMFs for the test systems used had no barriers; in the case when barriers in the experimental PMF would exist (see Fig. 2.1), a potential with the center placed at the peak of a barrier would have the $k_i^{(g)}$ value chosen to approximate the (negative) local curvature at the transition point. In the case of the pentapeptide studied here, (see Fig. 2.4) the choice of five w_i windows was obvious (cf. Figure 2.4, there are five regions in the curvature plots). In general applications, an iterative optimization procedure that minimizes a piecewise cost function describing the root-mean-square difference between the experimental PMF and the fitting windows can be done over the relevant set of parameters, i.e., number of windows, positions of the restraints ξ_i^g, and force constants k_i^g.

Regardless of their particular choice, the addition of the guiding potentials $w_i^{(g)}$ in Eq. (2.10) to the system's unperturbed Hamiltonian can be unbiased using the regular weighted histogram analysis method as presented in Section 2.2.

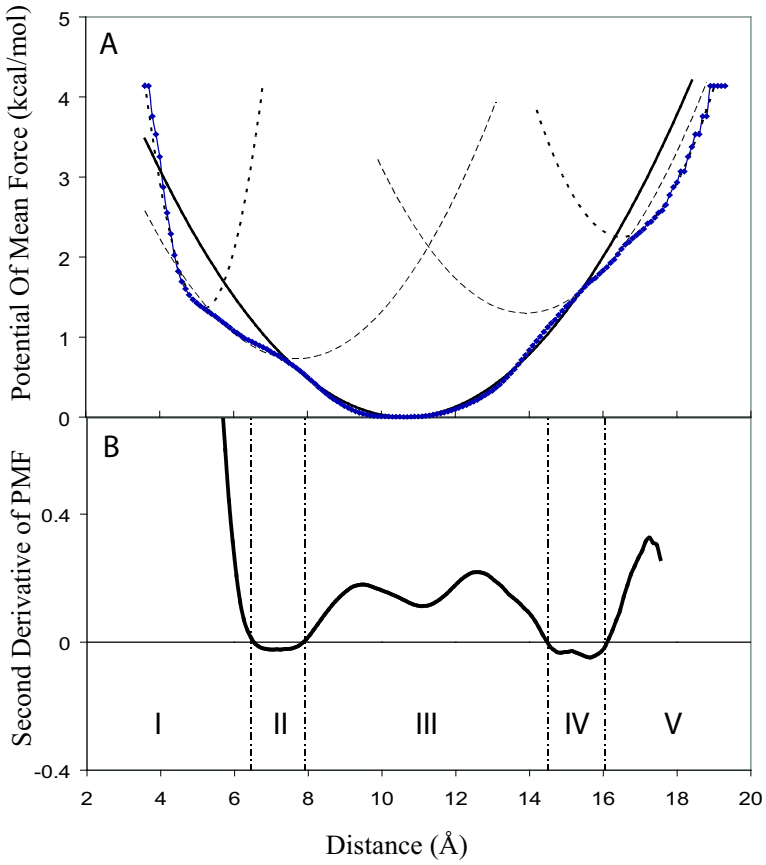

Figure 2.4: (A) "Experimental" potential of mean force for the end-to-end distance of pentapeptide with fitted biasing potentials. (B) Second derivative of the PMF calculated using a finite-difference scheme and by eliminating, with a running average filter of width 2.5 Å, changes in the curvature occurring on length scales smaller than the thermally accessible range for the harmonic potentials. The plot has five flat regions; correspondingly, five guiding umbrella potentials are chosen with the locations of their minima within the five regions placed according to a root-mean-square best fit to the "experimental" PMF.

2.4 Simulation Methods

Molecular dynamics simulations were performed on a five amino acid peptide with sequence Cys-Ala-Gly-Gln-Trp with the polar-hydrogen parameter set 19 of the CHARMM force field [82], using the ACE II implicit solvent potential [81], which is based on a generalized Born solvation model.

The starting structure of the peptide was built using typical internal coordinate values. All simulations were minimized with 1000 steps of steepest descent followed

by 2000 steps of the adopted-basis Newton-Raphson procedure and heated up to 300 K in 100 ps. All simulations used Langevin dynamics at 300 K with a (water) frictional coefficient set to 91 ps^{-1}. The reaction coordinate, ξ, was defined as the distance in Angstroms between the sulfur of cysteine and the center of mass of the tryptophan aromatic ring. The various sets of simulations we ran were as follows:

(a) exhaustive sampling simulations totaling 0.5 μs simulation time were run using constant-temperature molecular dynamics (in the absence of any restraining potential). We generated 25 initial conformations from a 10 ns long equilibration out of which originated 25 equilibrium trajectories, each 20 nanoseconds long, with different canonically distributed initial conditions; the conformational snapshots such obtained during the 0.5 μs cumulative time were thus representative conformations drawn from the canonical ensemble and were binned by the value of the reaction coordinate ξ to generate the "experimental" potential of mean force. In other words, this is, in the absence of a real experimental PMF, what we will use as the guiding PMF to which the guiding potentials will be fit.

(b) unguided (regular) equal-restraint umbrella sampling simulations run in

   i. seventeen windows; $w_i = k_i(\xi - \xi_i)^2,\ i = \overline{1,17}$, all with the same restraining potential constant, $k_i = 1.2$ kcal/(mol$Å^2$) were run for 15 ns each for a total of 255 ns. The windows were placed at positions ξ_i that ranged from 3 to 19 Å in 1 Å increments. The value of the reaction coordinate was recorded every 10 fs.

   ii. five windows; $w_i = k_i(\xi - \xi_i)^2,\ i = \overline{1,5}$; this second set of umbrella sampling simulations was run using the same range but fewer windows and wider

restraining potentials with force constant $k_i = 0.4$ kcal/(molÅ^2) placed at 3, 7, 11, 15, and 19 Å for 15 ns each for a total of 75 ns.

(c) experimentally-guided umbrella sampling with five guiding potentials $w_i^{(g)} = k_i^{(g)}(\xi - \xi_i^{(g)})^2$ fitted to the simulated "experimental" PMF obtained from the snapshots recorded during the 0.5 μs exhaustive sampling trajectory (Figure 2.4). The guided umbrella sampling potential force constants were: $k_1^{(g)} =$ 1.1 kcal/(molÅ^2), $k_2^{(g)} = 0.11$ kcal/(molÅ^2), $k_3^{(g)} = 0.07$ kcal/(molÅ^2), $k_4^{(g)} =$ 0.12 kcal/(molÅ^2), $k_5^{(g)} = 0.3$ kcal/(molÅ^2)y, placed at $\xi_1^{(g)} = 5.2$ Å, $\xi_2^{(g)} = 7.7$ Å, $\xi_3^{(g)} = 10.6$ Å, $\xi_4^{(g)} = 13.9$ Å and $\xi_4^{(g)} = 16.5$ Å. Each window was run for 15 ns for a total of 75 ns.

To further test our method, we calculated the potential of mean force for the forced unfolding of the I27 domain of titin. The free energy surface for forced unfolding of the I27 domain of titin has been experimentally determined from single-molecule pulling experiments[16], using the Jarzynski equation [77]. Both experimentally guided and unguided umbrella sampling simulations were run on the titin I27 domain using the CHARMM 22 parameter set with CMAP correction [83, 84] set in ace II implicit solvent. Umbrella sampling windows were equilibrated for 500ps followed by a production run of 5ns. For umbrella sampling the reaction coordinate used was the difference of the distance between the C_α atoms of the terminal residues from the equilibrium distance i.e. the equilibrium distance, 46 Å is designated r=0, 47 Å is r=1 and so on. For guided umbrella sampling, potentials were assigned every 1 Å or 0.5 Å with a force constant based on fitting to the experimental PMF. Distance between adjacent windows was determined based on overlap between the restraining potentials. For unguided umbrella sampling, potentials were assigned every 0.5 Å with a force constant of 10 kcal/molÅ. The total simulation time was

140 ns for guided umbrella sampling and 155 ns for the unguided umbrella sampling. The starting configuration for each window was taken from the previous window.

All umbrella sampling distributions were unbiased using a memory efficient implementation of the WHAM Eqs. (2.4) and (2.5) by Alan Grossfield [85]. PMFs were also generated for the exhaustive and guided umbrella sampling simulations of the pentapeptide at various times during the simulations in order to compare their evolution in time.

2.5 Results

2.5.1 Pentapeptide Test Case

Comparison of the Probability Distributions and PMFs. An efficient umbrella sampling simulation typically consists of a set of as few restraining windows as possible, with as much overlap between neighboring windows as possible. The histograms for the three umbrella sampling experiments are shown in Figure 2.5. While the overall shape looks similar for all three, the 5 window unguided umbrella sampling does not have a substantial amount of overlap between the histograms. This could be improved by adding more windows or altering the restraining potentials, but either would result in the need for more computational time.

The unbiased probability distribution for the reaction coordinate (i.e., the result of unbiasing the histograms in Fig. 2.5 with the WHAM formulae Eqs.(2.4) and (2.5)) were quantitatively compared to the "experimental" distribution from extensive sampling quantitatively using a Kolmogorov-Smirnov (K-S) test. The K-S statistic [86] is a first order measure of similarity defined as

$$D = \max_{-\infty<\xi<\infty} \mid N_1(\xi) - N_2(\xi) \mid \tag{2.11}$$

where $N_1(\xi)$ and $N_2(\xi)$ are the cumulative probability distributions for the two sets

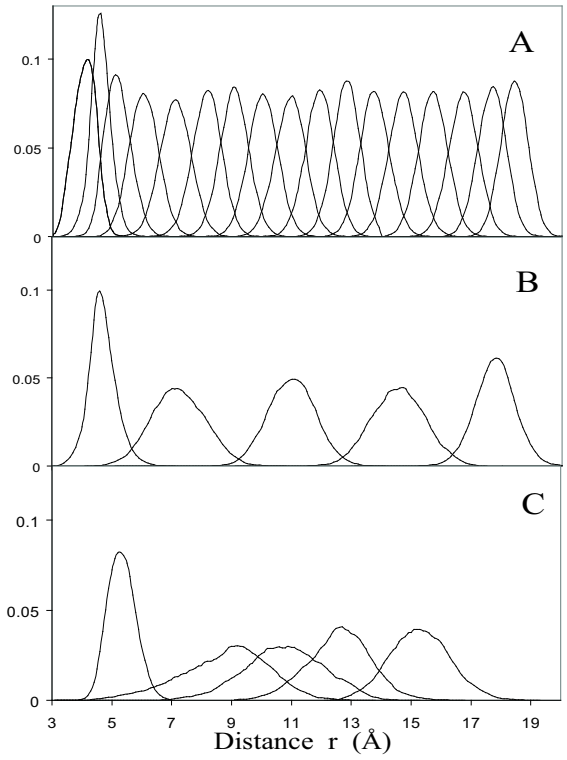

Figure 2.5: Biased histograms for (A) the unguided 17 window umbrella sampling, (B) 5 window unguided umbrella sampling, and (C) 5 window guided umbrella sampling after 10 ns per window for each. Note the poor overlap for the histograms in panel B relative to A, and the recovery of the overlap in C.

to be compared; it has been used previously [87] as a stringent test for accurately gauging conformational sampling efficiency.

The Kolmogorov-Smirnov statistics are plotted as a function of simulation time in Figure 2.6. A small value of the K-S statistic indicates a higher probability that the two data sets that are compared are drawn from the same underlying probability distribution, with a value of 0 indicating identical distributions. The results suggest that our guided umbrella sampling approaches the exhaustive sampling probability distribution much more quickly than the 17 window unguided umbrella sampling while the 5 window unguided umbrella sampling never comes as close to reproducing the results of the extensive sampling as the other methods, even after 75 ns. After about 30 ns the guided umbrella sampling gives a K-S value of 0.025 as compared to

extensive sampling, but after 30 ns the K-S value starts to increase. However, if the guided umbrella sampling is compared to the 17 window unguided umbrella sampling after 170 ns (shown in purple in Figure 2.6), the similarity of the guided umbrella sampling to the unguided begins to increase at the same time that its similarity to the exhaustive sampling begins to decrease, indicating that both the guided and unguided umbrella sampling simulations may have been converging toward the same final state. The K-S statistic for the final distributions of the guided and 17 window unguided umbrella sampling simulations is very low, only 0.018.

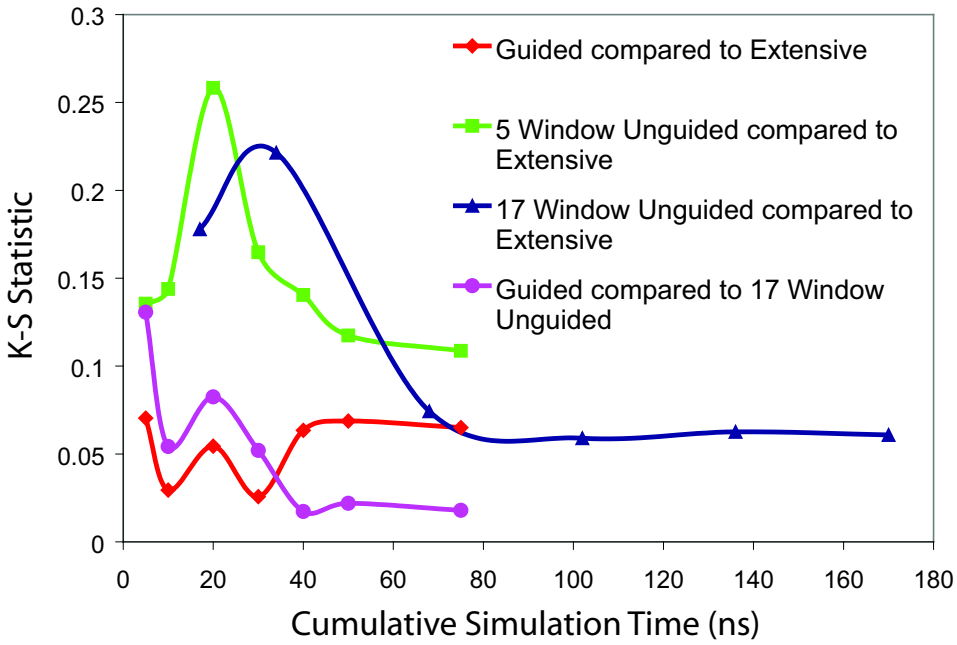

Figure 2.6: Kolmogorov-Smirnov test of similarity. The cumulative probability distribution functions for the distance between the sulfur of cysteine and the center of mass of the aromatic ring of tryptophan of the pentapeptide for the three umbrella sampling methods at different simulation times were compared to the exhaustive simulations at 0.5 μs. The K-S values for the 17 window umbrella sampling, 5 window umbrella sampling, and the 5 window guided umbrella sampling are shown. The K-S statistic for the guided umbrella sampling at various times as compared to the 17 window umbrella sampling at 170 ns is also shown. The lines are meant as a guide to the eye. A lower K-S value indicates higher likelihood that the distributions represent the same data.

Visual comparison of the probability curves and PMFs shows that the guided umbrella sampling is actually most like the exhaustive one at about 10 ns. The PMFs

of the extensive sampling after the full simulation time and, for the guided and 17 window unguided umbrella sampling simulations, at the point in the simulation at which they most closely resembled the guiding extensive sampling, at 10 ns and 68 ns respectively, are shown in Figure 3.4 A. The guided umbrella sampling result is the closest to the extensive sampling one. The 5 window unguided umbrella sampling, also shown after the full 75 ns, never resembles the extensive strongly. The final PMFs after convergence of the umbrella sampling techniques are compared to the extensive sampling in Figure 3.4 B. At this point the guided and 17 window unguided umbrella sampling techniques yield PMFs that are more similar to each other than to the one obtained from the extensive sampling, as indicated by the K-S calculations. These relationships are also shown in Figure 2.8, which shows the probability distributions of extensive sampling with the guided umbrella sampling at 10 ns and the 17 window unguided sampling at 68ns (A) and the guided sampling at 30 ns and unguided at 102 ns (B). Figure 2.8 C shows the final 17 window unguided umbrella sampling probability distribution with those for the guided at 50 and 75 ns.

In other words, after reaching a profile very close to that of the extensive sampling, the probability distributions of the guided and 17 window unguided umbrella sampling techniques begin to shift toward the right and eventually resemble each other more than the extensive.

A possible explanation for this drift away from the exhaustive sampling distribution is that there is a transition which is not seen after the 0.5 μs of "extensive" sampling, i.e., that the later did not have sufficient time to exhaustively sample to convergence of the right weight of the ξ variable. This possibility seems most likely given the large timescales involved in the relaxation dynamics of many biomolecules of similar size and the fact that the umbrella sampling distributions start out looking

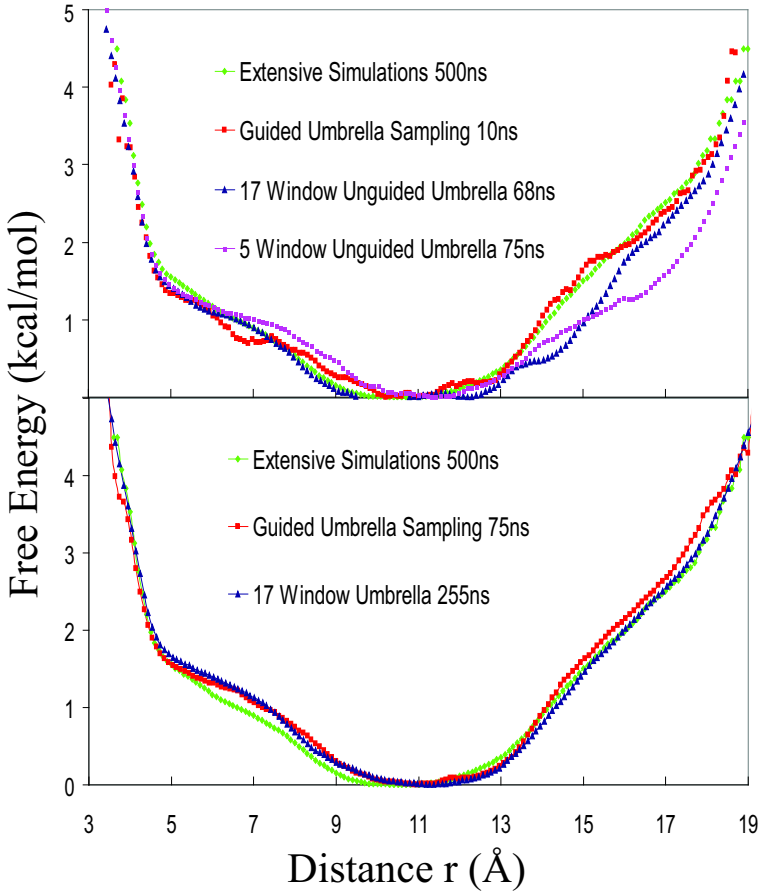

Figure 2.7: (A) Potentials of mean force for the end-to-end distance of the pentapeptide for the guiding extensive sampling at 0.5 μs and for the umbrella sampling methods at the time they most resemble the extensive run: i.e., at 68 ns for the 17 window unguided run and at 10 ns for the guided run. The 5 window unguided umbrella sampling never fully resembles the extensive simulation, not even after 75 ns. (B) PMFs for the 17 window unguided umbrella sampling after the full 255 ns and the guided umbrella sampling after the full 75 ns compared to 0.5 μs of extensive sampling.

similar to the exhaustive then gradually shift to a different distribution.

Time Scaling of the Sampling Efficiency. In order to test the possibility that there is a transition, the PMFs for the guided umbrella sampling were compared with the extensive simulations at various simulation times using the K-S statistic and visual comparison. The PMFs appear to line up at several points. Figure 2.9 A shows a plot of the extensive simulation times versus the guided umbrella sampling simulation times at which the PMFs match. The data was fit with an exponential of

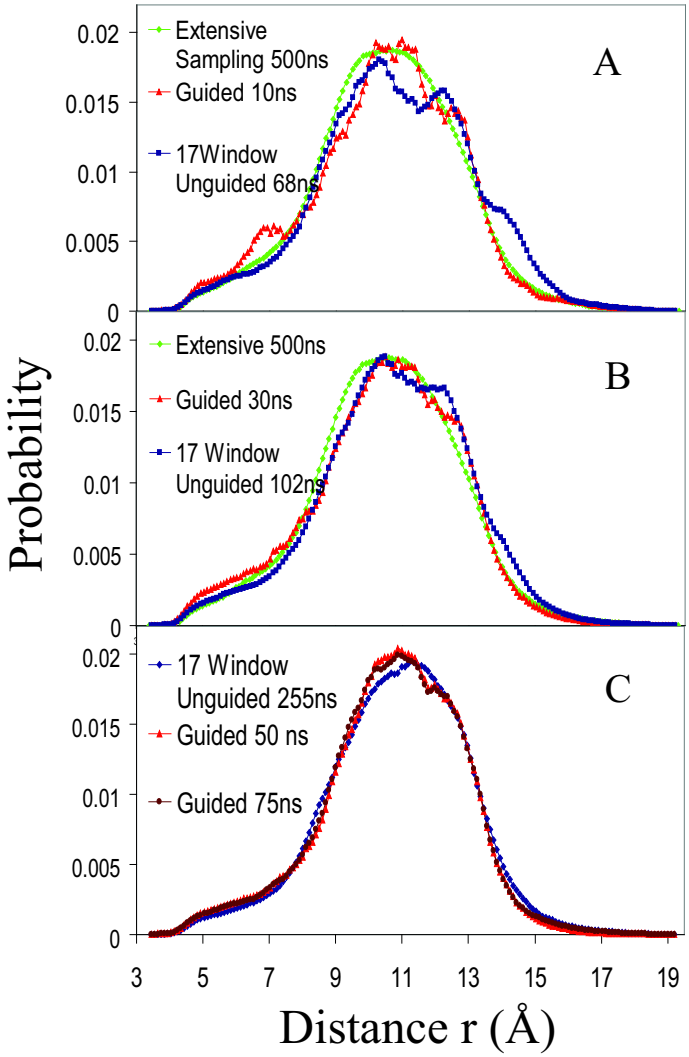

Figure 2.8: Comparison of probability curves of peptide end-to-end distance at various simulation times. (A) Exhaustive sampling at 0.5 μs, guided umbrella sampling at 10 ns (2 ns per window), and the 17 window unguided umbrella sampling at 68 ns (4 ns per window) . (B) Exhaustive sampling at 0.5 μs, guided umbrella sampling at 30 ns (6 ns per window), and 17 window unguided umbrella sampling at 102 ns (6 ns per window). It can be seen that the umbrella sampling probabilities are beginning to shift to the right. (C) Probability distribution for the unguided 17 window umbrella sampling is shown at 255 ns (15 ns per window) with the guided umbrella sampling at 50 ns and 75 ns. There is very little difference in the guided umbrella sampling between 50 and 75 ns.

the form $t_{\mathrm{ex}} = 1.707 \exp(0.485\, t_{\mathrm{gu}})$, where t_{ex} and t_{gu} are the total extensive sampling time and the total (cumulative) guided umbrella sampling time, respectively. This exponential was used to transform the guided umbrella sampling simulation with respect to time. A plot of the guided umbrella sampling PMF curve as a function

of cumulative simulation time is shown in Fig. 2.9 B. The PMF begins to shift after about 12 ns and reaches a converged state at approximately 40 ns. Figure 2.9 C shows the extensive PMF as a function of cumulative simulation time (purple) overlaid with the first 10 ns of the guided umbrella sampling PMF transformed with the exponential time relationship (red). The two plots line up extremely well with the exception of a slight decrease in the exhaustive around 12 Å from 50 to 200 ns. The plots do not match before 25 ns, most likely because the guided umbrella sampling simulations were started with the peptide in a conformation near the median reaction coordinate for each potential. There is no comparable relationship between the PMFs with respect to time for 17 window unguided umbrella sampling simulations and the extensive simulations.

Sampling of the "Perpendicular" Coordinates: Dihedral Angle Distributions. To determine if this method would yield accurate information for degrees of freedom to which no restraining potential was explicitly applied, we also compared the probability distributions of the peptide backbone dihedral angles Φ and Ψ.

Dihedral angle probability curves for all simulations were calculated using a modification of the WHAM equation described in equation 2.6. Since the peptide chosen contains only five amino acids, it was feasible to calculate the unbiased PMFs of all the Φ and Ψ dihedral angles. The individual angles from the 0.5 μs-long extensive sampling simulation were compared with those from the umbrella sampling methods over 68 ns for the 17-window unguided and 10 ns for the guided umbrella sampling protocols (Figure 2.10). The 5-window unguided umbrella sampling simulation was not compared as it failed to produce a valid distribution for the reaction coordinate.

In the particular case of backbone dihedrals for this system, a good amount of similarity is observed for the (unrestrained) dihedral distribution when comparing

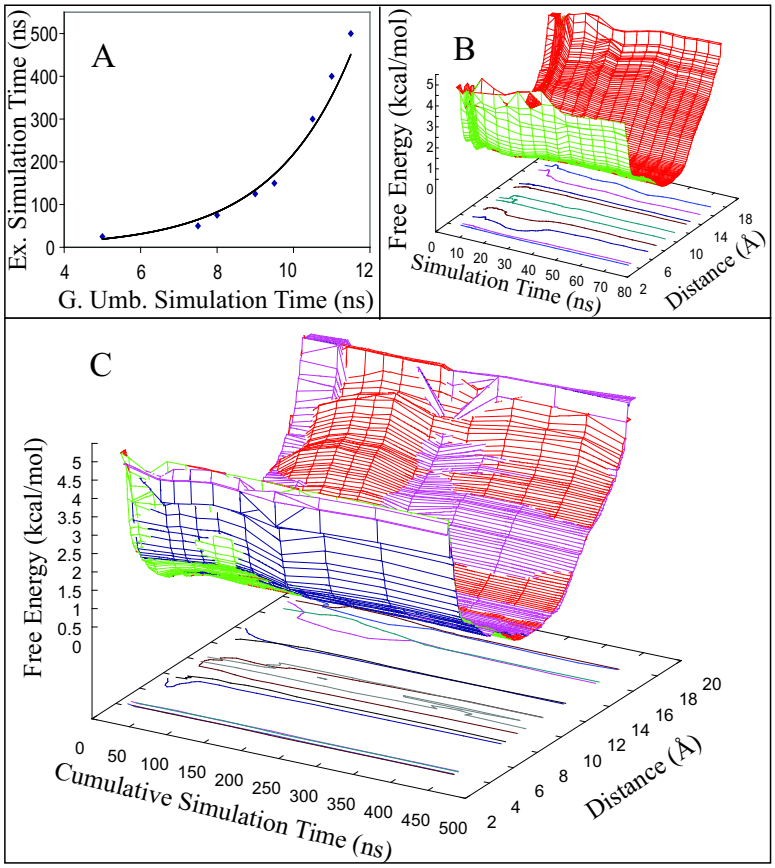

Figure 2.9: Comparison of the exhaustive and guided umbrella sampling PMFs for peptide end-to-end distance as a function of simulation time. (A) Plot of exhaustive sampling times versus the guided umbrella sampling times with a similar PMF; exponential fit yields $t_{\mathrm{ex}} = 1.707 \exp(0.485\ t_{\mathrm{gu}})$. (B) Guided umbrella sampling PMF as a function of simulation time. The time evolution of the potential of mean force of the reaction coordinate shows that the free energy well begins to narrow after about 25 ns and does not converge until about 45 ns. (C) Overlay of the PMFs from the exhaustive sampling (purple) and the guided umbrella sampling (red). The time of the guided umbrella sampling has been transformed using the exponential from panel A in order to make it line up with the exhaustive run results. The two simulations follow the same relative time path with respect to the reaction coordinate.

exhaustive sampling with umbrella sampling. However, there were exceptions, as detailed below. In general cases, the quality of the match will depend on the amount of coupling between the restrained and unrestrained coordinates. A suitable measure of the degree of relationship between two coordinates has been showed to be provided, when it can be calculated to convergence, by the mutual entropy [79], a high value for it indicating that the coordinates are correlated. This correlation is unlikely

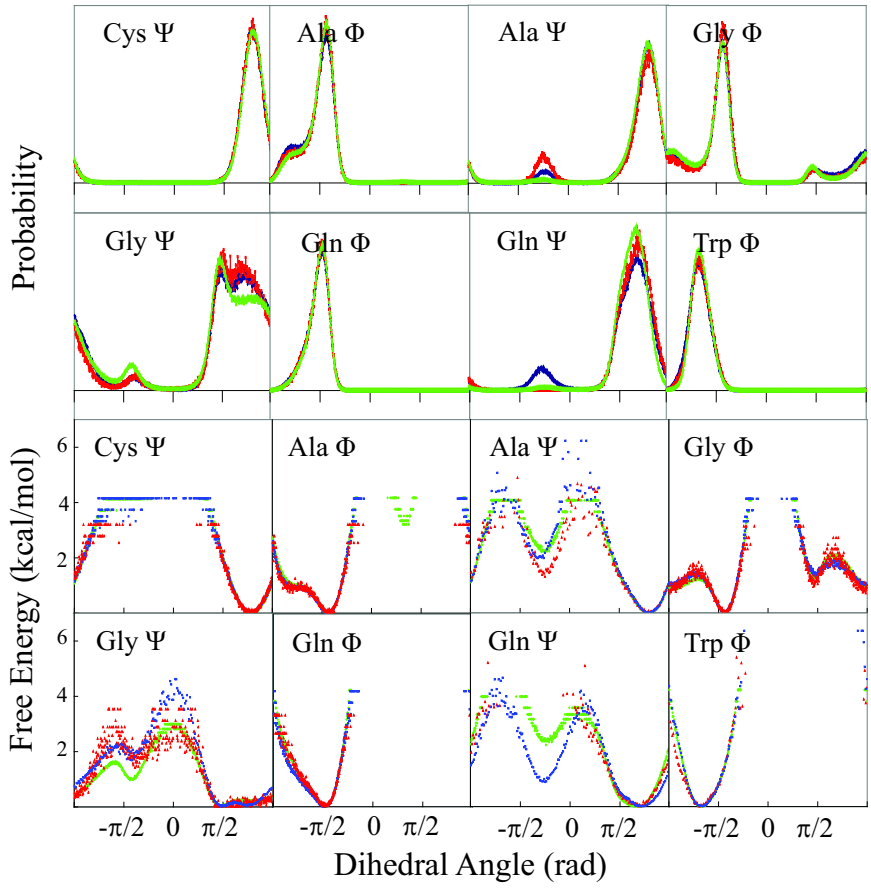

Figure 2.10: Probability distributions and free energy profiles for "perpendicular" degrees of freedom, i.e., the peptide dihedral angles. Exhaustive sampling distribution at 0.5 μs is shown in green, 17 window umbrella sampling (68 ns) in blue and 5 window guided umbrella sampling (10 ns) in red.

to be the case for the present two coordinates of choice (end-to-end distance and backbone dihedrals) because of the pliability of protein backbones. This means that guiding along the end-to-end distance will not enhance significantly the sampling of relevant backbone dihedral values. As described below, very rare events in the perpendicular degrees of freedom not coupled to the reaction coordinate may still not be seen with this method, and if properties of the perpendicular manifold are desired, then multi-dimensional umbrella sampling (i.e., sampling along two or more reaction coordinates) should be employed [88]. In this regard, it would also be interesting to apply multi-dimensional umbrella sampling using guiding data from three-color

FRET experiments [19, 20] that can report up to three interfluorophore distances.

Despite the general agreement, there were angles which differed in terms of their sampled distributions between the three numerical experiments significantly: the Φ and Ψ angles of alanine, the Ψ angle of glycine, and and the Ψ angle of glutamine. The differences between the Ψ angle of alanine for the three techniques are relatively small and may be due in part to the larger range of flexibility of this angle due to the small size of the side chain. The same is true of the glycine Ψ angle, although in this case the guided umbrella sampling yielded results that were more like those from extensive sampling than from the unguided one. The most significant differences are in the Φ dihedral of alanine and Ψ dihedral of glutamine. The extensive simulation gives a small well in the PMF of alanine Φ which is not seen in the other sampling methods. This well is the result of a single event during the 0.5 μs of simulations. The other difference which appears to be significant is in the Ψ angle of glutamine (see Fig. 2.10). The well at about -$\pi/4$ is overpopulated in the 17-window unguided umbrella sampling and not seen in the guided umbrella sampling. The guided umbrella sampling protocol appears to have partially brought about crossing of the barrier, but then returned to the more favorable angle. This well also represents occurrences of a rare event, which took place only four times during the extensive simulations. The fact that non-convergence of the the Φ dihedral of alanine and Ψ dihedral of glutamine did not affect the convergence of the distribution of ξ further emphasizes the fact that neither of these dihedrals is expected to be coupled to the reaction coordinate (i.e., that integration over the Boltzmann factor of the dihedral energy term comes out as a constant in the expression for the PMF).

Ideally, in general cases, any efficient umbrella sampling simulations would visit every possible configuration in the directions perpendicular to the reaction coor-

dinate. Unfortunately, given the limited simulation time, this may not always be possible. For example, as seen here, for singular events such as those which contribute to the differences in the PMFs for the alanine Φ and glutamine Ψ dihedrals, it may not be possible to obtain a completely accurate profile without enhancing sampling in those directions. It may also be the case that our extensive sampling was not exhaustive enough to produce accurate PMFs for these coordinates. Regardless, in the particular case of weak coupling between the perpendicular coordinates and the reaction coordinate, this lack of accuracy might not affect significantly the quality of the PMF along the reaction coordinate. A related example has to do with the calculation of order parameters that report on the motion of bond vectors in proteins as measured by solution NMR relaxation experiments [89]. Computer simulation of these experiments have shown that there is good agreement with experiment for backbone order parameters (which can be calculated well with a simple harmonic vibrational approximation [90, 91, 92] or by an even simpler contact model [93]). By contrast, for order parameters of side chains (which are the "perpendicular coordinates" in the language of our previous discussion) agreement to experiment is relatively poorer [94, 95] and a more thorough dynamical sampling is needed [96]. The relatively larger difficulty in converging side-chain vs. backbone parameters is because of the more complex motion of the side chains, but the lack of convergence for side chains does not preclude the ability to calculate accurate backbone parameters, which would indicate that there is relatively limited coupling between the backbone and the side chains, at least in what the order parameter estimator is concerned.

2.5.2 Experimentally guided umbrella sampling of forced unfolding pathway of Titin I27.

Guided umbrella sampling on the forced unfolding of titin I27 similarly result in faster and more accurate calculations of the potential of mean force than umbrella sampling simulations with arbitrarily chosen potentials.

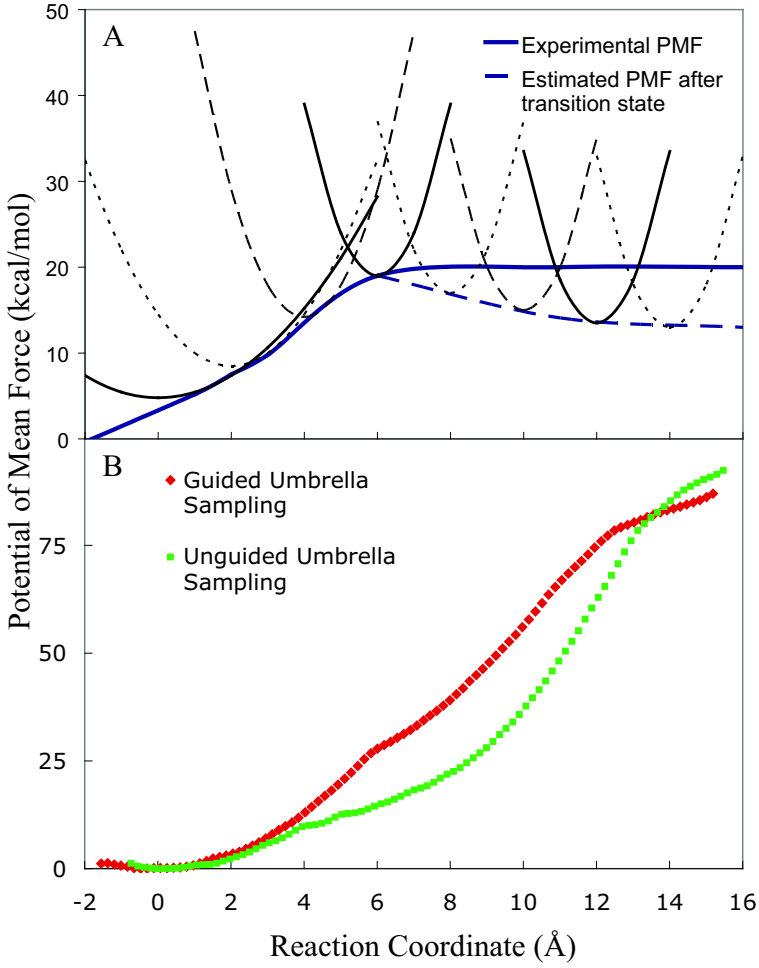

Figure 2.11: A. Experimental PMF and a selection of umbrella sampling potentials used in experimentally guided umbrella sampling simulations of unfolding of titin I27. As the simulations progress toward higher energy states, the force constant increases. The Potential immediately after the transition state cannot be accurately determined from pulling experiments due to the cantilever snapping at this point. The estimated free energy for this region is shown in dashed lines. B. Potentials of mean force for guided and unguided umbrella sampling simulations. The two methods produce substantially different curves.

Figure 2.11 A shows the experimental PMF with a sample of the restraining

potentials used. In order to insure adequate sampling, potentials were fit to the curve every Å initially and every 0.5 Å where 1 Å did not provide sufficient overlap. For these simulations we have used the assumption that the equilibrium state of titin I27 is approximately 6 Å from the transition state, as stated in the Harris et al. As can be seen from the figure, at the low energy equilibrium state the force constant of the umbrella potential is low. As the free energy increases, so do the force constants. This makes intuitive sense, as increasingly strong potentials will be needed to drive the protein away from its native state. The area of the curve immediately after the transition state can not be determined directly from experiment, due to snapping of the cantilever after the domain begins to unfold. In this region a constant k value of 5 $kcal/mol/\mathring{A}^2$ was used. Figure 2.11 B shows the PMFs calculated from the guided umbrella sampling simulation and the unguided umbrella sampling with a constant k value of 10 kcal/mol/Å^2 every 0.5 Å. The guided umbrella sampling PMF shows an inflection point at 6 Å but continues to increase after this point until reaching a plateau at 12.5 Å. The guided umbrella sampling PMF also has a slight inflection point at 1.7 Å (see Figure 2.12). The free energy change at this distance for the guided umbrella sampling is 2.4 kcal/mol. This is in good agreement with AFM experiments by Williams et al. [56], which predict an intermediate state at 2.2 Å with a free energy difference of 3.2 kcal/mol between the intermediate and native state. This intermediate corresponds to the stretching and breaking of the first hydrogen bonds between the A and G β strands. Previous constant force simulations have also shown the first event in titin unfolding corresponds to a discontinuous breaking of these hydrogen bonds from 2 to 6.6 Å [57]. The free energy value at r=6 is 26.8 kcal/mol, considerably larger than the experimental value of 11.4 kcal/mol [16]. This is not entirely surprising, since our simulations use an implicit solvent

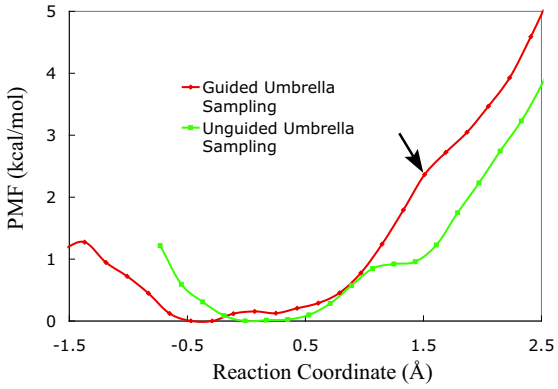

Figure 2.12: Detail of PMFs for guided and unguided umbrella sampling of titin. The guided PMF has an inflection point at 1.7 Å (arrow) and a free energy change of 2.4 kcal/mol, in agreement with experimental predictions.

model optimized for folded proteins, and similar differences between free energies calculated by simulation and experiment have been reported previously [97, 98]. On the other hand, denaturation experiments predict a ΔG of 22.2 kcal/mol [99], and the free energy reconstructed from biased MD simulations yield a value for ΔG of 23 kcal/mol [70]. The PMF of the unguided umbrella sampling has a significantly different shape from the guided PMF, with an inflection point at 4 Å and a slight plateau beginning at 13.5 Å.

Snapshots at different values of r for the two sets of umbrella sampling simulations (Figure 2.13 A and B) reveal the structural basis for the differences in the PMFs. In the both the guided and unguided umbrella sampling simulations, at r=2 Å h-bonds between β strand A and β strand G are stretched. The pathways, however, have diverged by r=7 Å. In the guided umbrella sampling simulations, the N terminal strand, consisting of β strands A and A' is beginning to pull away from the C-terminal β strand G, and at 13 Å the N terminal strand is completely disassociated from the rest of the domain. The guided simulations are in agreement with the forced unfolding pathway as determined by experimental ϕ values and forced unfolding simulations [71, 100]. In this previously determined pathway, titin I27 initially unfolds to an

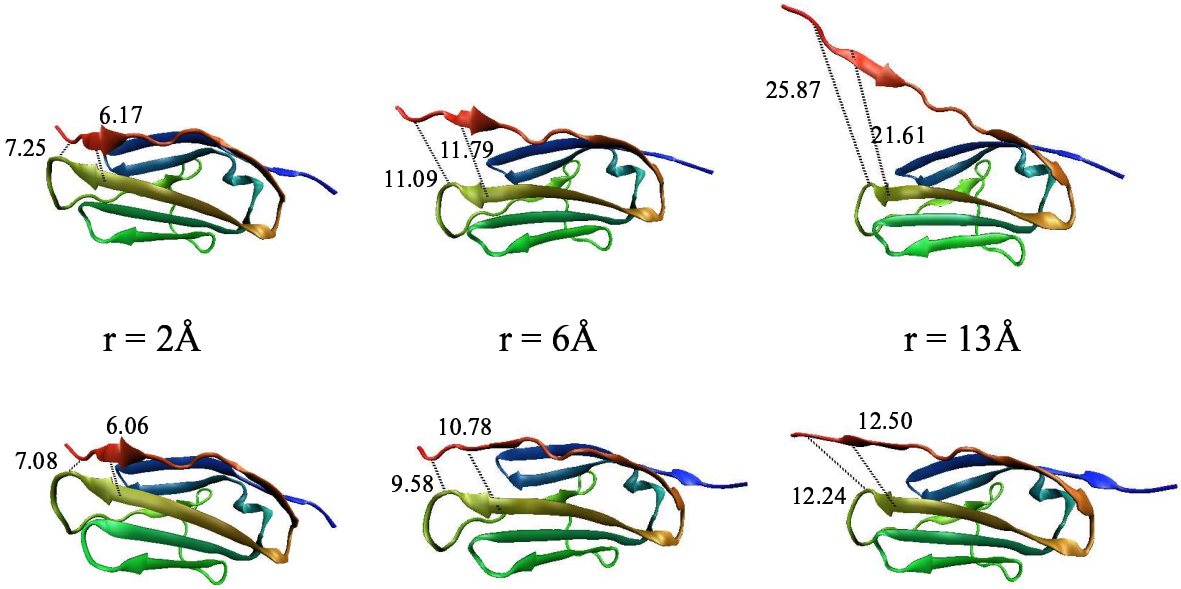

Figure 2.13: Snapshots from the guided (A) and unguided (B) umbrella sampling simulations of forced unfolding of titin I27 at r values of 2 Å, 7 Å, and 13 Å where r is the distance between the N and C terminus of the protein minus the native distance. The guided umbrella sampling simulations reveal the intermediate and transition states seen in forced unfolding simulations by Best et al. [100]. The unguided umbrella sampling simulations fail to replicate these results. The C_α to C_α distances between residues 2 and 26 and 4 and 24 are highlighted with black lines. The native structure distances for these residues are 5.72 and 6.20 Å respectively.

intermediate state with the N-terminal beta sheet partially separated from the rest of the domain before reaching the transition state, in which the N and C-terminal strands are completely separated from each other. The unguided umbrella sampling simulations (Figure 2.13, B), on the other hand, do not replicate this pathway. At 6 Å the N-terminal beta strand is stretched but not separated from the domain. At 13 Å the N terminal strand has begun to pull away, but the overall structure looks more like the guided umbrella sampling intermediate at 6 Å than the more extended transition state. In these simulations, the higher forces seem to be creating a shearing effect, in which the A and A' β strands slide along the length of the G strand rather than becoming disassociated from it as in the guided simulations. Our

results indicate that despite the fact that the guiding PMF was incomplete beyond the transition state, using the experimental PMF as a guide for simulations resulted in significant improvement in the accuracy of umbrella sampling.

These results also highlight the need to be careful when choosing umbrella sampling potentials, as naively chosen restraining potentials resulted in the system following an apparently artificial pathway.

2.6 Concluding Discussion

We have presented a theoretical analysis of the feasibility of an experimentally guided umbrella sampling calculation protocol. Using guiding restraining potentials based on an experimental free energy profile derived from measured equilibrium distributions of a conformational variable, we suggest a method that is able to "guide" umbrella sampling simulations to a converged potential of mean force along that variable. Using the distribution of the end-to-end distance of a pentapeptide model system, the guided umbrella sampling was more efficient than regular, unguided umbrella sampling, requiring fewer windows and giving the same results in less than one sixth the total simulation time. This does not count the simulation time of discarded unguided umbrella sampling runs with restraining potentials which were not effective. The new method also outperformed an extensive sampling done with regular molecular dynamics of 0.5 μs total time. Our guided umbrella sampling also provided potentials of mean force for unbiased reaction coordinates, the dihedral angles of the peptide, which were similar to those of the extensive sampling, with a few notable exceptions. Given the fact that they did not affect the convergence of the PMF along the constrained coordinate, the exceptions were attributed to negligible coupling between the reaction coordinate values and the respective dihedral angle

values.

Our results also indicated that, with the guided umbrella sampling procedure, there was a transition of the converged free energy profile of the peptide which went beyond that obtained from the 0.5 μs of direct sampling, but which is seen with regular umbrella sampling. This is consistent with explicit water simulations of the same peptide and others [45], which needed μs scale simulations to converge the results of contact formation observables in peptide loops such that they matched the corresponding experimental values [44]. Interestingly, although the guided umbrella simulations on the peptide were based on the PMF from the extensive simulations, the distribution of the reaction coordinate began to look more like the 17 window regular umbrella sampling simulations after about 30 ns cumulative time. This implies that the experimentally guided umbrella sampling protocol we propose will be effective even when the experimental PMF used to determine the restraining potentials is not an exactly accurate PMF. This holds true for the titin test case, in which experimentally guided umbrella sampling produces better results than unguided, despite errors inherent in both experimental and simulation methods.

The comparison of the guided umbrella sampling and direct exhaustive molecular dynamics of the pentapeptide as a function of simulation time shows that the guided umbrella sampling simulations follow roughly the same time course as the exhaustive simulation result, but does so exponentially faster. The time plots also show how the guided PMF evolves in time past the profile reached by exhaustive sampling. Similarly, for the forced unfolding of titin I27, guided umbrella sampling simulations follow an unfolding pathway in agreement with previous simulations and experiment [71, 100], whereas umbrella sampling with arbitrarily chosen force constants does not.

Although the usefulness of our proposed technique obviously depends on the availability of experimental data, we believe that the potential applications are promising. The recent advances in single-molecule techniques will no doubt lead to the determination of more experimental potentials of mean force in the coming years, expanding the potential uses of our method. While errors in both FRET and force manipulation experiments can lead to inaccuracies in the resulting PMFs [101, 102], our work shows that umbrella sampling is improved even when the guiding potential is not the exact PMF, in the case of the peptide, or incomplete, in the case of titin I27. Having the experimental guiding potential for large complex biomolecules might prove useful because it can guide sampling in regions of configuration space that would not even be visited by the other methods (because of long time sampling problems). As such, properties of perpendicular degrees of freedom might be revealed that would not have been gauged before. Just as techniques such as transition path sampling [103] or targeted molecular dynamics [104, 102] use information about the final state of a system to direct simulations that reveal information about a pathway, this technique can use experimental information about a reaction coordinate to direct simulations in a way that can reveal more detailed information about a system in which details of the final state may not be known.

In the same vein, there may be exciting applications that have to do with heterogeneity and the apparent non-ergodicity [105] observed in single molecule studies of complex molecules. For example, in the case of the hairpin ribozyme, single molecule FRET revealed complex structural dynamics that would have been difficult to detect in ensemble measurements. This ribozyme, which undergoes a conformational change from an undocked to a docked conformation has several distinct docked/undocked substrates with different kinetic rates of undocking, and the ribozyme exhibited

a memory effect, where switching between different docked substates was rarely seen [106]. This apparent non-Markovian, memory effect in the hairpin ribozyme is presumably due to different conformations of two loops within the molecule [106]. These different conformations cannot be directly detected experimentally, but could be probed during a molecular dynamics simulation that is guided along a distinct experimental (single-molecule) PMF.

This work has been published in the Journal of Chemical Physics [107].

BIBLIOGRAPHY

[1] J.G. Kirkwood. Statistical mechanics of fluid mixtures. *Journal of Chemical Physics*, 3:300–313, 1935.

[2] B. Roux and M. Karplus. Ion-transport in a model gramicidin channel - structure and thermodynamics. *Biophysical Journal*, 59:961–981, 1991.

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

[4] D. Andrews and A. Demidov. *Resonance Energy Transfer*. John Wiley Sons, 1999.

[5] A. Cooper. Thermodynamics of fluctuations in protein molecules. *Proceedings of The National Academy of Sciences of The United States of America*, 73:2740–2741, August 1976.

[6] A. A. Deniz, T. A. Laurence, G. S. Beligere, M. Dahan, A. B. Martin, D. S. Chemla, P. E. Dawson, P. G. Schultz, and S. Weiss. Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proceedings of The National Academy of Sciences of The United States of America*, 97:5179–5184, 2000.

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

[8] N. G. Walter, D. A. Harris, M. J. B. Pereira, and D. Rueda. In the fluorescent spotlight: Global and local conformational changes of small catalytic RNAs. *Biopolymers*, 61:224–242, 2001.

[9] M. Carrion-Vazquez, A. F. Oberhauser, T. E. Fisher, P. E. Marszalek, H. B. Li, and J. M. Fernandez. Mechanical design of proteins-studied by single-molecule force spectroscopy and protein engineering. *Progress In Biophysics & Molecular Biology*, 74:63–91, 2000.

[10] 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.

[11] T. R. Strick, M. N. Dessinges, G. Charvin, N. H. Dekker, J. F. Allemand, D. Bensimon, and V. Croquette. Stretching of macromolecules and proteins. *Reports on Progress In Physics*, 66:1–45, 2003.

[12] G. Hummer and A. Szabo. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proceedings of The National Academy of Sciences of The United States of America*, 98:3659–3661, 2001.

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

[14] J. Nummela, F. Yassin, and I. Andricioaei. Entropy-energy decomposition from nonequilibrium work trajectories. *Journal of Chemical Physics*, 128:024104–024107, 2008.

[15] J. Nummela and I. Andricioaei. Exact low-force kinetics from high-force single-molecule unfolding events. *Biophysical Journal*, 93:3373–3381, 2007.

[16] N. C. Harris, Y. Song, and C. H. Kiang. Experimental free energy surface reconstruction from single-molecule force spectroscopy using Jarzynski's equality. *Physical Review Letters*, 99:068101–068104, 2007.

[17] J. Preiner, H. Janovjak, C. Rankl, H. Knaus, D. A. Cisneros, A. Kedrov, F. Kienberger, D. J. Muller, and P. Hinterdorfer. Free energy of membrane protein unfolding derived from single-molecule force measurements. *Biophysical Journal*, 93:930–937, 2007.

[18] M. T. Woodside, P. C. Anthony, W. M. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. M. Block. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science*, 314:1001–1004, 2006.

[19] S. Hohng, C. Joo, and T. Ha. Single-molecule three-color FRET. *Biophysical Journal*, 87:1328–1337, 2004.

[20] J. P. Clamme and A. Deniz. Three-color single molecule FRET studies of biomolecules. *Biophysical Journal*, 86:326A–326A, 2004.

[21] M. Karplus. Molecular dynamics simulations of biomolecules. *Accounts of Chemical Research*, 35:321–323, 2002.

[22] 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.

[23] A. M. Ferrenberg and R. H. Swendsen. Optimized Monte-Carlo data-analysis. *Physical Review Letters*, 63:1195–1198, 1989.

[24] A. M. Ferrenberg. New Monte Carlo technique for studying phase transitions. *Physical Review Letters*, 63:1658–1658, 1989.

[25] 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 .1. the method. *Journal of Computational Chemistry*, 13:1011–1021, 1992.

[26] M. Mezei. Adaptive umbrella sampling - self-consistent determination of the non-boltzmann bias. *Journal of Computational Physics*, 68:237–248, 1987.

[27] R. W. W. Hooft, B. P. Vaneijck, and J. Kroon. An adaptive umbrella sampling procedure in conformational-analysis using molecular-dynamics and its application to glycol. *Journal of Chemical Physics*, 97:6690–6694, 1992.

[28] C. Bartels and M. Karplus. Multidimensional adaptive umbrella sampling: Applications to main chain and side chain peptide conformations. *Journal of Computational Chemistry*, 18:1450–1462, 1997.

[29] C. Bartels and M. Karplus. Probability distributions for complex systems: Adaptive umbrella sampling of the potential energy. *Journal of Physical Chemistry B*, 102:865–880, 1998.

[30] M. Schaefer, C. Bartels, and M. Karplus. Solution conformations and thermodynamics of structured peptides: Molecular dynamics simulation with an implicit solvation model. *Journal of Molecular Biology*, 284:835–848, 1998.

[31] E. Darve, M. A. Wilson, and A. Pohorille. Calculating free energies using a scaled-force molecular dynamics algorithm. *Molecular Simulation*, 28:113–144, 2002.

[32] A. Laio and M. Parrinello. Escaping free-energy minima. *Proceedings of The National Academy of Sciences of The United States of America*, 99:12562–12566, 2002.

[33] B. A. Berg and T. Neuhaus. Multicanonical algorithms for 1st order phase-transitions. *Physics Letters B*, 267:249–253, 1991.

[34] B. A. Berg and T. Neuhaus. Multicanonical ensemble - A new approach to simulate 1st-order phase-transitions. *Physical Review Letters*, 68:9–12, 1992.

[35] U. H. E. Hansmann and Y. Okamoto. Prediction of peptide conformation by multicanonical algorithm - New approach to the multiple-minima problem. *Journal of Computational Chemistry*, 14:1333–1338, 1993.

[36] F. G. Wang and D. P. Landau. Efficient, multiple-range random walk algorithm to calculate the density of states. *Physical Review Letters*, 86:2050–2053, 2001.

[37] J. Lee. New Monte-Carlo algorithm - Entropic sampling. *Physical Review Letters*, 71:211–214, 1993.

[38] N. . Rathore, T. .A. . Knotts IV, and J. .J. . de Pablo. Density-of-states based Monte Carlo techniques for simulation of proteins and polymers. *AIP Conference Proceedings*, no.690:289–98, 2003.

[39] M. S. Shell, P. G. Debenedetti, and A. Z. Panagiotopoulos. An improved Monte Carlo method for direct calculation of the density of states. *Journal of Chemical Physics*, 119:9406–9411, 2003.

[40] Y. Kim and H. Ling. On the optimal sampling strategy for model-based parameter estimation using rational functions. *IEEE Transactions on Antennas and Propagation*, 54:762–765, 2006.

[41] T. C. Beutler and W. F. van Gunsteren. The computation of a potential of mean force - choice of the biasing potential in the umbrella sampling technique. *Journal of Chemical Physics*, 100:1492–1497, 1994.

[42] R. B. Best and M. Vendruscolo. Determination of protein structures consistent with NMR order parameters. *Journal of The American Chemical Society*, 126:8090–8091, 2004.

[43] M. Vendruscolo, E. Paci, C. M. Dobson, and M. Karplus. Rare fluctuations of native proteins sampled by equilibrium hydrogen exchange. *Journal of The American Chemical Society*, 125:15686–15687, 2003.

[44] L. J. Lapidus, W. A. Eaton, and J. Hofrichter. Measuring the rate of intramolecular contact formation in polypeptides. *Proceedings of The National Academy of Sciences of The United States of America*, 97:7220–7225, 2000.

[45] I. C. Yeh and G. Hummer. Peptide loop-closure kinetics from microsecond molecular dynamics simulations in explicit solvent. *Journal of The American Chemical Society*, 124:6563–6568, 2002.

[46] K. Sudhakar, C. M. Phillips, C. S. Owen, and J. M. Vanderkooi. Dynamics of parvalbumin studied by fluorescence emission and triplet absorption-spectroscopy of tryptophan. *Biochemistry*, 34:1355–1363, 1995.

[47] G. B. Strambini and M. Gonnelli. Tryptophan phosphorescence in fluid solution. *Journal of The American Chemical Society*, 117:7646–7651, 1995.

[48] D. V. Bent and E. Hayon. Excited-state chemistry of aromatic amino-acids and related peptides .3. Tryptophan. *Journal of The American Chemical Society*, 97:2612–2619, 1975.

[49] M. Gonnelli and G. B. Strambini. Phosphorescence lifetime of tryptophan in proteins. *Biochemistry*, 34:13847–13857, 1995.

[50] D. Roccatano, W. M. Nau, and M. Zacharias. Structural and dynamic properties of the cagqw peptide in water: A molecular dynamics simulation study using different force fields. *Journal of Physical Chemistry B*, 108:18734–18742, 2004.

[51] M. J. Feige and E. Paci. Rate of loop formation in peptides: A simulation study. *Journal of Molecular Biology*, 382:556–565, 2008.

[52] L. Tskhovrebova and J. Trinick. Properties of titin immunoglobulin and fibronectin-3 domains. *Journal of Biological Chemistry*, 279:46351–46354, 2004.

[53] J. Trinick and L. Tskhovrebova. Titin: a molecular control freak. *Trends in Cell Biology*, 9:377–380, 1999.

[54] W. A. Linke, M. R. Stockmeier, M. Ivemeyer, H. Hosser, and P. Mundel. Characterizing titin's I-band IG domain region as an entropic spring. *Journal of Cell Science*, 111:1567–1574, 1998.

[55] L. Tskhovrebova, J. Trinick, J. A. Sleep, and R. M. Simmons. Elasticity and unfolding of single molecules of the giant muscle protein titin. *Nature*, 387:308–312, 1997.

[56] P. M. Williams, S. B. Fowler, R. B. Best, J. L. Toca-Herrera, K. A. Scott, A. Steward, and J. Clarke. Hidden complexity in the mechanical properties of titin. *Nature*, 422:446–449, 2003.

[57] P. E. Marszalek, H. Lu, H. B. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten, and J. M. Fernandez. Mechanical unfolding intermediates in titin modules. *Nature*, 402:100–103, 1999.

[58] H. Li, A. Oberhauser, S. Fowler, J. Clarke, P. E. Marszalek, and J. M. Fernandez. Mechanical unfolding of titin is under kinetic control. *Biophysical Journal*, 78:2371, 2000.

[59] M. Rief, M. Gautel, and H. E. Gaub. Unfolding forces of titin and fibronectin domains directly measured by afm. *Elastic Filaments of The Cell*, 481:129–141, 2000.

[60] K. Wang, J. G. Forbes, and A. J. Jin. Single molecule measurements of titin elasticity. *Progress in Biophysics & Molecular Biology*, 77:1–44, 2001.

[61] M. S. Z. Kellermayer, S. Smith, C. Bustamante, and H. L. Granzier. Mechanical manipulation of single titin molecules with laser tweezers. *Elastic Filaments of The Cell*, 481:111–128, 2000.

[62] M. S. Z. Kellermayer, S. B. Smith, C. Bustamante, and H. L. Granzier. Complete unfolding of the titin molecule under external force. *Journal of Structural Biology*, 122:197–205, 1998.

[63] H. S. C. Martin, S. Jha, S. Howorka, and P. V. Coveney. Determination of free energy profiles for the translocation of polynucleotides through alpha-hemolysin nanopores using non-equilibrium molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 5:2135–2148, 2009.

[64] A. Minajeva, M. Kulke, J. M. Fernandez, and W. A. Linke. Unfolding of titin domains explains the viscoelastic behavior of skeletal myofibrils. *Biophysical Journal*, 80:1442–1451, 2001.

[65] J. L. Marin, J. Muniz, M. Huerta, and X. Trujillo. Folding-unfolding of immunoglobulin domains in titin: A simple two-state model. *General Physiology and Biophysics*, 18:305–309, 1999.

[66] B. Zhang, G. Z. Xu, and J. S. Evans. A kinetic molecular model of the reversible unfolding and refolding of titin under force extension. *Biophysical Journal*, 77:1306–1315, 1999.

[67] G. Pabon and L. M. Amzel. Mechanism of titin unfolding by force: Insight from quasi-equilibrium molecular dynamics calculations. *Biophysical Journal*, 91:467–472, 2006.

[68] R. D. Toofanny and P. M. Williams. Simulations of multi-directional forced unfolding of titin I27. *Journal of Molecular Graphics & Modelling*, 24:396–403, 2006.

[69] M. Cieplak, A. Pastore, and T. X. Hoang. Mechanical properties of the domains of titin in a Go-like model. *Journal of Chemical Physics*, 122:54906, 2005.

[70] P. C. Li and D. E. Makarov. Theoretical studies of the mechanical unfolding of the muscle protein titin: Bridging the time-scale gap between simulation and experiment. *Journal of Chemical Physics*, 119:9260–9268, 2003.

[71] S. B. Fowler, R. B. Best, J. L. T. Herrera, T. J. Rutherford, A. Steward, E. Paci, M. Karplus, and J. Clarke. Mechanical unfolding of a titin IG domain: Structure of unfolding intermediate revealed by combining AFM, molecular dynamics simulations, NMR and protein engineering. *Journal of Molecular Biology*, 322:841–849, 2002.

[72] M. Gao, H. Lu, and K. Schulten. Simulated refolding of stretched titin immunoglobulin domains. *Biophysical Journal*, 81:2268–2277, 2001.

[73] M. Gao, B. Isralewitz, H. Lu, and K. Schulten. Steered molecular dynamics simulation of the unfolding of multiple titin immunoglobulin domains. *Biophysical Journal*, 78:156, 2000.

[74] H. Lu and K. Schulten. The key event in force-induced unfolding of titin's immunoglobulin domains. *Biophysical Journal*, 79:51–65, 2000.

[75] H. Lu and K. Schulten. Steered molecular dynamics simulation of conformational changes of immunoglobulin domain I27 interpret atomic force microscopy observations. *Chemical Physics*, 247:141–53, 1999.

[76] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophysical Journal*, 75:662–71, 1998.

[77] C. Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78:2690–2693, 1997.

[78] E. Gallicchio, M. Andrec, A. K. Felts, and R. M. Levy. Temperature weighted histogram analysis method, replica exchange, and transition paths. *Journal of Physical Chemistry B*, 109:6722–6731, 2005.

[79] D. D. L. Minh. Multidimensional potentials of mean force from biased experiments along a single coordinate. *Journal of Physical Chemistry B*, 111:4137–4140, 2007.

[80] M. N. Kobrak. Systematic and statistical error in histogram-based free energy calculations. *Journal of Computational Chemistry*, 24:1437–1446, 2003.

[81] M. Schaefer and M. Karplus. A comprehensive analytical treatment of continuum electrostatics. *Journal of Physical Chemistry*, 100:1578–1599, 1996.

[82] 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. *Journal of Computational Chemistry*, 4:187–217, 1983.

[83] 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.

[84] 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.

[85] A. Grossfield. http://dasher.wustl.edu/alan/.

[86] H. Lilliefo. On Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of The American Statistical Association*, 62:399–402, 1967.

[87] 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.

[88] 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.

[89] G. Lipari, A. Szabo, and R. M. Levy. Protein dynamics and NMR relaxation - comparison of simulations with experiment. *Nature*, 300:197–198, 1982.

[90] E. R. Henry and A. Szabo. Influence of vibrational motion on solid-state line-shapes and NMR relaxation. *Journal of Chemical Physics*, 82:4753–4761, 1985.

[91] R. Bruschweiler and D. A. Case. Characterization of biomolecular structure and dynamics by NMR cross-relaxation. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 26:27–58, 1994.

[92] S. Sunada and N. Go. Calculation of nuclear magnetic resonance order parameters in proteins by normal mode analysis .2. contribution from localized high frequency motions. *Journal of Chemical Physics*, 105:6560–6564, 1996.

[93] F. L. Zhang and R. Bruschweiler. Contact model for the prediction of NMR N-H order parameters in globular proteins. *Journal of The American Chemical Society*, 124:12654–12655, 2002.

[94] L. J. Smith, A. E. Mark, C. M. Dobson, and W. F. van Gunsteren. Comparison of MD simulations and NMR experiments for hen lysozyme - analysis of local fluctuations, cooperative motions, and global changes. *Biochemistry*, 34:10918–10931, 1995.

[95] D. M. Ming and R. Bruschweiler. Prediction of methyl-side chain dynamics in proteins. *Journal of Biomolecular NMR*, 29:363–368, 2004.

[96] R. B. Best, J. Clarke, and M. Karplus. The origin of protein sidechain order parameter distributions. *Journal of The American Chemical Society*, 126:7734–7735, 2004.

[97] R. H. Zhou and B. J. Berne. Can a continuum solvent model reproduce the free energy landscape of a beta-hairpin folding in water? *Proceedings of The National Academy of Sciences of The United States of America*, 99:12777–12782, 2002.

[98] H. Nymeyer and A. E. Garcia. Simulation of the folding equilibrium of alpha-helical peptides: A comparison of the generalized born approximation with explicit solvent. *Proceedings of The National Academy of Sciences of The United States of America*, 100:13934–13939, 2003.

[99] M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez. Mechanical and chemical unfolding of a single protein: A comparison. *Proceedings of The National Academy of Sciences of The United States of America*, 96:3694–3699, 1999.

[100] R. B. Best, S. B. Fowler, J. L. T. Herrera, A. Steward, E. Paci, and J. Clarke. Mechanical unfolding of a titin IG domain: Structure of transition state revealed by combining atomic force microscopy, protein engineering and molecular dynamics simulations. *Journal of Molecular Biology*, 330:867–877, 2003.

[101] B. Schuler, E. A. Lipman, and W. A. Eaton. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, 419:743–747, 2002.

[102] K. A. Merchant, R. B. Best, J. M. Louis, I. V. Gopich, and W. A. Eaton. Characterizing the unfolded states of proteins using single-molecule fret spectroscopy and molecular simulations. *Proceedings of The National Academy of Sciences of The United States of America*, 104:1528–1533, 2007.

[103] C. Dellago, P. G. Bolhuis, and P. L. Geissler. Transition path sampling. *Advances in Chemical Physics, Vol 123*, 123:1–78, 2002.

[104] 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.

[105] Z. Xie, N. Srividya, T. R. Sosnick, T. Pan, and N. F. Scherer. Single-molecule studies highlight conformational heterogeneity in the early folding steps of a large ribozyme. *Proceedings of The National Academy of Sciences of The United States of America*, 101:534–539, 2004.

[106] X. W. Zhuang, H. Kim, M. J. B. Pereira, H. P. Babcock, N. G. Walter, and S. Chu. Correlating structural dynamics and function in single ribozyme molecules. *Science*, 296:1473–1476, 2002.

[107] M. Mills and I. Andricioaei. An experimentally guided umbrella sampling protocol for biomolecules. *Journal of Chemical Physics*, 129:114101–11413, 2008.

CHAPTER III

Sources of Heterogeneity in the Forced Unfolding Pathway of Streptokinase β

3.1 Introduction

The development of sophisticated instruments for the manipulation of individual molecules such as optical and magnetic tweezers [1, 2, 3, 4, 5, 6], atomic force microscopes [7], and microfluidics [8, 9] have made it possible to measure the elastic and mechanical properties of a variety of systems. These instruments allow for the determination of molecular extension on the nm scale as a function of applied forces at sub-picoNewton levels [10, 11, 12]. Single molecule force pulling experiments have been used to great advantage to reveal details about the mechanical properties of both load-bearing proteins [13, 14, 15, 16, 17, 18, 19, 20, 21] and non-load-bearing proteins [22, 23, 24, 25]. Theoretical methods developed to analyze the force-extension curve produced in these experiments mean that such experiments can be used to determine both energetic [26, 27, 29, 30, 31, 32, 33] and kinetic properties [35, 36, 37] of the forced unfolding pathways of the systems studied. Furthermore, because these experiments pull apart protein domains one at a time, they allow for detection of heterogeneity in unfolding pathways.

At the same time, simulation techniques have been developed to complement these experiments and provide further details of structural unfolding [38, 39]. Most notably,

steered molecular dynamics [40, 41, 42, 43, 44, 45], in which one end of a molecule is held fixed while the other is pulled at either constant force or constant velocity, directly mimics force pulling experiments and provides similar force-extension data, although much higher forces and velocities are necessary to unfold proteins within computationally accessible time-scales.

Although a number of systems have now been studied extensively using force pulling experiments and simulations, there are still many potentially mechanically stable proteins for which there is little data. Recently, Cieplek et al. used unfolding simulations of Go-like model proteins, in which the residues are represented as beads at the C_α position and interactions are modeled such that only native contacts are favorable, to identify protein domains with high mechanical stability [46]. The potential stability of 30,000 structures obtained from the PDB was tested in these Go-model simulations by tracking the number of broken native contacts as the protein was stretched at constant velocity. Of the structures studied, 134 were identified as having high mechanical strength.

3.1.1 Streptokinase β

The protein domain predicted by these simulations to be the most mechanically stable, with a force peak nearly twice that of the I27 domain of the muscle protein titin, was the β domain of Streptokinase. Streptokinase is a promoter enzyme involved in the fibrin degradation pathway [47, 48]. Its structure consists of three homologous domains, α, β, and γ, which cooperatively bind plasminogen [49, 50]. Due to flexibility of the β domain, the structure has only been determined for a 137 residue segment, shown in figure 3.1 [51]. This 137-residue sequence has a β grasp structure with 5 β strands flanking a single α helix. The domain also has two flexible loop regions. The load bearing region is between the C- and N-terminal parallel β

strands 1 and 7 which are connected by a network of 10 hydrogen bonds. The core of the domain is largely nonpolar, with the interaction between the helix and β strands primarily due to hydrophobic contacts.

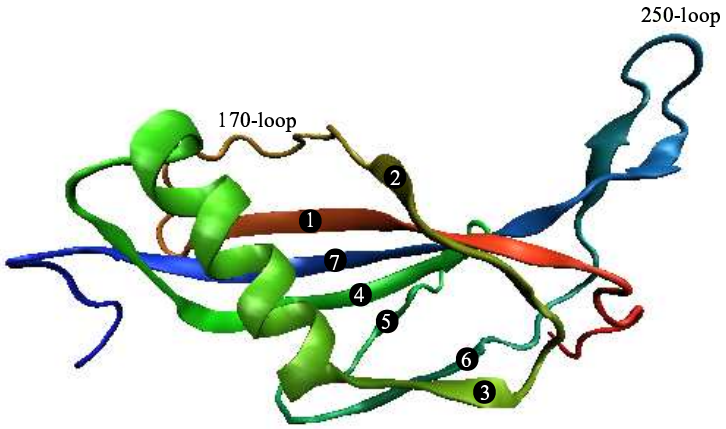

Figure 3.1: Structure of streptokinase β.

Here, we present evidence from AFM force pulling experiments and Steered Molecular Dynamics simulations that the β domain of streptokinase unfolds via three distinct transition states. The experimental results show that the distribution of unfolding pathways is velocity dependent, with one pathway becoming dominant at high velocity. Since SMD simulations are carried out at much higher velocities than used in experiment, the chance of seeing more than one unfolding pathway in straightforward SMD simulations is very low. To get around this problem, simulations were run at high temperatures, in order to soften the protein and lower the free energy barriers, increasing the chance that the domain would explore the full unfolding free energy surface. The relationship between temperature and forced unfolding has been assessed experimentally, and it has been shown that increasing temperature results in lower unfolding forces [52, 53, 54].

3.2 Methods

3.2.1 AFM Pulling Experiments

AFM force curves were kindly provided by Dr. Dora Guzman and Dr. Zhibin Guan. Force measurements were carried out at room temperature with a Multimode-Nanoscope IV atomic force microscope (Veeco, Santa Barbara, CA). Silicon nitride cantilevers (Olympus, TR400PB) with an average spring constant of 0.028 N/m and resonant frequency between 10.796 and 10.970 kHz were used. Gold coated silicon wafers were used as the substrate for AFM experiments. A 50 μL sample was incubated on the substrate for one hour and then gently rinsed with fresh PBS ($2x25$ μL). Finally, a droplet of 25 μL PBS was deposited on the sample prior to beginning the measurements. Additionally, 25 μL of PBS was deposited on the cantilever tip, avoiding formation of air bubbles, to reduce surface tension upon contact with the protein sample. Force measurements were conducted at pulling speeds of 0.2, 0.5, 1.0, 2.0, and 4.0 nm/s. The data sets were acquired in triplicate using multiple cantilevers to reduce systematic errors of the measured forces attributed to spring constant calibration.

3.2.2 Steered Molecular Dynamics Simulations

Steered molecular dynamics (SMD) simulations were run in NAMD [55] using the CHARMM 22 parameter set with CMAP correction [56, 57]. The system was solvated in a TIP3P water bubble of radius 60 Å. The system was minimized in CHARMM for 500 steps of steepest descent minimization with the protein held fixed, followed by 1000 steps of adopted basis Newton-Raphson minimization with the protein harmonically restrained, and 1000 steps ABNR minimization with no constraints. The system was then equilibrated for 50ps. SMD simulations were

run at two temperatures. As a control, 20 simulations were run at 300K, and 100 simulations were run at 350K. The simulations were run using Berendsen pressure controls to keep the pressure constant and a flexible cell to allow for the deformation of the water bubble as the protein was pulled. The C-terminus of the protein was held fixed, while the N-terminus was pulled at a constant velocity of 0.1 Å/ps using a harmonic constraint with a force constant of 5.95 kcal/mol/Å^2. The simulations were run for 4.5 ns, the approximate time necessary for the protein to reach full extension.

3.3 Results

3.3.1 Analysis of AFM force curves

Comparison of coarse-grained [46, 58] and all-atom [42] steered molecular dynamics simulations for the β domain of Streptokinase with the I27 domain of titin predicted that SKβ was roughly twice as strong. However, AFM pulling experiments yielded results similar to those for titin I27. One possible explanation for this is that SKβ may unfold via different pathways under force, and because the SMD simulations were run at much higher velocities and forces than the experiments, they represent an alternate pathway.

Close examination of the AFM data supports the hypothesis that the streptokinase β domain unfolds via multiple pathways. When the first force extension curves from each set are aligned using just the initial slack region, three distinct types of curve are observed (Fig. 3.2). These curve types are consistent across experiments at different pulling velocities. Furthermore, the different curve types are present within the same sets of experiments, which indicates this is an intrinsic feature of the domain rather than an experimentally induced difference. Figure 3.3 shows an example of a single AFM experiment. The relative curvatures of the force peaks are

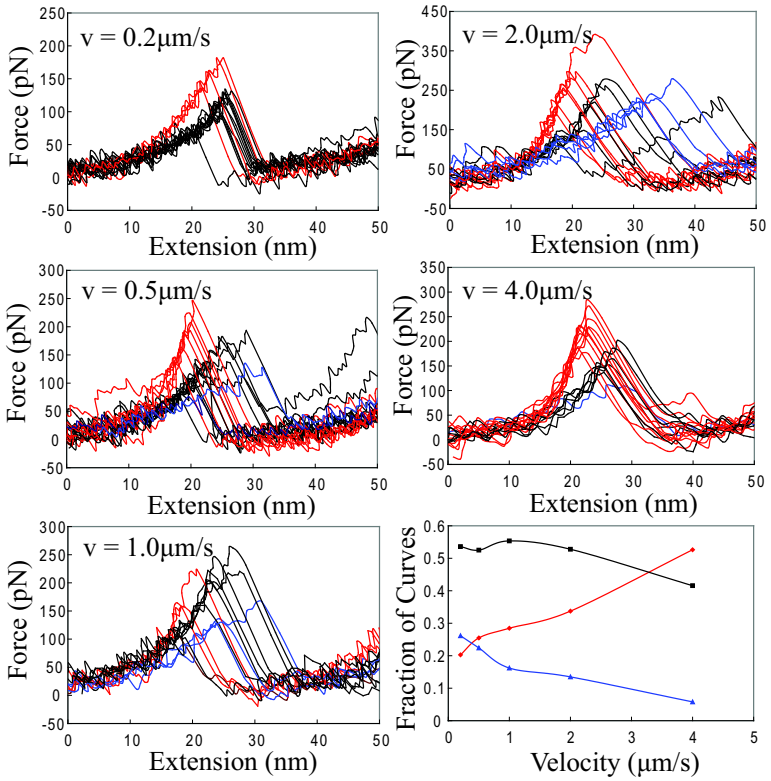

Figure 3.2: Aligned force extension curves for SKβ at various pulling rates. Three distinct types of curves are visible when the curves are aligned. Type 1 is highlighted in red, type 2 is black, and type 3 blue. The final panel shows the fraction pf each curve type as a function of pulling velocity. Alignments are for the first curve out of each set of FECs, the distribution of curve types however, is calculated from all force curves.

highlighted with red lines The three types of force curve exhibit different degrees of sharpness and average peak values. Interestingly, the relative distribution of these curves appears to change as a function of velocity (Figure 3.2, last panel, Table 3.1). At low velocities, the second type of curve is clearly dominant. However, as the pulling rate increases, the percentage of the first type of curve increases, while the second type decreases, until at 4.0 μm/s the first curve type appears in higher proportions than the second. The third curve type also decreases as a function of velocity, with the fraction of type 3 curves approaching zero at 4.0 μm/s. The first curve type has the steepest rise and highest average peak, and as this curve becomes dominant at high pulling rates, we theorize that this curve type represents the path-

way seen in previous SMD simulations. Pulling rate has been shown previously to have an effect on the unfolding pathways of FnIII in Monte Carlo simulations[59] and an actin-binding domain in Go-model simulations [60]. Experimentally, it has been shown that the unfolding pathway differs for low and high pulling forces for ubiquitin [61].

Table 3.1: Fraction of force curve types out of total unfolding events for AFM pulling experiments on SKβ

Velocity	Force Curve Type 1	Force Curve Type 2	Force Curve Type 3
0.2 μm/s	0.2026	0.5359	0.2614
0.5 μm/s	0.2548	0.5251	0.2239
1.0 μm/s	0.2846	0.5534	0.1621
2.0 μm/s	0.5263	0.4158	0.1349
4.0 μm/s	0.5263	0.4158	0.0579

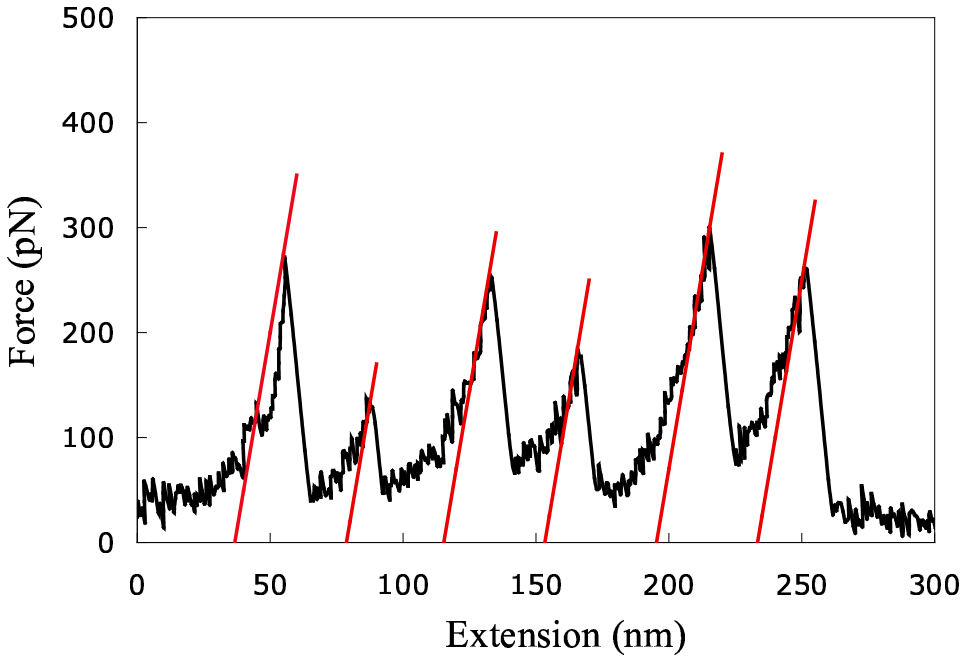

Figure 3.3: Example of a series of FECs from a single pulling experiment. The red lines highlight the different curvatures of individual unfolding events.

The potential of mean force for a nonequilibrium distribution can be calculated from the work put into the system using the Jarzynski equality [33], where

$$e^{-\beta G} = \langle e^{-\beta W} \rangle \tag{3.1}$$

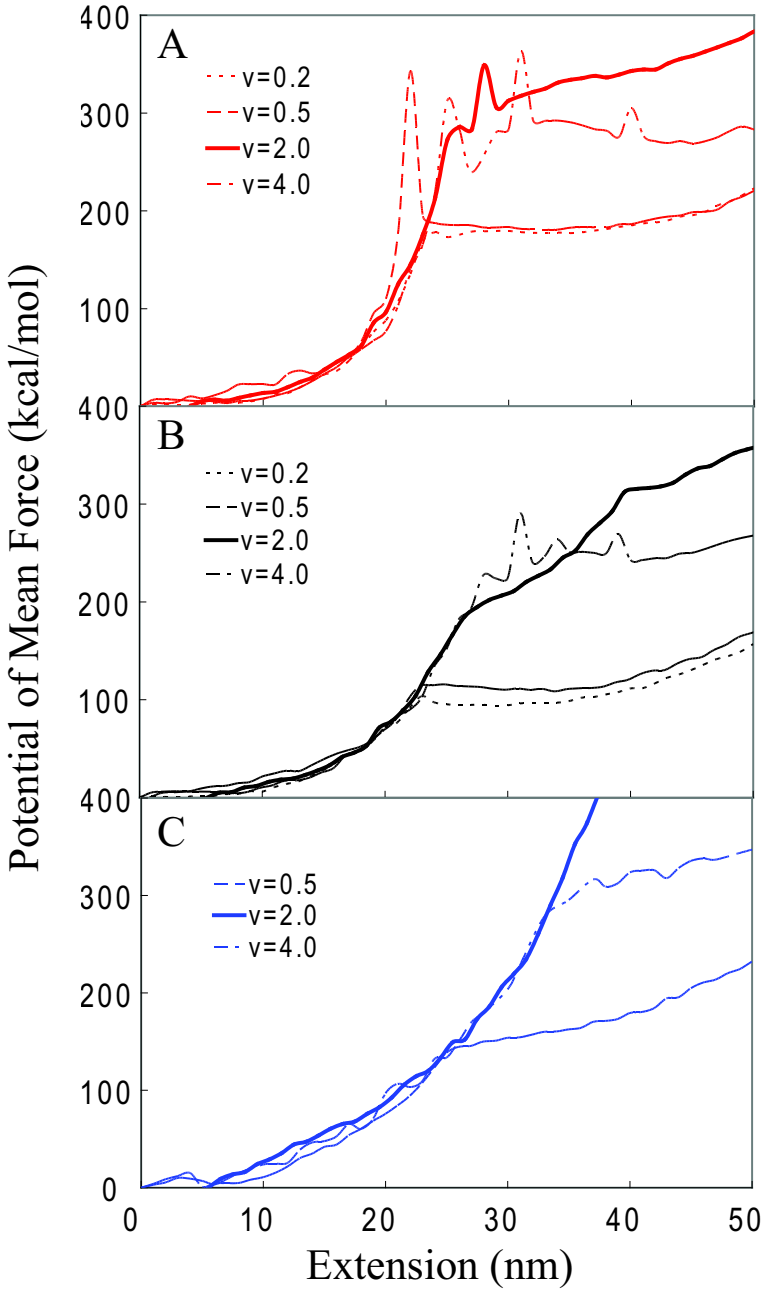

Figure 3.4: Approximate potentials of mean force calculated for the different curve types calculated from force extension curves at several pulling rates using the Jarzynski equation.

For both AFM force curves and SMD simulations, the work can be calculated from the force and extension[32, 31, 62, 27]. Unfortunately, we do not have a large enough ensemble of curves to calculate accurate PMFs, since for most accurate results it is necessary to use only the first unfolding curve and there were only 12-20 usable

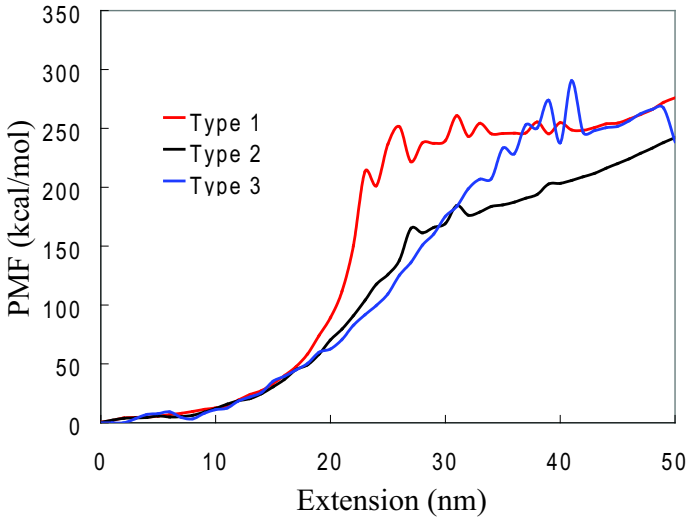

Figure 3.5: PMFs for the different FEC types averaged over all pulling rates.

curves for each velocity. Furthermore, because of the variance of curve distributions with velocity, the overall PMF would differ for the various pulling rates.

However, we can still calculate approximate PMFs for each curve type for comparison using the Jarzynski equation. Figure 3.4 shows these values for the different curve types overlaid for the different experimental velocities. The averages for the PMFs of the different curve types over all velocities are shown in figure 3.5. The height of the barrier is highly variable, due to poor sampling, but the overlap between the different velocities is very good, confirming that the distinct curves we see in the force extension plots are the same for the different pulling rates. The averaged curves (Figure 3.5) show that the PMFs are nearly identical for the first 17nm before the first curve type diverges from the second and third, and the second and third curves diverge at approximately 19nm (These values are relative, since the amount of slack varies. The alignment is based on the lowest value observed in the curves used for the PMF). We take this overlap as evidence that the alignments used in distinguishing the curves are reasonable.

3.3.2 High Temperature Steered Molecular Dynamics Simulations

In order to investigate the source of the heterogeneity seen in the AFM data, we ran a series of regular and high-temperature SMD simulations. An ensemble of 100 simulations was run at 350K, and a control group of 20 simulations was run at 300K. Running SMD simulations at high temperature has the effect of flattening the free energy landscape, allowing the protein to explore pathways more readily than at lower temperatures. Because the CHARMM parameters were developed for physiological temperatures, temperature in an MD simulation does not perfectly represent experimental temperatures. Proteins do not melt in MD simulations until temperatures on the order of 450-600K are reached [63, 64, 65, 66]. Consequently, we do not expect the temperatures used in these simulations to cause thermal denaturation of the proteins.

As expected from the results of the AFM analysis, three distinct unfolding pathways were observed for SKβ in our high-temperature SMD simulations. The force extension curves for these three pathways are shown in Figure 3.6. As in the AFM data, the three types have different slopes and average heights. The average force peak is 2445.6 pN for type 1, 2298.3 pN for type 2, and 2178.7 pN for type 3 and in all types occurs between 55 and 60 Å. The relatively small difference in the height of the peaks mirrors the experimental data. Of the 100 350K simulations, 61 followed the first pathway, 22 followed the second pathway, and 17 followed the third pathway. All 20 of the 300K simulations followed the first pathway as predicted.

Surprisingly, the origin of the difference in unfolding forces is not a difference in the sequence of hydrogen bonds broken. In all of our simulations, the first set of hydrogen bonds to break is the same; bonds between the first and seventh β strands, the load-bearing region of the protein. This sequence of events is illustrated

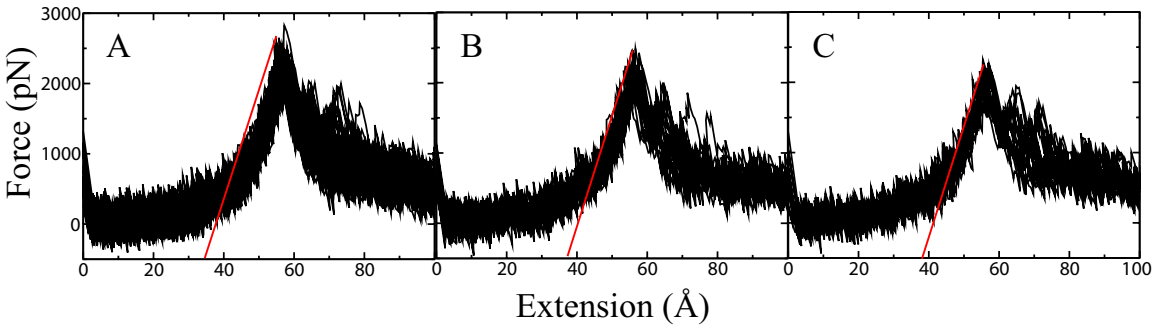

Figure 3.6: Force extension curves for the three unfolding pathways seen in SMD simulations, corresponding to the three types of curve seen in the AFM experiments. The difference in slope is highlighted with a red line.

in Figure 3.7 A. This process involves not only breaking of hydrogen bonds, but also formation of new bonds as the strand 1 is threaded between strands 2 and 7. This shearing effect may in part explain the large difference in relative peak height seen in simulations versus experiments. Despite the increased temperature, our simulations still resulted in much higher force curves than expected. At experimental velocities, interactions with water molecules and reorientation of the domain, both of which are are likely to be affected by high velocities, may allow these strands to separate completely rather than sliding along each other, resulting in lower unfolding forces.

Hydrophobic Interactions Responsible for Heterogeneity in FECs The differences in unfolding pathway are in fact due to hydrophobic interactions in the core of the domain. The residues involved are shown in Figure 3.7 B. In the top structure, the residues that distinguish the second unfolding pathway from the first, which involves interactions between the α helix and β strand 1 are highlighted. The specific residues responsible for this interaction are LEU 199, ALA 203, LEU 207, TYR 162, VAL 164, and PHE 166. The bottom structure shows the residues involved in the third unfolding pathway, LEU 199, LEU 235, PHE 241, and ILE 270. These interactions are responsible for the association of the α helix with the N-terminal

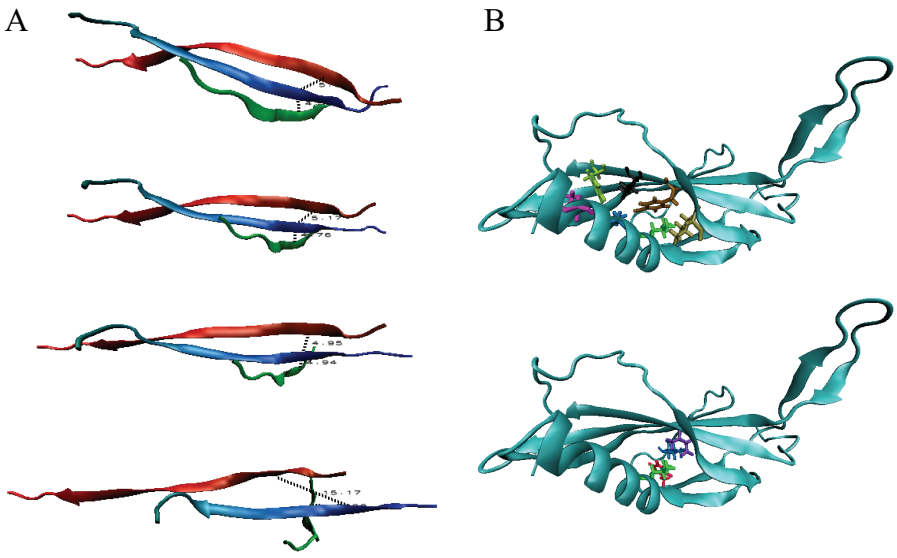

Figure 3.7: A. Time evolution of interaction between β strands 1, 2, and 7. Strand 1 is threaded through strands 2 and 7. B. Regions of hydrophobic interaction responsible for differences in forces required to unfold protein.

side of the β grasp structure.

The global differences in behavior for all three pathways are shown in Figure 3.8. The most obvious difference in the pathways is the position of the α helix with respect to β strand 1 on the C-terminal side of the protein and strands 4, 5 and 6 on the N-terminal side of the protein. In the first pathway, the position of the helix with respect to the rest of the domain is relatively stable. In the second pathway, the helix shifts toward a position roughly perpendicular to the pulling force, and in the third pathway, the helix moves in the opposite direction, approaching an orientation near the direction of the pulling force.

Interactions in the first hydrophobic region, residues 199, 203, 207, 162, 164, and 166, lead to the differences in the FECs for pathways 1 and 2. The red structure (Figure 3.8, 3.9) is representative of the first pathway 1, in which β strand 1 "slides" along the helix and stays associated with it long after the first major unfolding

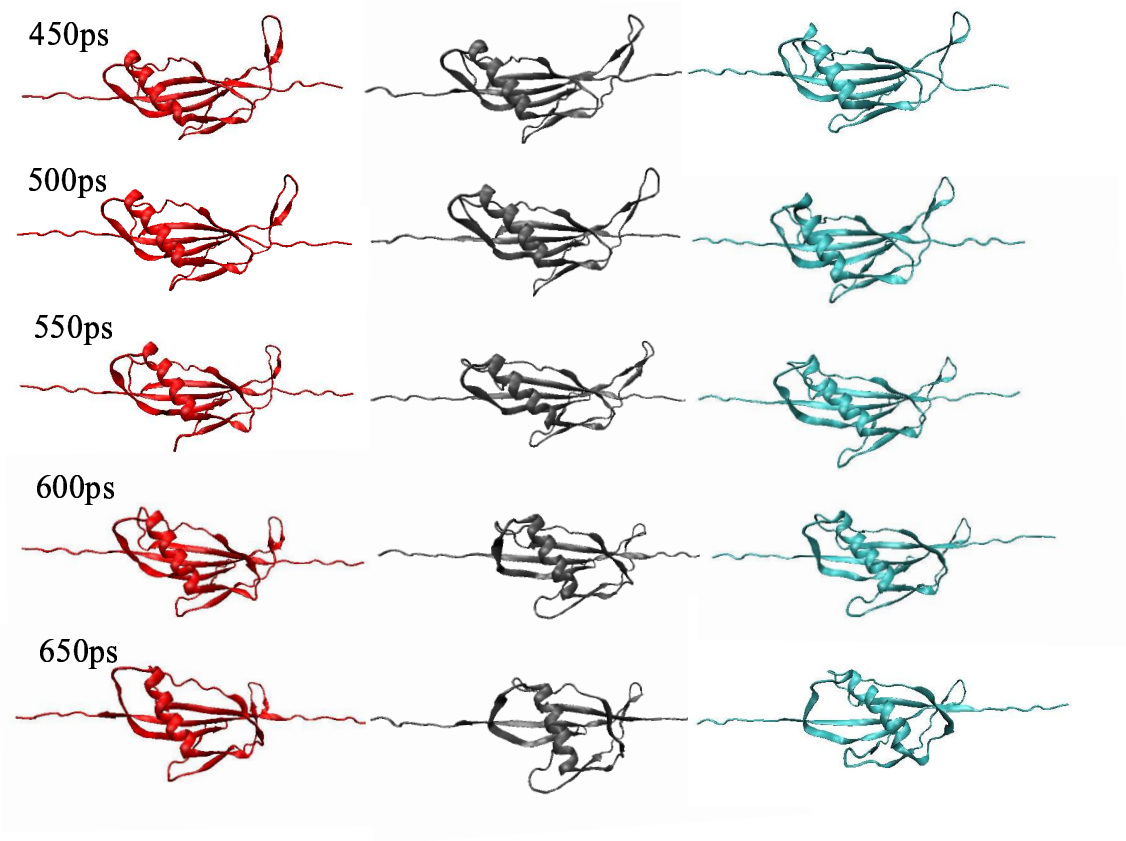

Figure 3.8: Global differences in the unfolding pathway for type 1 (red), type 2 (gray), and type 3 (cyan).

event, which typically happens between 600 and 650ps. This "sliding" involves the dissociation of the original contacts, between 199 and 162, 203 and 164, and 207 and 166, followed by the formation of new contacts as the protein is stretched. As β strand 1 is pulled away from the core of the protein, the contact between residues 199 and 162 is replaced by interaction between 199 and 164 and finally 199 and 166. In the second unfolding pathway, represented by the gray protein in Figures 3.8,3.9, this "sliding" effect is not present. Rather, β strand 1 pulls away from the helix completely at approximately the same time as the first hydrogen bonds are breaking.

Representative graphs of the distances between the side chains of residue 199 and

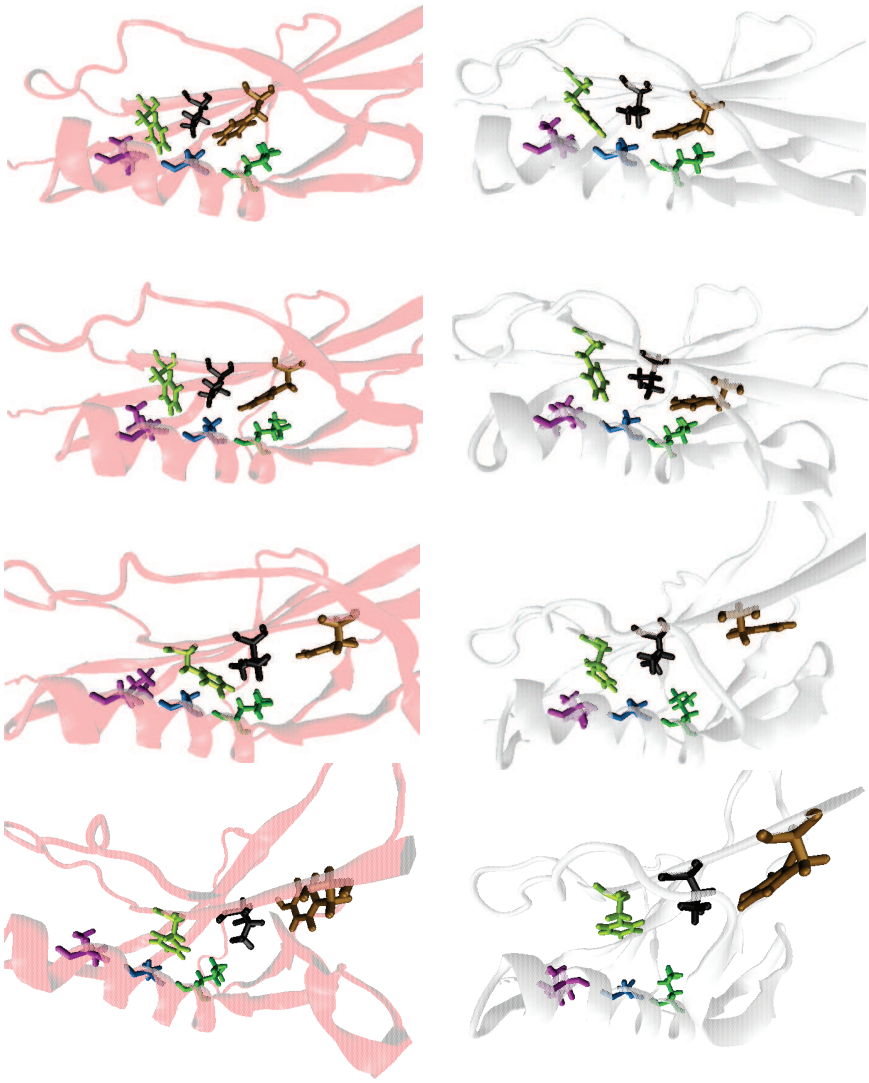

Figure 3.9: Closeup view of residues involved in the hydrophobic interactions which distinguish pathway 1 from pathway 2.

residues 162, 164, and 166 are shown in Figure 3.11 A (pathway 1) and B (pathway 2).

The differences in the second hydrophobic region, residues 199, 235, 241, and 270, are shown in Figure 3.10. In unfolding pathway 1 (red), the helix stays associated with the hydrophobic residues on β strands 4, 5, and 7, while in the third unfolding pathway, represented in cyan, the helix dissociates from this region, again at the

same time as the first hydrogen bonds are breaking. We note that in some cases rather than dissociating completely, the interaction of residues 235, 241, and 270 with LEU199 is replaced by interaction with LEU200. In these cases, the force extension curve resembles that for pathway 1, indicating yet another way that non-native hydrophobic interactions contribute to the unfolding force.

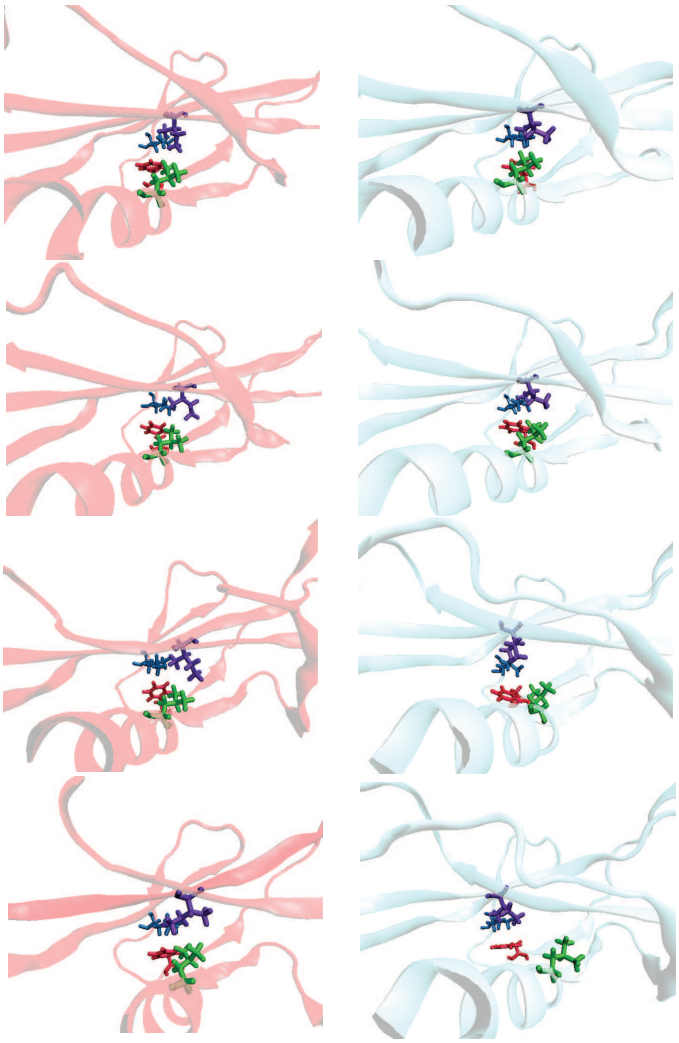

Figure 3.10: Closeup view of residues involved in hydrophobic interactions which distinguish pathway 1 from pathway 3. The key residue, 199, is represented in green

Representative graphs of the distances between the side chains of residue 199 and residues 235, 241, and 270 are shown in Figure 3.11 C (pathway 1) and D (pathway 3). In the few cases where the behaviors of both pathways were seen, the curves were

assigned to type 3, since the third type is associated with the lowest unfolding force.

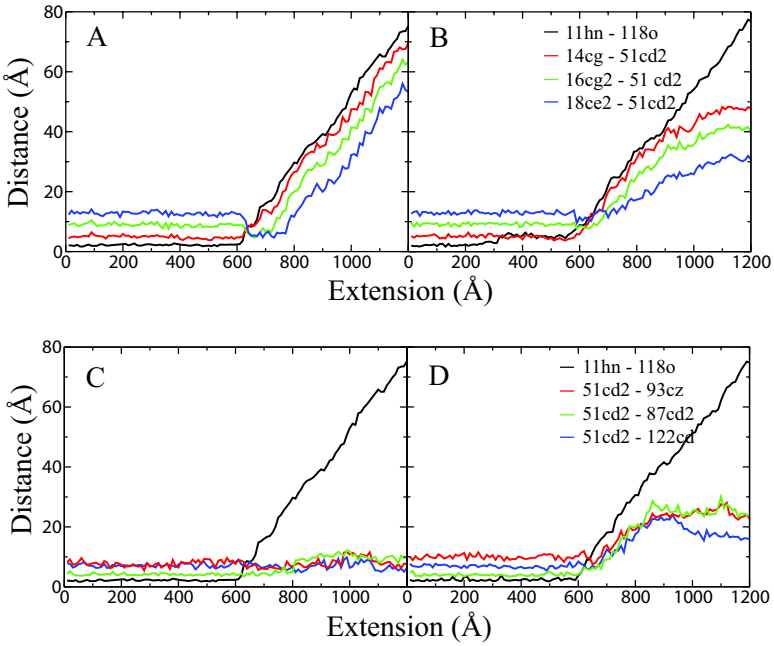

Figure 3.11: Comparison of interactions between hydrophobic regions for the different unfolding pathways. Representative distances between atoms involved in hydrophobic interactions. The distance between hydrogen bond pair, residue 159 atom HN and residue 266 O, representing the first major unfolding event, is shown in black for reference. A. region 1, unfolding pathway type 1. B. Region 1 unfolding pathway type 2. C. Region 2, unfolding pathway type 1. D. Region 2, unfolding pathway type 3.

As the hydrophobic regions described above are not directly involved in interactions in the load bearing regions, it appears that they instead affect the force indirectly by stabilizing the domain's core. We theorize that these hydrophobic interactions represent opposing forces on the α helix. The release of the helix from one side or the other most likely eases the strain on the β strands involved in the hydrogen bonding interactions that are the first to break, resulting in a lower force extension curve.

Potentials of Mean Force from SMD simulations

It is possible to calculate PMFs from SMD simulations in the same was as for experiments. Potentials of mean force from the SMD simulations are shown in figure 3.12. PMFs for the SMD simulations were calculated using the method developed

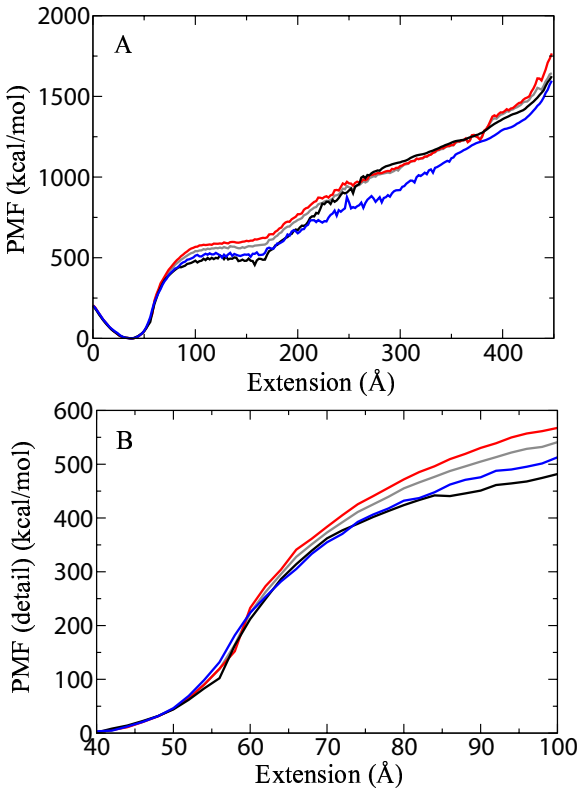

Figure 3.12: A. Potentials of Mean Force calculated from the SMD force curves. The total PMF is shown in gray, PMFs for the distinct unfolding pathways are shown in red (type 1), black (type2), and blue (type3). B. Detail from PMFs showing the region in which the first major folding event occurs.

by Hummer and Szabo [27]. In this method, the PMF is calculated using a method similar to the Weighted Histogram Analysis Method used for unbiasing umbrella sampling simulations. An initial guess of PMF is first calculated using the equation

$$G_0(x) = -\beta ln \frac{\sum_t \frac{\langle \delta(x-x_t) e^{-\beta W_t} \rangle}{\langle e^{-B W_t} \rangle}}{\sum_t \frac{e^{[-\beta U(x,t)]}}{\langle e^{-\beta W_t} \rangle}} \tag{3.2}$$

. The normalizing factor $\langle e^{-\beta W_t} \rangle$ is then replaced with $e^{-\beta[U(x,t)+G_0(x)}dx/e^{-\beta G_0(x)}$. The equations are solved iteratively until self-consistent. This method has been shown to result in more accurate PMFs than more simplistic approximations [28].

The PMF for all 100 simulations is shown in gray, while the PMFs for the individual pathways are shown in red, black, and blue. The number of simulations of each type is once again too few for an accurate calculation of the PMF, and the

temperature issues in MD discussed above, as well as the very large pulling velocity, also contribute to uncertainty in our calculations, so these curves should be taken as qualitative values for comparison, rather than an exact calculation of the free energy differences. It is obvious from the curves that the overall PMF is overly influenced by the most dominant pathway. This is consistent with our attempts to calculate a single PMF for the AFM data (data not shown), which resulted in significantly different curves at each velocity. The PMF calculations for the different pathways do. however, agree qualitatively with our rough calculations of PMFs from the experimental data.

The effect of the distribution of pathways on the overall PMF reveals a more general problem with relying on a single reaction coordinate to define the free energy landscape of a complex system. Although a single coordinate can be used to accurately describe a system which follows a single pathway along the chosen coordinate, for any system in which multiple behaviors are present, more degrees of freedom are necessary to accurately describe the free energy landscape.

The equilibrium calculation for a potential of mean force is

$$G(\xi) = -\frac{1}{\beta} ln \langle P(\xi) \rangle \tag{3.3}$$

where $P(\xi)$ is the probability distribution of the reaction coordinate ξ. Jarzynski's equality replaces this equilibrium distribution for a non-equilibrium system with the exponential of the work done on the system, $e^{-\beta W(\xi)}$. The resulting PMF, it should be noted, is still for the equilibrium state. In essence, the Jarzynski equality is a way to remove the bias of the applied force from the system. It is possible to calculate the unbiased potential of mean force from a biased system along a secondary reaction coordinate, i.e. a coordinate other than the one to which the biasing force is applied, by calculating the probability distribution of the secondary coordinate, ϕ, given the

primary coordinate, ξ (the coordinate which the force is applied to). Since the force is applied in the direction of our primary reaction coordinate, we can rewrite the Jarzynski equality to describe two degrees of freedom, such that

$$G(\phi|\xi) = -\frac{1}{\beta} ln\langle e^{-\beta W(\phi|\xi)}\rangle \tag{3.4}$$

$W(\phi|\xi)$ is calculated by binning values of ϕ for each value of ξ and calculating the work for each pair of values. The resulting work values can then be used in equation 3.2 to produce a two-dimensional PMF. This expansion has been previously shown to provide accurate results for the secondary coordinate in a simple model system [28].

In our case, the most useful secondary reaction coordinate is the solvent accessible surface area (SASA) of the nonpolar residues involved in the alternative unfolding pathways. Since the interaction between hydrophobic groups is really driven by solvent entropy, the surface area of the residues exposed to water is the most relevant coordinate to explore their contribution to the free energy. Figure 3.13 shows the two-dimensional free energy surfaces for two sets of additional coordinates, the SASA of residues 164, 166 and 199, and the SASA of residues 199, 235, 241 and 270. The total PMFs along the full extension of the domain are shown in A and D. The first pathway has a sharp increase in SASA from 65 Å to 130 Å followed by a more gradual increase from approximately 130 Å to 190 Å. The change in SASA of the second region, on the other hand, is more linear overall, with the bulk of the increase happening between roughly 65 Å and 380 Å. The second region also exhibits a broader range of SASA values at all extensions.

Closeup views of the PMFs around the extension corresponding to the force peak (3.13 C,D and E,F) show that there are indeed two distinguishable energetic pathways for each hydrophobic region. For the first region, the free energy surface splits into

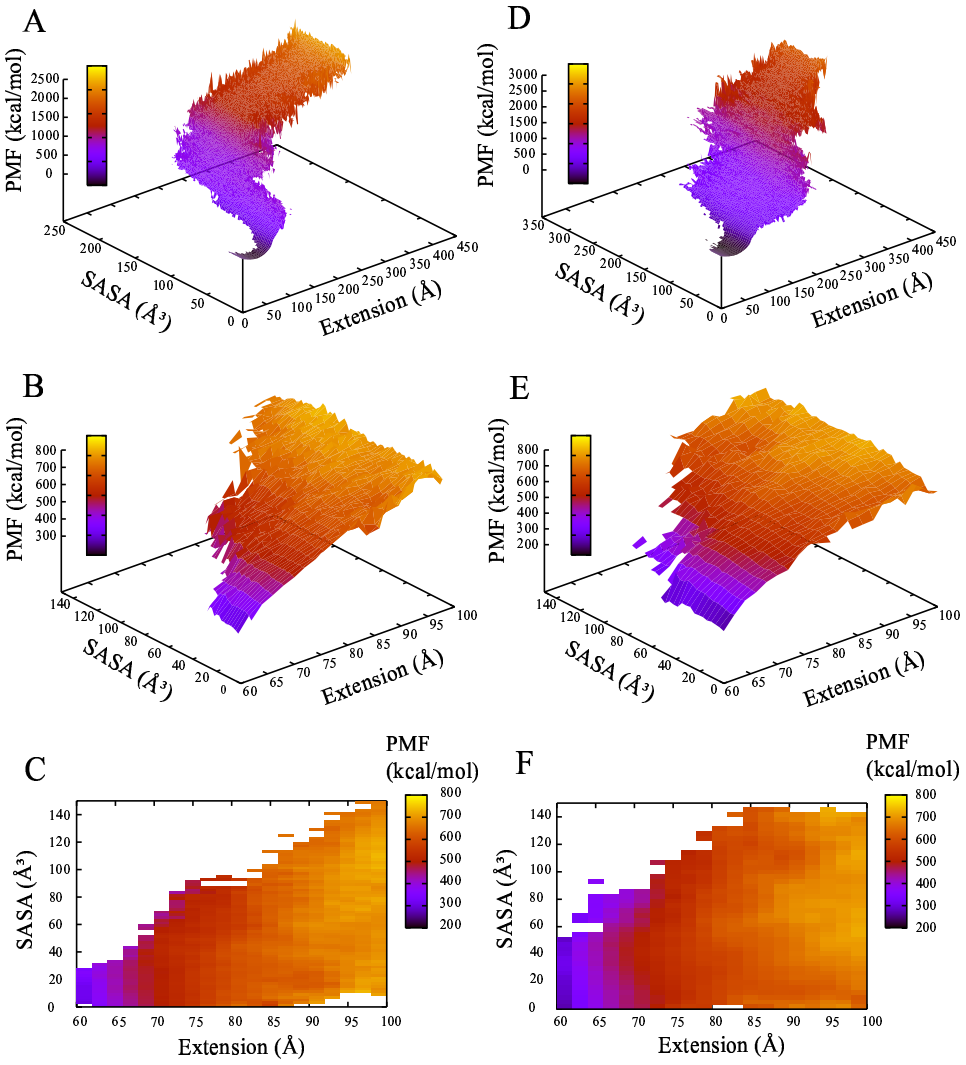

Figure 3.13: Multi dimensional potentials of mean force. A. The PMF as a function of extension and the solvent exposed surface area of residues involved in distinguishing the first and second pathways, 164, 166 and 199. B. and C. Closeup view of the curve for the extension corresponding to the force peak. D. PMF as a function of extension and the SASA of the residues involved in distinguishing the first and third pathways, 199, 235, 241, 270. Closeup views of the region around the force peak are shown in E and F.

two pathways around 60 Å, with type 2 represented by the pathway with higher surface area. The type 2 pathway splits a second time at about 85 Å. One branch returns to a lower SASA, similar to pathway 1, while the SASA of the other continues to increase steadily. This splitting is consistent with behavior observed in a subset of the trajectories that followed this pathway, in which the β strand returns to interaction with the helix after the primary load-bearing h-bonds are broken. The

barrier between the different pathways is on the order of 50-100 kcal/mol. In the second region, the difference between the pathways is more distinct, with the SASA of pathway 3 increasing substantially over that seen in pathway 1 from 60 Å on. The free energy of the third pathway is fairly irregular, possibly due to poor sampling, as this pathway was the most rare of the three. The barrier between pathways 1 and 3 is again on the order of 50-100 kcal/mol.

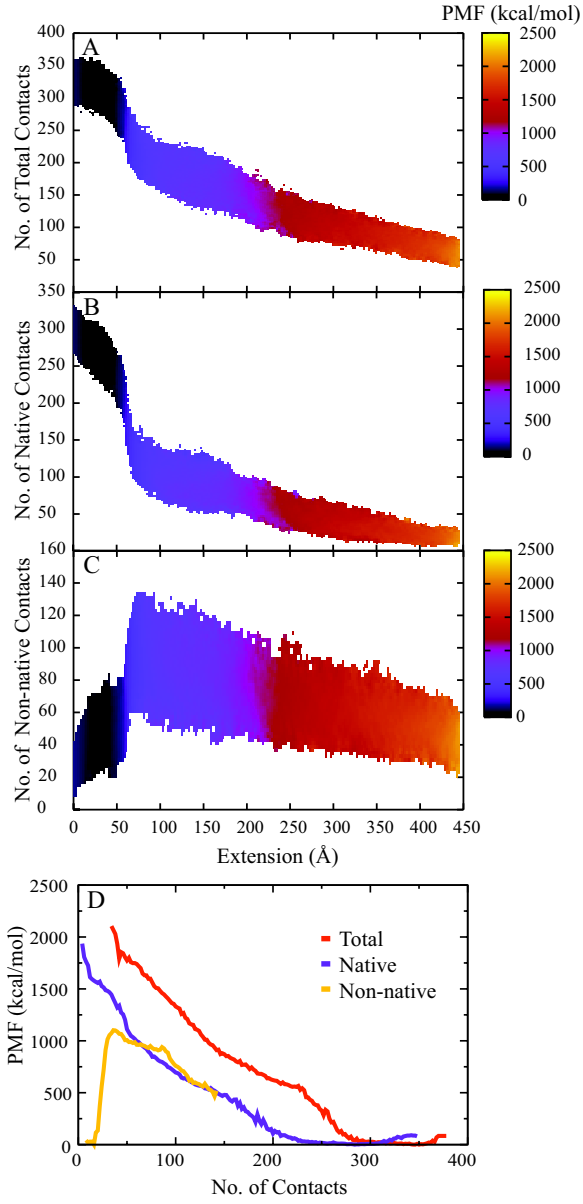

Figure 3.14: Two-dimensional PMFs for total contacts (A), native contacts (B), and non-native contacts (C). Collapsed 1D PMFs are shown in D.

For the case of protein unfolding, it is also interesting to look at the free energy as a function of contacts, both native and non-native. Two dimensional PMFs for total, native, and non-native contacts are shown in figure 3.14. Non-neighboring residues were considered in contact if their side chains were within 6 Å of each other. While the different pathways are not distinguishable in these PMFs, there are interesting global features. There is a sharp decrease in the number of native contacts at approximately 65 Å, after the initial unfolding event. Interestingly, there is a similar sharp *increase* in the number of non-native contacts at this distance. This reflects the shearing behavior of the primary load-bearing strands as well as the sliding hydrophobic interactions of pathway 2, and seems to indicate that, at least at high forces, the stress on the domain caused by the pulling force not only leads to breaking of bonds but also to the formation of alternative contacts as the domain is deformed. One-dimensional PMFs for the number of contacts, calculated by integrating the two-dimensional PMF along the primary coordinate, are shown in Figure 3.14 D. For the total and native contacts, these free energies have a shape similar that of the pulling coordinate, with a slight plateau at the break-point. For non-native contacts the structure resembles a double well more than the cusp-like structure seen in the other PMFs.

3.4 Concluding Discussion

The agreement between the AFM experiments and our simulations implies that the heterogeneity observed in the unfolding of SKβ is not due to artifacts of either method. Both methods predict three unfolding pathways, which result in force extension curves with varying curvature. Both methods predict three unfolding pathways, which result in force extension curves with varying curvature and peak height. Al-

though experiments were able to detect the presence of these different pathways, experiments alone could not easily reveal the sources of such differences. The additional information available in MD simulations enables us to propose a possible explanation for the experimental observations. The simulations further can be used to design future experiments, as they indicate that residue 199 is a prime candidate for mutation experiments.

The contribution of hydrophobic contacts to the shape and height of FECs is surprising, since such contacts are not usually considered to be very strong. Since there is little actual attractive force between hydrophobic residues, this contribution is most likely due to solvent entropy effects, as exposing the side chains of hydrophobic residues to solvent will cause ordering of local water molecules. It has been shown that such entropy contributions may be sufficient to overcome negative enthalpic contributions to protein folding [68, 69, 70, 71, 72]. It would therefore appear reasonable that such effects could contribute to protein unfolding as well. This hydrophobic contribution is especially interesting in light of the fact that many structurally stable protein domains have large number of hydrophobic residues in their core. The structural domain, titin I27 has 41 hydrophobic residues out of 89. Ubiquitin's core is also hydrophobic, with 28 hydrophobic residues out of 76. For most systems, these interactions do not appear to lead to the sort of heterogeneity observed here, although it is known that hydrophobic residues contribute to structural stability. For example, replacing Leucine 60 of titin I27 with alanine reduces the stability and causes an unfolding free energy decrease of 5.3 kcal/mol, although it does not affect the kinetics of forced unfolding [73]. Our results indicate that the importance of these residues should not be overlooked in the study of mechanical proteins.

Continuing advances in computing power have made it possible to generate statistically significant ensembles of complex systems. Such ensembles allow for the ability to not only identify distinct pathways, but to define such pathways energetically. The expansion of the Jarzynski equality described here and in [28] for the calculation of free energy pathways over multiple reaction coordinates from SMD simulations means that the energetic effect of secondary coordinates on the unfolding pathway can be quantified. This expansion can be used to calculate PMFs for any relevant degree of freedom, so long as it can be adequately sampled within the timescales required for the exploration of the primary reaction coordinate. While the application of this extension to computational studies is obvious, there is also potential for the determination of such multi-dimensional PMFs experimentally. Techniques which combine force manipulation and fluorescence have recently been developed [74] which could provide the necessary information on secondary coordinates for the type of calculations described here.

While our SMD simulations have made it possible to determine the possible source of heterogeneity observed in the AFM experiments, the difference in force peaks for the three unfolding pathways is insufficient to explain the two-fold differences observed in forces required for unfolding of titin I27 and SKβ in Go-model and all-atom SMD simulations. This may have to do with the way the hydrogen bonds in the load-bearing region interact. At lower velocities, it is likely that water has sufficient time to enter the region between strands 1 and 7 as they separate. These waters could weaken the bonds between side chains and, once these bonds are broken, sequester the hydrogen-bonding groups of the strands, preventing the formation of new hydrogen bonds between the strands. Such water interactions are known to be important to the unfolding of titin I27 [42]. It appears that even with elevated

temperatures, the velocity used in our simulations is too great to allow water time to interact with the separating strands, leading to the shearing interaction observed in simulations.

Finally, the work presented here demonstrates the need to design simulations carefully, and with experimental input where possible. Without prior knowledge of the velocity dependence of the domain's behaviour, the need to modify the standard SMD procedure would not have been obvious, and the presence of alternative pathways would likely not have been detected.

BIBLIOGRAPHY

[1] D. E. Smith, G. J. Gemmen, R. Millin, J. P. Rickgauer, A. L. Schweitzer, and D. N. Fuller. Using optical tweezers to study protein-DNA interactions. *Proceedings of the SPIE - The International Society for Optical Engineering*, vol.5930, no.1:593012–1–10, 2005.

[2] A. Ashkin, J. M. Dziedzic, and T. Yamane. Optical trapping and manipulation of single cells using infrared-laser beams. *Nature*, 330:769–771, 1987.

[3] K. Svoboda and S. M. Block. Biological applications of optical forces. *Annual Review of Biophysics and Biomolecular Structure*, 23:247–285, 1994.

[4] D. G. Grier. A revolution in optical manipulation. *Nature*, 424:810–816, 2003.

[5] C. Haber and D. Wirtz. Magnetic tweezers for DNA micromanipulation. *Review of Scientific Instruments*, 71:4561–4570, 2000.

[6] C. Gosse and V. Croquette. Magnetic tweezers: Micromanipulation and force measurement at the molecular level. *Biophysical Journal*, 82:3314–3329, 2002.

[7] J. Zlatanova, S. M. Lindsay, and S. H. Leuba. Single molecule force spectroscopy in biology using the atomic force microscope. *Progress in Biophysics & Molecular Biology*, 74:37–61, 2000.

[8] S. K. Sia and G. M. Whitesides. Microfluidic devices fabricated in poly(dimethylsiloxane) for biological studies. *Electrophoresis*, 24:3563–3576, 2003.

[9] C. Q. Yi, C. W. Li, S. L. Ji, and M. S. Yang. Microfluidics technology for manipulation and analysis of biological cells. *Analytica Chimica Acta*, 560:1–23, 2006.

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

[11] Y. E. Pak, M. Marimuthu, and S. Kim. Nanomechanics of biomolecules: A review. *Biochip Journal*, 2:235–241, 2008.

[12] K. C. Neuman and A. Nagy. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nature Methods*, 5:491–505, 2008.

[13] M. Sotomayor and K. Schulten. Single-molecule experiments in vitro and in silico. *Science*, 316:1144–1148, 2007.

[14] R. B. Best, D. J. Brockwell, J. L. Toca-Herrera, A. W. Blake, D. A. Smith, S. E. Radford, and J. Clarke. Force mode atomic force microscopy as a tool for protein folding studies. *Analytica Chimica Acta*, 479:87–105, 2003.

[15] M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez. Mechanical and chemical unfolding of a single protein: A comparison. *Proceedings of The National Academy of Sciences of The United States of America*, 96:3694–3699, 1999.

[16] P. E. Marszalek, H. Lu, H. B. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten, and J. M. Fernandez. Mechanical unfolding intermediates in titin modules. *Nature*, 402:100–103, 1999.

[17] M. S. Z. Kellermayer, S. B. Smith, C. Bustamante, and H. L. Granzier. Complete unfolding of the titin molecule under external force. *Journal of Structural Biology*, 122:197–205, 1998.

[18] W. A. Linke and A. Grutzner. Pulling single molecules of titin by AFM - recent advances and physiological implications. *Pflugers Archiv-European Journal of Physiology*, 456:101–115, 2008.

[19] M. Cieplak and P. E. Marszalek. Mechanical unfolding of ubiquitin molecules. *Journal of Chemical Physics*, 123:194903–194910, 2005.

[20] C. L. Chyan, F. C. Lin, H. B. Peng, J. M. Yuan, C. H. Chang, S. H. Lin, and G. L. Yangy. Reversible mechanical unfolding of single ubiquitin molecules. *Biophysical Journal*, 87:3995–4006, 2004.

[21] M. Carrion-Vazquez, H. B. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez. The mechanical stability of ubiquitin is linkage dependent. *Nature Structural Biology*, 10:738–743, 2003.

[22] Y. Cao, C. Lam, M. J. Wang, and H. B. Li. Nonmechanical protein can have significant mechanical stability. *Angewandte Chemie-International Edition*, 45:642–645, 2006.

[23] D. J. Brockwell, G. S. Beddard, E. Paci, D. K. West, P. D. Olmsted, D. A. Smith, and S. E. Radford. Mechanically unfolding the small, topologically simple protein l. *Biophysical Journal*, 89:506–519, 2005.

[24] D. J. Brockwell, E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham, and S. E. Radford. Pulling geometry defines the mechanical resistance of a beta-sheet protein. *Nature Structural Biology*, 10:731–737, 2003.

[25] R. B. Best, B. Li, A. Steward, V. Daggett, and J. Clarke. Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation. *Biophysical Journal*, 81:2344–2356, 2001.

[26] C. Hyeon and D. Thirumalai. Measuring the energy landscape roughness and the transition state location of biomolecules using single molecule mechanical unfolding experiments. *Journal of Physics-Condensed Matter*, 19:113101–113127, 2007.

[27] G. Hummer and A. Szabo. Free energy surfaces from single-molecule force spectroscopy. *Accounts of Chemical Research*, 38:504–513, 2005.

[28] D. D. L. Minh. Multidimensional potentials of mean force from biased experiments along a single coordinate. *Journal of Physical Chemistry B*, 111:4137–4140, 2007.

[29] S. Kirmizialtin, L. Huang, and D. E. Makarov. Topography of the free-energy landscape probed via mechanical unfolding of proteins. *Journal of Chemical Physics*, 122:234915, 2005.

[30] G. Hummer and A. Szabo. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proceedings of The National Academy of Sciences of The United States of America*, 98:3658–3661, 2001.

[31] R. W. Friddle. Experimental free energy surface reconstruction from single-molecule force spectroscopy using Jarzynski's equality - comment. *Physical Review Letters*, 100:019801, 2008.

[32] N. C. Harris, Y. Song, and C. H. Kiang. Experimental free energy surface reconstruction from single-molecule force spectroscopy using Jarzynski's equality - reply. *Physical Review Letters*, 100:019802,2008.

[33] C. Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78:2690–2693, 1997.

[34] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophysical Journal*,75:662–71, 1998.

[35] O. K. Dudko, G. Hummer, and A. Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proceedings of The National Academy of Sciences of The United States of America*, 105:15755–15760, 2008.

[36] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. Reversible unfolding of individual titin immunoglobulin domains by afm. *Science*, 276:1109–1112, 1997.

[37] M. Rief, J. M. Fernandez, and H. E. Gaub. Elastically coupled two-level systems as a model for biopolymer extensibility. *Physical Review Letters*, 81:4764–4767, 1998.

[38] D. K. West, D. J. Brockwell, P. D. Olmsted, S. E. Radford, and E. Paci. Mechanical resistance of proteins explained using simple molecular models. *Biophysical Journal*, 90:287–297, 2006.

[39] P. C. Li and D. E. Makarov. Theoretical studies of the mechanical unfolding of the muscle protein titin: Bridging the time-scale gap between simulation and experiment. *Journal of Chemical Physics*, 119:9260–9268, 2003.

[40] M. Gao, B. Isralewitz, H. Lu, and K. Schulten. Steered molecular dynamics simulation of the unfolding of multiple titin immunoglobulin domains. *Biophysical Journal*, 78:156, 2000.

[41] B. Isralewitz, M. Gao, H. Lu, and K. Schulten. Forced unfolding or mutant titin immunoglobulin domains with steered molecular dynamics. *Biophysical Journal*, 78:157, 2000.

[42] H. Lu and K. Schulten. The key event in force-induced unfolding of titin's immunoglobulin domains. *Biophysical Journal*, 79:51–65, 2000.

[43] H. Lu, A. Krammer, B. Isralewitz, V. Vogel, and K. Schulten. Computer modeling of force-induced titin domain unfolding. *Elastic Filaments of The Cell*, 481:143–162, 2000.

[44] H. Lu and K. . Schulten. Steered molecular dynamics simulation of conformational changes of immunoglobulin domain I27 interpret atomic force microscopy observations. *Chemical Physics*, 247:141–53, 1999.

[45] K. Schulten. Manipulating proteins by steered molecular dynamics. *Journal of Molecular Graphics & Modelling*, 16:289–289, 1998.

[46] J. I. Sulkowska and M. Cieplak. Mechanical stretching of proteins - a theoretical survey of the protein data bank. *Journal of Physics-Condensed Matter*, 19:283201, 2007.

[47] L. V. Medved, D. A. Solovjov, and K. C. Ingham. Domain structure, stability and interactions in streptokinase. *European Journal of Biochemistry*, 239:333–339, 1996.

[48] P. Rodriguez, P. Fuentes, M. Barro, J. G. Alvarez, E. Munoz, D. Collen, and H. R. Lijnen. Structural domains of streptokinase involved in the interaction with plasminogen. *European Journal of Biochemistry*, 229:83–90, 1995.

[49] A. C. Tharp, M. Laha, P. Panizzi, M. W. Thompson, P. Fuentes-Prior, and P. E. Bock. Plasminogen substrate recognition by the streptokinase-plasminogen catalytic complex is facilitated by arg(253), lys(256), and lys(257) in the streptokinase beta-domain and kringle 5 of the substrate. *Journal of Biological Chemistry*, 284:19511–19521, 2009.

[50] R. R. Bean, I. M. Verhamme, and P. E. Bock. Role of the streptokinase alpha-domain in the interactions of streptokinase with plasminogen and plasmin. *Journal of Biological Chemistry*, 280:7504–7510, 2005.

[51] X. Q. Wang, J. Tang, B. Hunter, and X. J. C. Zhang. Crystal structure of streptokinase beta-domain. *Febs Letters*, 459:85–89, 1999.

[52] M. Schlierf and M. Rief. Temperature softening of a protein in single-molecule experiments. *Journal of Molecular Biology*, 354:497–503, 2005.

[53] E. Botello, N. C. Harris, J. Sargent, W. H. Chen, K. J. Lin, and C. H. Kiang. Temperature and chemical denaturant dependence of forced unfolding of titin 127. *Journal of Physical Chemistry B*, 113:10845–10848, 2009.

[54] Y. X. Mejia, H. B. Mao, N. R. Forde, and C. Bustamante. Thermal probing of E-coli RNA polymerase off-pathway mechanisms. *Journal of Molecular Biology*, 382:628–637, 2008.

[55] 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.

[56] 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.

[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. Cieplak, A. Pastore, and T. X. Hoang. Mechanical properties of the domains of titin in a Go-like model. *Journal of Chemical Physics*, 122:54906, 2005.

[59] S. Mitternacht, S. Luccioli, A. Torcini, A. Imparato, and A. Irback. Changing the mechanical unfolding pathway of FNIII(10) by tuning the pulling strength. *Biophysical Journal*, 96:429–441, 2009.

[60] M. S. Li and M. Kouza. Dependence of protein mechanical unfolding pathways on pulling speeds. *Journal of Chemical Physics*, 130:145102, 2009.

[61] M. Schlierf, H. B. Li, and J. M. Fernandez. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proceedings of The National Academy of Sciences of The United States of America*, 101:7299–7304, 2004.

[62] S. Park and K. Schulten. Calculating potentials of mean force from steered molecular dynamics simulations. *Journal of Chemical Physics*, 120:5946–5961, 2004.

[63] G. Settanni and A. R. Fersht. High temperature unfolding simulations of the trpz1 peptide. *Biophysical Journal*, 94:4444–4453, 2008.

[64] D. Paschek, S. Gnanakaran, and A. E. Garcia. Simulations of the pressure and temperature unfolding of an alpha-helical peptide. *Proceedings of The National Academy of Sciences of The United States of America*, 102:6765–6770, 2005.

[65] W. Gu, T. T. Wang, J. Zhu, Y. Y. Shi, and H. Y. Liu. Molecular dynamics simulation of the unfolding of the human prion protein domain under low pH and high temperature conditions. *Biophysical Chemistry*, 104:79–94, 2003.

[66] V. Daggett and M. Levitt. Molecular-dynamics simulations of helix denaturation. *Journal of Molecular Biology*, 223:1121–1138, 1992.

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

[68] M. C. Stumpe and H. Grubmuller. Urea impedes the hydrophobic collapse of partially unfolded proteins. *Biophysical Journal*, 96:3744–3752, 2009.

[69] L. J. Lapidus, S. H. Yao, K. S. McGarrity, D. E. Hertzog, E. Tubman, and O. Bakajin. Protein hydrophobic collapse and early folding steps observed in a microfluidic mixer. *Biophysical Journal*, 93:218–224, 2007.

[70] R. H. Zhou, X. H. Huang, C. J. Margulis, and B. J. Berne. Hydrophobic collapse in multidomain protein folding. *Science*, 305:1605–1609, 2004.

[71] V. R. Agashe, M. C. R. Shastry, and J. B. Udgaonkar. Initial hydrophobic collapse in the folding of barstar. *Nature*, 377:754–757, 1995.

[72] C. Simmerling and R. Elber. Hydrophobic collapse in a cyclic hexapeptide - computer-simulations of CHDLFC and CAAAAC in water. *Journal of The American Chemical Society*, 116:2534–2547, 1994.

[73] P. M. Williams, S. B. Fowler, R. B. Best, J. L. Toca-Herrera, K. A. Scott, A. Steward, and J. Clarke. Hidden complexity in the mechanical properties of titin. *Nature*, 422:446–449, 2003.

[74] S. Hohng, R. B. Zhou, M. K. Nahas, J. Yu, K. Schulten, D. M. J. Lilley, and T. J. Ha. Fluorescence-force spectroscopy maps two-dimensional reaction landscape of the Holliday junction. *Science*, 318:279–283, 2007.

CHAPTER IV

Chemical, Structural, and Energetic Analysis of Polyamidoamine Dendrimer-DNA interactions

4.1 Introduction

Polyamidoamine (PAMAM) dendrimers [1] are functionalized nanoparticles with promise in several biomedical applications involving the targeted delivery of drugs and genetic material into the cell. Due to their chemical similarity to DNA-binding proteins, they are also instructive model systems for the hierarchical organization of genes [2]. PAMAM dendrimers consist of a diamine core with four branched units (Figure 4.1). Their size can be easily controlled by adding successive generations of branches and their surface chemistry can be manipulated by conjugating different functional groups to the terminations, [3], allowing the dendrimers to be targeted to specific cell types [4]. Because the amine terminations of the dendrimers are protonated at physiological pH, their positive charge allows them to effectively bind negatively-charged DNA. Dendrimers have also been shown to be effective vectors for transfection [5], i.e., for the delivery inside the cell of therapeutic genes, antisense oligonucleotides or ribozymes. Furthermore, dendrimers have been shown to not only bind DNA [6, 7, 8, 9], but to condense it [10, 11, 12, 13, 14], and to serve as effective transfection vectors [5, 15]. As such, a detailed quantitative analysis of the physical interactions between dendrimers and nucleic acids is a crucial first step in

understanding the delivery mechanisms.

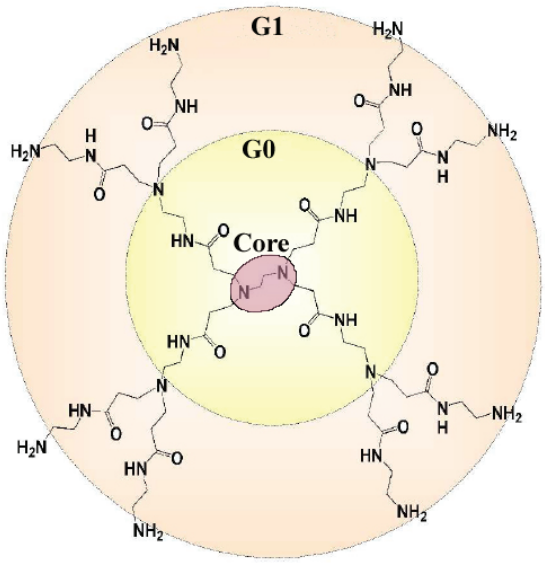

Figure 4.1: Chemical structure of a generation 1 PAMAM dendrimer.

The interaction between dendrimers and double stranded DNA has been studied by a variety of experimental techniques [7, 13, 16, 17]. Relatively fewer theoretical studies have been reported, including a molecular dynamics simulation of dendrimers interacting with single stranded DNA [18] and mathematical modeling of the electrostatics of duplex DNA-dendrimer interactions [19, 9]. Small angle x-ray scattering studies [20] and cryo-TEM analysis [21] indicate that low-generation dendrimers condense DNA by forming toroidal structures. On the other hand, the similarity of dendrimers to histone proteins in structure and charge has lead to speculation that DNA may wrap around higher-generation dendrimers. Dynamic light scattering [6, 8] and single molecule data [14] on the interaction between amine dendrimers and DNA indicate that DNA may be condensed by low generation dendrimers, or may wrap around higher generation dendrimers, but details of the respective compaction mechanisms are absent. Moreover, the interplay between various components such as electrostatic interactions, solvent structure and dynamical changes upon binding

have not been described previously. The effect of binding on the local structure of both DNA and dendrimer is unknown, as is the nature of the chemical interactions. In order to determine the details of the structure of the dendrimer-DNA complex, including the deformation of both the DNA and dendrimer, as well as free energies of complex formation, and the effect of termination type, atomistic molecular dynamics simulations were run on four dendrimer-dsDNA systems.

In order to determine the structural detail of the dendrimer-DNA complex, including deformation of the DNA and the dendrimer, as well as the free energies of complex formation, we set up to run atomistic molecular dynamics simulations and free energy calculations on a dendrimer-dsDNA system. We chose a generation-3 (G3) dendrimer because it is expediently small for intracellular delivery purposes, flexible enough to be interesting (otherwise large dendrimers are almost rigid and bind like "spheres"), and because it is computationally feasible, so that convergence in the calculation of free energy can be achieved with available computational resources.

We find that the interaction between the dendrimer and DNA is non-specific, and dependent on the charge and orientation of the dendrimer. Local deformation in both the DNA and dendrimer was observed in at least one set of interactions. Additionally, we characterize the energetic contributions to complexation, including the effect of ordered waters between the molecules.

4.1.1 Dendrimer-DNA Systems

Interactions were examined for a double-stranded, 24 base-pair segment of DNA and a generation 3 (G3) dendrimer. The terminations and position of the dendrimer were varied in order to probe the effects of dendrimer charge and orientation on the interaction. The dendrimers studied were as follows:

(a) a G3 dendrimer with 32 amine terminations, positioned such that the widest side of the dendrimer was facing the DNA, hereafter referred to as all-amine, orientation 1.

(b) a G3 dendrimer with 32 amine terminations. Because the dendrimer may approach the DNA at different angles, which may affect the overall interaction, a second orientation of the all-amine dendrimer was also studied. The dendrimer was rotated 90^o with respect to the core so that the dimension with the smallest surface area faced the DNA, hereafter referred to as all-amine, orientation 2.

(c) a G3 dendrimer with one side containing 16 amine terminations and the other side containing 16 acetamide terminations, in the same conformation as the all-amine orientation 1, here after referred to as amine-acetamide half-and-half.

(d) a G3 dendrimer with 16 amine and 16 acetamide terminations randomly distributed, also positioned with it's widest side facing the DNA, hereafter referred to as amine-acetamide random.

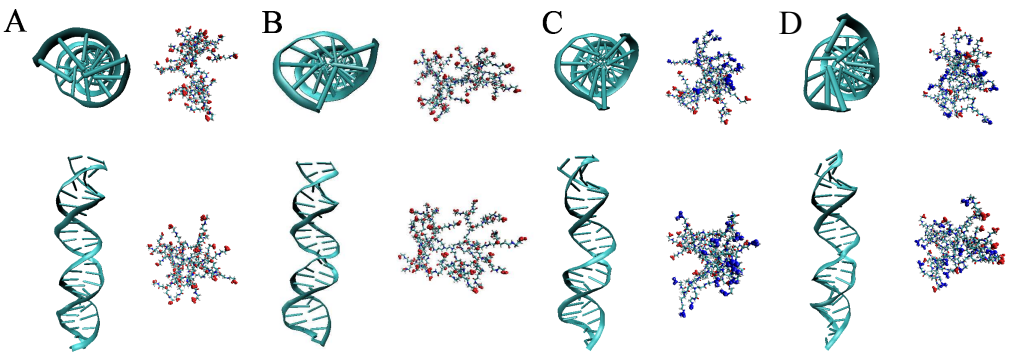

Figure 4.2: Front and top views of the starting structures with an initial center-to-center distance of 50 Å for dendrimers with amine terminations, A. orientation 1, and B. orientation 2, and C. amine terminations on one lobe and acetamide terminations on the other, and D. amine-acetamide mixed terminations. The amine termination groups are represented in blue and acetamide groups are represented in red.

Starting structures are shown in figure 4.2 A-D. While the amine termination has a positive charge at neutral pH, the acetamide has no charge, allowing us to control the overall charge of the dendrimer. Since the dendrimer has a large number of positively charged groups and the DNA has a large number of negatively charged groups, it is expected that the interaction will be driven largely by electrostatics. Thus, the neutral acetamide termination serves as a control for understanding these interactions, as well as the relative contribution of non-electrostatic interactions such as hydrogen bonding and van der Waals interactions.

4.2 Simulation Methods

Three starting structures were generated for each dendrimer-DNA system, one each with the molecules' centers of mass 35 Å, 50 Å, and 70 Å apart, or 45 Å and 60 Å for the 90^o, second orientation of the all-amine dendrimer, corresponding to a minimum distance between the molecules of approximately 10, 25, and 45 Å. The DNA-dendrimer systems were fully solvated using TIP3P water molecules [22] in a box of dimensions 110 × 100 × 100 $Å^3$ and 125 × 100 × 100 $Å^3$, and 145 × 100 × 100 $Å^3$, respectively. 46 sodium ions were added to the system to balance the charge of the DNA and 32 or 16 chlorine ions, depending on the number of amine terminations of the dendrimer, were also added, yielding electrically neutral unit cells. Simulations were run using constant number of particle, volume and temperature conditions and all used periodic boundary conditions. Simulations were run in NAMD using the CHARMM 27 parameter set [23] with a timestep of 2 fs. Hydrogens were held fixed using SHAKE [24]. Electrostatics were calculated using the particle-mesh Ewald method [25]. Nonbonded interactions were truncated with a cutoff of 14 Å and a switching function at 8 Å to smooth the truncation [26]. The

systems were minimized for 1000 steps of steepest-descent minimization with the DNA and dendrimer held fixed and, subsequently, for 4000 steps of the adopted basis set Newton Raphson method with decreasing harmonic restraints on the dendrimer and DNA. The system was then equilibrated for 50 ps with dendrimer and DNA fixed with a harmonic restraint of 0.5 kcal/mol/Å^2 applied to the heavy atoms of the dendrimer and DNA, so that only the solvent and ions were free to move. The system was then equilibrated for another 50 ps with a temperature coupled to a heat bath of 300K under constant volume conditions with no restraints. In all simulations the base pairs at the ends of the DNA segment were restrained with a harmonic potential to prevent fraying; while fraying is physically possible for short stretched of DNA duplexes, this boundary restraint on the end basepairs was deemed appropriate to model the interaction of the dendrimer with stretched of DNA longer than we could include in the atomistic simulation.

For umbrella sampling simulations (see Chapter 2.2 for a detailed description of the method), the reaction coordinate used was the distance between the center of mass of the dendrimer and the center of the DNA, defined as the center of mass of the middle two base pairs. A harmonic potential with a force constant of 2.5 kcal/mol/Å^2 was applied to this reaction coordinate over a series of windows starting from the initial structures and progressing along the reaction coordinate in 1 Å increments. Each window was run for 200 ps of simulation, yielding total simulation times ranging from 11.4 ns to 20.8 ns. Two parallel sets of umbrella sampling simulations were run to increase sampling. In all simulations, a structure from the previous window was used to start each successive window. Coordinates were saved every picosecond.

4.3 Results

4.3.1 Molecular Details of DNA-Dendrimer Complexation

Final complexed structures for the four dendrimer-DNA systems are shown in Figure 4.3

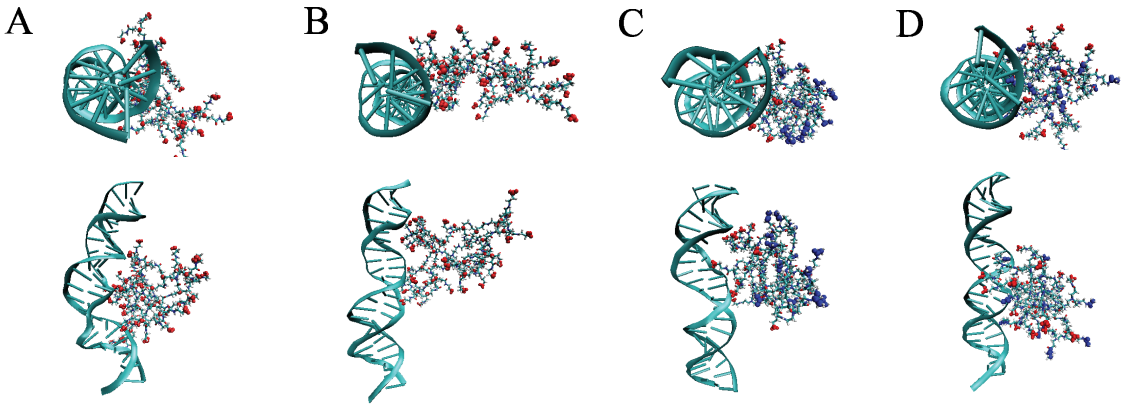

Figure 4.3: Front and top views of the complexed structures for dendrimers with amine terminations, A. orientation 1, and B. orientation 2, and C. amine terminations on one lobe and acetamide terminations on the other, and D. amine-acetamide mixed terminations.

Table 4.1 details the hydrogen bond interactions between the dendrimers and DNA. A hydrogen bond was defined to exist if the donor-acceptor distance was smaller than 2.4 Å. The average lifetime of hydrogen bonds between the DNA and dendrimer was 86.8ps for the all-amine terminated dendrimers and 54.0ps for the mixed termination dendrimers. Most likely this difference is because the multiple contacts between the molecules in the all amine case reinforce each other. In all the simulations, only the amine terminations of the dendrimer interacted with the DNA, no hydrogen bond contacts were detected for the acetamide groups. These interactions were primarily with the phosphate oxygens of the backbone, although there were a few instances in which the dendrimer amine groups interacted directly with the DNA bases and sugars. In particular, contacts were made with the O2 of

Cytosine, and O6 and N7 of Guanine and O3' of the sugar.

Table 4.1: Hydrogen Bond Interactions between Dendrimer amines and DNA

	Avg. Number of Contacts	%Contacts to Phosphate	%Contacts to Sugar	%Contacts to Base
All-amine, orientation 1	15.85	94.4	3.7	1.9
All-amine, orientation 2	11.00	95.4	4.1	0.0
Amine-acetamide, half-and-half	4.00	89.4	8.1	2.5
Amine-acetamide, random	2.10	89.8	4.6	4.6

The type of contacts detected in these simulations indicate that the interaction between dendrimers and DNA is likely to be nonspecific and is primarily driven by electrostatic contacts between charged groups, as one would expect from the high charge-density of the molecules. There are very few interactions between the dendrimer and the DNA bases and no obvious specificity to those interactions. Furthermore, all interactions occur on the surface of the DNA, either along the backbone or within a groove. We see no evidence of dendrimer terminations intercalating into the DNA. Surface structures of DNA with dendrimer interaction sites highlighted in red are shown in Figure 4.4. The first orientation of the all-amine dendrimer interacts with about 9-10 bases at a time, with a span of 16 base pairs. The second orientation has slightly fewer interactions, with hydrogen bonds to about 8 bases and a span of 12bp. The mixed termination dendrimers both interact with 3-6 bases and span of 8bp.

Effect of Complexation on DNA structure As shown in the final complexed structures (Figure 4.3), the first orientation of the all-amine terminated dendrimer has the greatest effect on the structure of both molecules in complex. The dendrimer stretches out to cover as much of the DNA as possible, inducing a kink in the DNA in the process. The primary bend occurs in the first major groove of the DNA between the 3rd and 12th base pairs, although at times a second bend occurs between bases

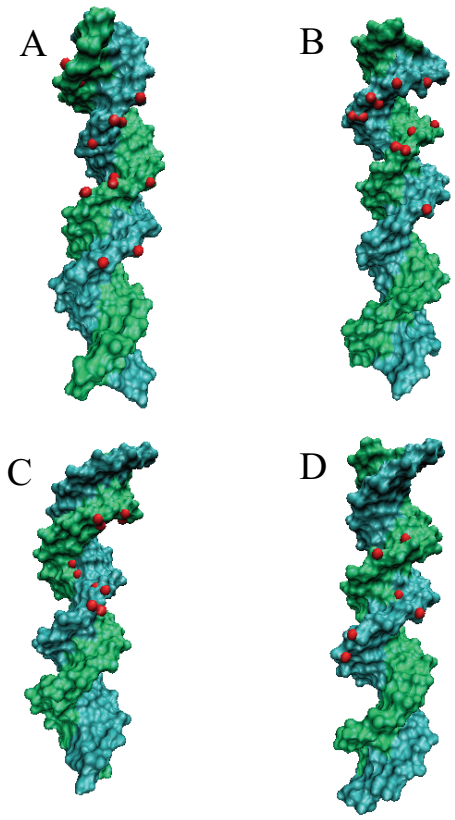

Figure 4.4: DNA structures with dendrimer contact sites highlighted in red. The first orientation of the all-amine dendrimer (A) has contacts that span 10-15bp. The second orientation (B) spans 8bp, and the mixed termination half-and-half (C) and randomly distributed (D) dendrimers span 6-8bp

19 and 24. The total degree of backbone curvature, measured as the total of all angles formed by the DNA backbone, ranges from 19.91° to 108.81°, with an average value of 51.78° (Figure 4.5). Base-pairing is maintained in the bound structure. The second orientation of the all-amine terminated dendrimer and the mixed termination dendrimers do not have this effect, most likely because there are fewer contacts with the DNA.

The extent of bending seen in the DNA is energetically unfavorable, and must be facilitated in some way by the binding of the dendrimer. It has been shown that DNA bending can be induced by asymmetric charge neutralization, even in the absence of

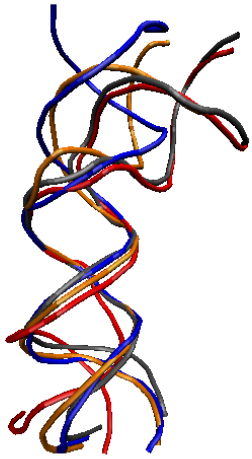

Figure 4.5: Trace structures of the DNA backbone in complex with the first orientation of the all-amine dendrimer, showing the extent and location of curvature induced by the dendrimer.

other attractive forces, possibly due to the removal of local electrostatic repulsions between phosphates [27, 28]. Similar affects have been seen in dendrimers, in which changes in solvent quality such as exposure to acetone or an increase in pH can lead to a collapse of structure [29, 30].

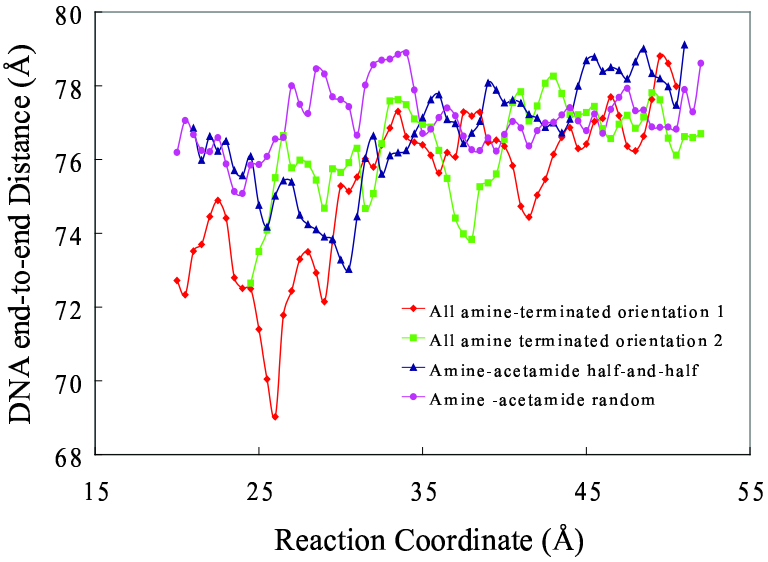

Figure 4.6: Distance between the ends of the DNA molecule as a function of DNA-dendrimer center-to-center distance.

In order to characterize any changes in the structure of the DNA segment as a

function of the approach distance, the distance between the terminal base pairs of the DNA was calculated for snapshots from the simulations and averaged for each value of the reaction coordinate (Figure 4.6). F was calculated for snapshots from the simulations and averaged for each value of the reaction coordinate for the amine terminated dendrimer, the average distance between the ends of the DNA decreases as the molecules come into contact with each other, at the same time as dendrimer contraction, with the total range of DNA length becoming 63 to 78 Å, with an average value of 69 Å. The DNA end-to-end distance also decreases slightly upon interaction with the half-acetamide dendrimer, but the decrease does not appear to be greater than normal fluctuations. Sample structures of the final DNA conformation for the all-amine, orientation 1 dendrimer simulations are shown in Figure 4.5. The dendrimer decreases the distance between the ends of the DNA by inducing a kink in a major groove.

It is interesting to compare the DNA bend angle upon dendrimer binding to what is structurally known in molecular biophysics regarding DNA-binding proteins that bend DNA. Several proteins induce bending of DNA to varying degrees as part of their function, including histones [31, 32], viral DNA packing proteins [33], and many transcription factors [34, 35, 36]. In the first two cases, the purpose of DNA bending is to decrease the size of the molecule. The purpose of DNA bending by transcription factors, which can induce bend angles ranging from 23 to 130° [36], is less clear. DNA bending by promoter proteins may affect transcription by facilitating binding of co-regulatory proteins [37], promoting interaction between RNA polymerase and upstream DNA [38], or increasing the binding affinity of the promoter [39]. It has also been proposed that DNA bending may store energy that could be used to induce structural changes in upstream DNA [40].

It is known that dendrimers condense DNA, and that this condensation is necessary to promote transfection [41]. However, G3 does not appear to pack DNA in the same fashion as nucleosomes, in which DNA is bent into a circular structure around large protein complexes [32]. In contrast, the dendrimer only induces a small local bend in the DNA. It is more likely that dendrimers condense DNA by multiple discrete bends and by bridging to two segments of non-adjacent DNA [14]. In this way dendrimers are more like other DNA packing proteins, such as the mitochondrial protein Abf2p. Abf2p is a 20 kDa protein that organizes DNA in the yeast mitochondria by bending the molecule at multiple sites by an average of 78° [42].

The bend angle induced by the dendrimer is also similar to those induced by transcription factors. Catabolite activator protein (CAP), for instance, induces an overall bend angle between 60 and 90° by kinking the DNA in a major groove and to a lesser extent in the minor groove [43]. CAP is a dimeric protein with a molecular weight of 22.5 kDa [44], much larger than the G3 dendrimer, which has a molecular weight of approximately 6.9 kDa. That the dendrimer is able to bend DNA to a similar extent as a much larger protein most likely has to do with the number of electrostatic contacts between the dendrimer and the DNA, as opposed CAP which has only 6 [43]. The flexibility of the dendrimer compared to most proteins may also play a role, as the dendrimer is able to stretch laterally along the DNA strand to interact with regions far apart from each other.

Novel dendrimer contraction mechanism electrostatically driven DNA-dendrimer condensation Not only does the DNA contract on scales of the order of its persistence length in our simulations, but the dendrimer also contracts on scales comparable to its size. The deformation of the dendrimer upon interaction with the DNA was characterized by calculating the dendrimer's radius of gyration (Fig. 4.7),

defined as

$$R_g = \sqrt{\frac{1}{N}\sum_{i=1}(r_i - r_0)^2} \tag{4.1}$$

where N is the number of atoms in the molecule and r_i is the position of each atom relative to the center of the molecule, r_0. As with DNA end-to-end distance, R_g was calculated for each saved snapshot and averaged over values of the reaction coordinate.

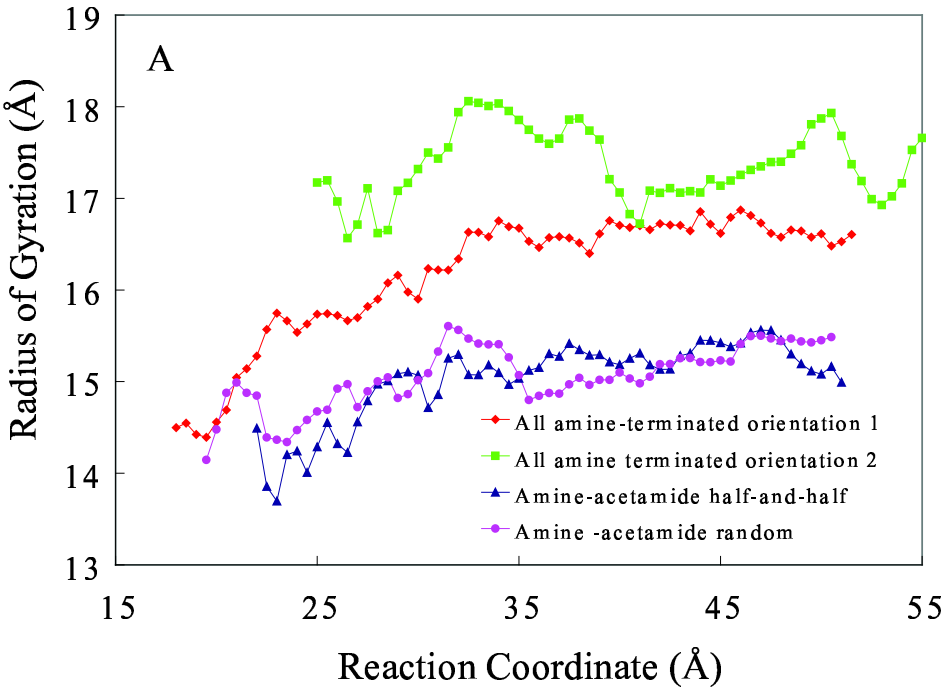

Figure 4.7: Radii of Gyration for the amine-terminated and half amine-half acetamide dendrimers.

As can be seen in the snapshots in Fig. 4.2 and as gauged by the values in Fig. 4.7, when not interacting with DNA, the all-amine terminated dendrimer has a significantly more spread-out structure than the mixed termination dendrimers, due to electrostatic repulsion between the terminations. This difference is reflected in the R_g for the equilibrated stand-alone dendrimers (i.e, in the absence of interactions with the DNA). The first orientation of the all-amine dendrimer deforms considerably upon binding the DNA, flattening against the DNA to maximize the number of con-

tacts. The mixed termination dendrimers also deform to accommodate interactions with the DNA. Only the second orientation of the all-amine dendrimer experienced no significant change in structure upon binding. The radius of gyration for the amine terminated dendrimer decreases most substantially upon interaction with the DNA, from an average of 16.7 Å to 14.4 Å. This suggests that the interaction between the dendrimer terminations and the DNA is strong enough to overcome the electrostatic repulsion between the dendrimers terminations and allow the dendrimer to adopt a more condensed form. The R_g for the half-acetamide dendrimers also decreases, from an average of 15.3 Å to 13.7 - 14.1 Å. In addition, the dendrimer's flexibility allows it to adopt a structure that maximizes its contacts with the DNA. Interestingly, the R_g of the complexed dendrimer for the first orientation all-amine dendrimer is the same as that for the mixed termination dendrimers, while in solution it is much more extended.

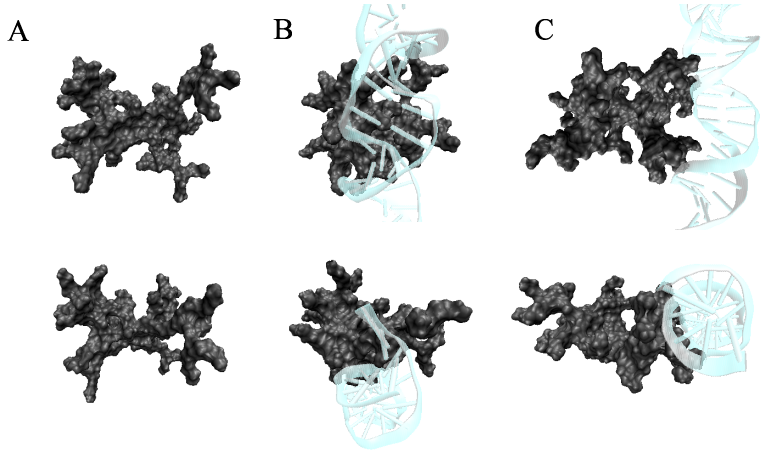

Figure 4.8: Details of Dendrimer structure upon complexation with DNA. A. side and top view of dendrimer in solution. B. Orientation 1 all-amine dendrimer in complex with DNA. C. Orientation 2 all-amine dendrimer in complex with DNA.

Of the second orientation of the amine-terminated dendrimer, only the lobe fac-

ing the DNA at the start of the simulation interacts with the DNA. Although the dendrimer is able to deform in directions perpendicular to its core, its structure does not change much in the direction parallel to it. The other side of the orientation 2 all amine dendrimer is therefore unable to reach the DNA. Surface structures for the all-amine dendrimers are shown in figure 4.8. From the structures it is clear that the decrease in size in the first orientation is primarily along one direction.

The half-and-half dendrimer also only interacts preferentially with one of its sides. The strong attractive force between the DNA's phosphates and the dendrimer's amine-terminations causes the half-and-half dendrimer to reorient itself over the course of the simulations so that the amine-terminated lobe is facing the DNA, leaving the acetamide terminated lobe unable to interact the DNA.

Small angle x-ray scattering experiments on the interaction between generation 3 dendrimers and DNA indicate that G3 condenses DNA into two structural phases, a hexagonal columnar phase, in which the dendrimer is surrounded by 6 strands of DNA, and a square columnar phase, in which 4 strands of DNA surround the dendrimer [20]. Details of the dendrimer structure, based on the DNA spacing, indicate that the dendrimer is more condensed in the hexagonal phase, with an aspect ratio near 1 for the two primary axes. In the square phase, the dendrimer structure appears more ellipsoidal, with a much larger difference between the sizes of the two primary axes. These values are in good agreement with the amount of deformation we see in the all-amine dendrimers upon complexation. For the all-amine dendrimers in our simulations, the ratio of the half lengths of the axes for the face perpendicular to the DNA and the face parallel to the DNA is approximately 0.84 for orientation one and 1.4 for orientation 2 of the all-amine dendrimer. The SAXS experiments predict ratios of 0.88-1.14 for the hexagonal phase and 0.9-1.78

for the square phase. Based on these values, we would expect the dendrimer bound in the orientation 1 configuration to form the hexagonal phase structure with DNA and the dendrimer bound in the orientation 2 configuration to form the square phase.

Taken collectively, the observations presented above regarding the conformational changes upon binding seen in the simulation reveal a novel insight into the mechanism of DNA-dendrimer condensation. In addition to the "macroscopic" condensation that is seen in the experimental studies [14], our simulations reveal a "microscopic" condensation in the DNA-dendrimer systems, a molecular level condensation by which both the dendrimers and DNA contract locally to decrease their sizes in comparison to the sizes they had when free (see the radius of gyration plots and DNA end-to-end distance plots and previous discussion).

It is additionally interesting to compare the contraction of the dendrimer to what happens in protein-DNA interactions. Unlike transcription factors [36] and proteins that bind to DNA to condense it inside viral capsids [33] or in the nucleosome [31], which are already compact when folded, dendrimers suffer also local condensation in addition to DNA condensation. Future experiments using NMR structural measurements are possible to look at this microscopic effect, in addition to further experimental and simulation work to gauge the dependence of the extent of microscopic dendrimer contraction on the size (i.e., generation) of the dendrimer.

4.3.2 Energetic Analysis of Complexation

Potentials of Mean Force To compare the free energy profiles along the approach coordinate for the various dendrimer-DNA systems and orientations, the potentials of mean force for the reaction coordinate, the COM to COM distance were calculated using the weighted histogram analysis method, as described in Ch. 2.2 (Fig. 4.9). The first orientation of the amine-terminated dendrimer, with its largest

side facing the DNA, has the most favorable change in free energy, about -13.5 kcal/mol. The other orientation of the all-amine dendrimer has a slightly lower total free energy change, of -10.9 kcal/mol. Because the free energy difference exceeds four times k_BT at room temperature, we expect, on the basis of thermodynamic considerations, a significantly larger population of these dendrimers bound with the large-side interface to DNA.

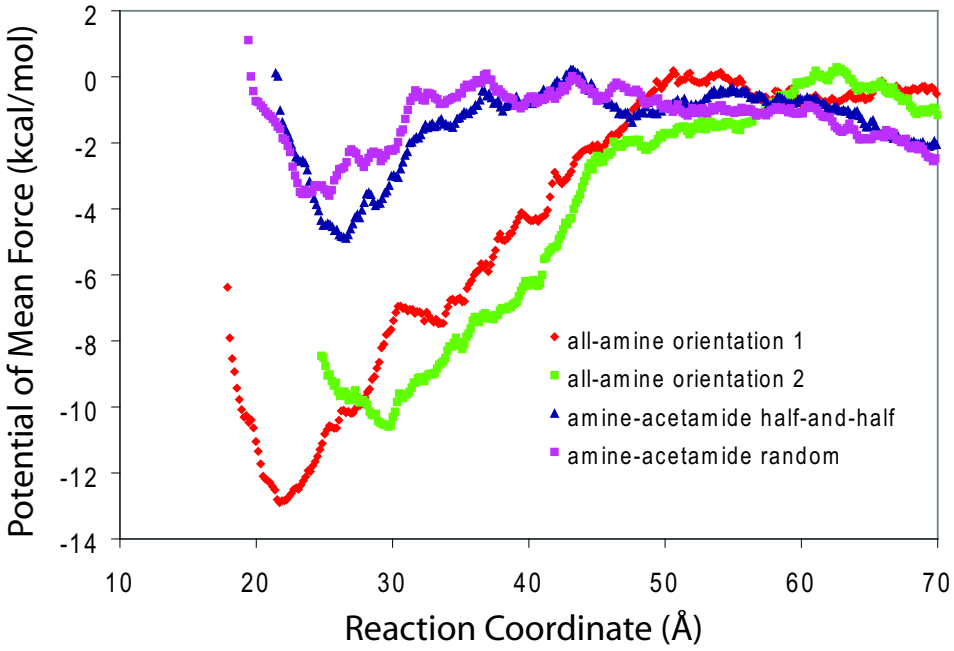

Figure 4.9: Potentials of Mean Force as a function of the distance between the central base pairs of the DNA and center of mass of the dendrimer. The differences in the location of the minimums have to do with the orientation and change in structure of the dendrimers upon interaction with the DNA.

The two half-acetamide dendrimers have lower free energy changes, of -5.1 kcal/mol for the half-and-half dendrimer and -4.6 kcal/mol for the dendrimer in which the acetamide and amine terminations were randomly distributed. The similarity in free energy change for the amine-acetamide dendrimers implies that the distribution of charges on the dendrimer matters less than the total number of charges. The difference between the all-amine terminated dendrimer and the dendrimers with mixed

terminations, however, is slightly larger than would be expected for a purely electrostatic interaction. The reasons for this difference, which we attribute to hydration effects, are discussed in more detail in a later section.

Electrostatic Energies It is expected that for systems with such high charge-density, the interaction will be largely driven by electrostatics. Figure 4.10 A shows the electrostatic potential between the two molecules for all DNA-dendrimer systems as a function of COM to COM distance. Energies were calculated for snapshots of the DNA and dendrimer isolated from the rest of the system. As a consequence, these energies do not include screening effects from water and ion molecules, but represent only the maximum electric potential energy between the molecules at a given distance and configuration. The energy change decreases as a function of distance squared for all systems. The total electrostatic change for the amine-acetamide mixed termination dendrimers is on the order of half the energetic change for the all-amine dendrimers.

The correlation between the electric potential energies of the isolated molecules and the free energy change calculated from the simulations is shown in Figure 4.10 B. The relationship is roughly linear. Regression line fit to the data yield a slope of 0.0012 for all four DNA-dendrimer systems. The fit is best for the mixed termination dendrimers, while the second orientation of the all-amine dendrimer experiences a dip in the otherwise straight line near the middle of the curve, possibly due to changes in the dendrimer's orientation in this region. Interestingly, the last 4.2 kcal/mol of the first orientation of the all-amine dendrimer does not correlate at the same rate as the rest of the curve with the electrostatic potential between the molecules, but rather jumps dramatically. This long-range interaction which does not correlate with electrostatic energy, corresponds to the region of the PMF which is attributed

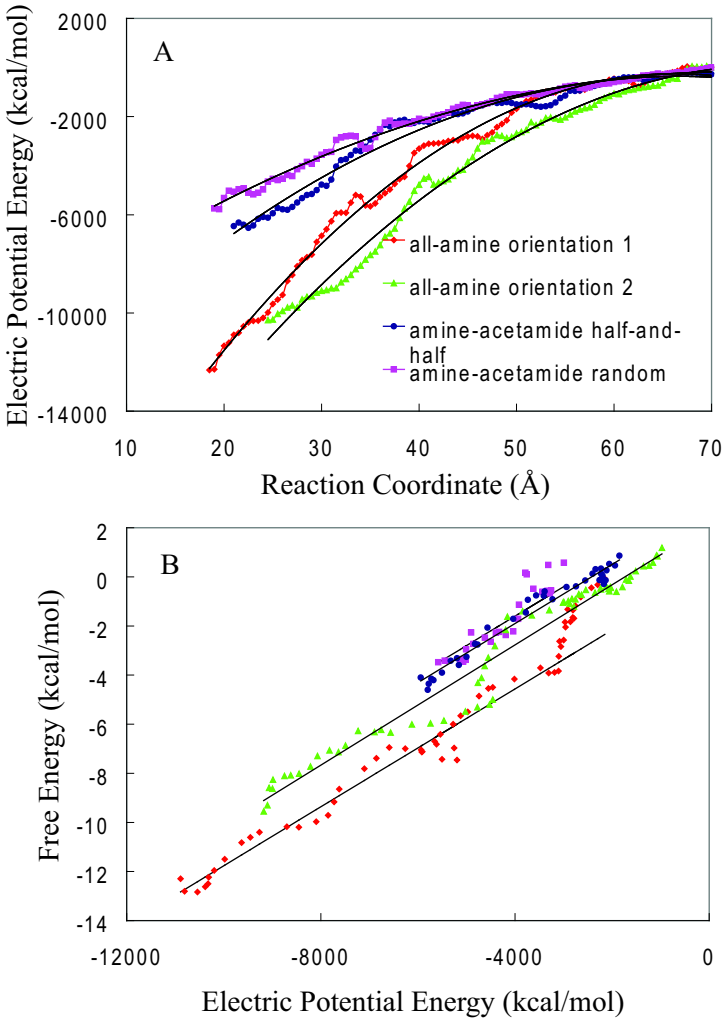

Figure 4.10: A. Electrostatic interaction energies between the dendrimer and DNA for all four systems. The energy was calculated from snapshots of the dendrimer and DNA isolated from the rest of the system and does not include shielding effects from water and ions. The energy change decreases as a function of R^2. B. Correlation between electrostatic energy and free energy of interaction. For all the systems, the electrostatic energy correlates roughly linearly with the free energy of interaction, with a slope of 0.0012 (black lines). The exception is the last 3.4 kcal/mol of the first orientation of the all-amine dendrimer.

to ordered waters.

Other Contributions to the Interaction Free Energy Although the distance between the centers of the dendrimer and DNA is a good reaction coordinate for describing the interaction of a particular DNA-dendrimer pair, it is not best suited to study how different DNA-dendrimer pairs compare with each other. This is be-

cause the COM to COM distances for different dendrimers interacting with DNA span different ranges depending on the size of the dendrimer. In order to better compare the systems, a potential of mean force was also calculated for the minimum distance between the dendrimer and DNA, where minimum distance was defined as the shortest distance between any atom of the DNA and any atom of the dendrimer. The potential of mean force for a value other than the reaction coordinate can be calculated using WHAM from the conditional probability distribution of the coordinate of interest with respect to the biased reaction coordinate [45]. According to the method described in Chapter 2.2, a two-dimensional PMF was calculated for the probability function and collapsed along the biased reaction coordinate to give the PMF of the secondary coordinate.

While one might naively think that the COM to COM plot should give the same kind of description as the closest distance plot, i.e., that the difference between COM to COM and closest distance is just a distance shift (as it would be for rigid objects), that is not quite true. This is because the dendrimer has a highly flexible structure and can 'reach out' with one or more of its branches toward the DNA. The minimum distance is also affected by the structures of the different systems. The all-amine dendrimers are more spread out than the mixed termination dendrimers, so COM to COM distance is not directly comparable for these systems. The minimum distance, on the other hand, does not depend on the structure or orientation of the dendrimer. Moreover, the minimum distance enables us to determine different contributions to the free energy. For instance, one can predict how much of the total free energy change is due to the formation of hydrogen bonds by looking at the free energy contribution within hydrogen-bonding distance.

The minimum distance based free energy values are plotted in Figure 4.11.

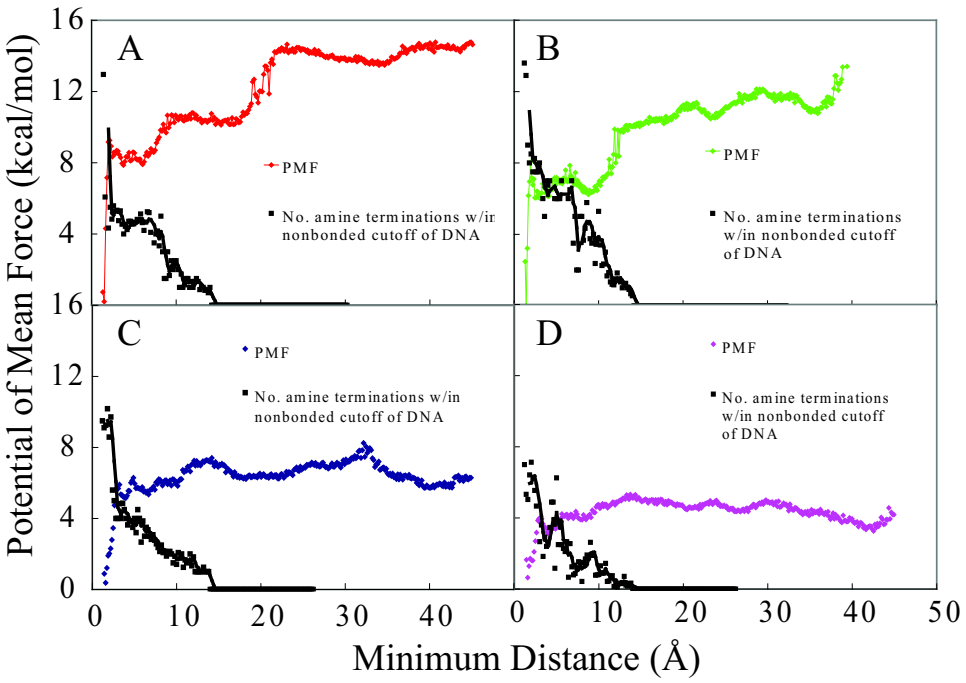

Figure 4.11: Potentials of Mean Force for the minimum distance between the dendrimer and DNA A. all-amine dendrimer orientation 1. B. All-amine dendrimer orientation 2. C. Amine-acetamide half-and-half. D. Amine-acetamide random. The number of amine terminations within the nonbonded cutoff distance of 14 Åis also plotted to show the relationship between the PMF and the number of dendrimer terminations within range of van der Waals forces.

The greatest free energy change for all the systems occurs at hydrogen bonding distance, between 1.9 and 5.0 Å. The all-amine, orientation 1 dendrimer has the largest change, of -8.4 kcal/mol. The second orientation is only slightly smaller, -7.2 kcal/mol, and the mixed termination dendrimers are much more so, -5.6 kcal/mol for the half-and-half dendrimer, and -3.8 kcal/mol for the randomly distributed dendrimer. These values are consistent with the number of hydrogen bonds formed in the simulations, corresponding to a free energy change of 0.59-0.71 kcal/mol per H-bond.

The next decrease in free energy, between 14.0 and 5.9 Å, corresponds to the distance at which a significant number of amine terminations are within electrostatic range of the DNA. The plateau and slight peak between these two areas is most likely

due to desolvation, ion loss, and possibly a loss in entropy. The various decreases and plateaus in this region correspond to increases and plateaus of the number of terminations within the nonbonded cutoff, which also corresponds roughly to the electrostatic range in the presence of solvent (black lines, Figure 4.11), which supports the assumption that this second region of free energy change is caused by electrostatic attraction between the molecules. The change in this region is about 3.0 kcal/mol for the first orientation of the all-amine terminated dendrimer, and 3.3 kcal/mol for the all-amine orientation 2 dendrimer. The shape of the free energy curves is similar for these two dendrimers, with the PMF beginning to decrease when two terminations are within electrostatic range and leveling off when there are approximately four terminations within the this range. For the randomly distributed amine-acetamide dendrimer the total change here is 1.7 kcal/mol, and the number of terminations within range is only one to two in this region. The half-and-half dendrimer has almost no change in this region. Unfortunately, the PMF of this system is problematic due to a tendency of the dendrimer to reorient so that its amine terminated side was closest to the DNA, and the unusual shape of the PMF, starting at a lower free energy and then increasing is probably due to this rearrangement.

The orientation 1 amine terminated dendrimer has a third large free energy change of about 4.3 kcal/mol between 22.3 and 17.2 Å, beyond the expected electrostatic range. The other orientation of the amine terminated dendrimer also has a smaller free energy change at this distance, of about 1.9 kcal/mol. This final change is apparently due to the contribution of ordered waters between the DNA and dendrimer.

4.3.3 Attractive Hydration Forces

Evidence of Ordered Waters A few words are here in place about long-range electrostatic interactions. We have used a particle-Mesh Ewald summation method

that in principle accounts for an infinite range of interactions with no cut-off. Nillson and coworkers have studied extensively the range beyond electrostatics make little contribution to stand-alone DNA molecules in solution. It was found that a cutoff of at least 12 Å is required to maintain the structure of B-DNA, but increasing the cutoff to 18 Å does not improve the results [46]. Similarly, our calculations of electrostatic energy between the molecules shows only a slight interaction beyond these distances, even in the absence of screening molecules. Consequently, we expect little effect from electrostatic interactions beyond a minimum distance of 14 to 16 Å. Strangely, while the electric energy change for the mixed termination dendrimers is approximately half that for the all-amine dendrimer, exactly as expected, yet the free energies for the mixed terminations dendrimers are only approximately one third of the free energy change for the all-amine dendrimer in the same orientation. Furthermore, the free energy changes for the mixed termination dendrimers do correspond to roughly half that of the all-amine dendrimer in the first two regions of the minimum-distance PMFs (Figure 4.11). However, interactions at large distances (from COM to COM distances of 40 - 50 Å) account for -4.3 kcal/mol for the orientation 1 all-amine dendrimer, nearly a third of the entire free energy change. It is these long-range interactions that cannot be explained with simple electrostatic attraction

As described above, there is little effect that is expected from electrostatic interactions beyond 14 Å from individual charges, or COM to COM distances of about 40 Å, and yet nearly a third of the total free energy change for the all-amine dendrimer occurs at distances beyond the expected elctrostatic range of these simulations. We attribute this contribution to the effects of ordered waters between the DNA and dendrimer. Contour plots of the solvent-site dipole moment as a function of center-to-center distance between the dendrimer and DNA are shown in Figure 4.12,4.13.

The solvent-site dipole moment is calculated by dividing the system into a grid with spacing 3.0 Å× 3.0 Å× 3.0 Åand taking the average dipole moment of all the waters to occupy a cube on the grid during the 200ps simulation for each umbrella sampling window. In bulk water, these dipole moments should average to 0.0. A dipole moment significantly greater than 0.0 indicates that the waters in a given position are orienting in a certain direction due to an electric field.

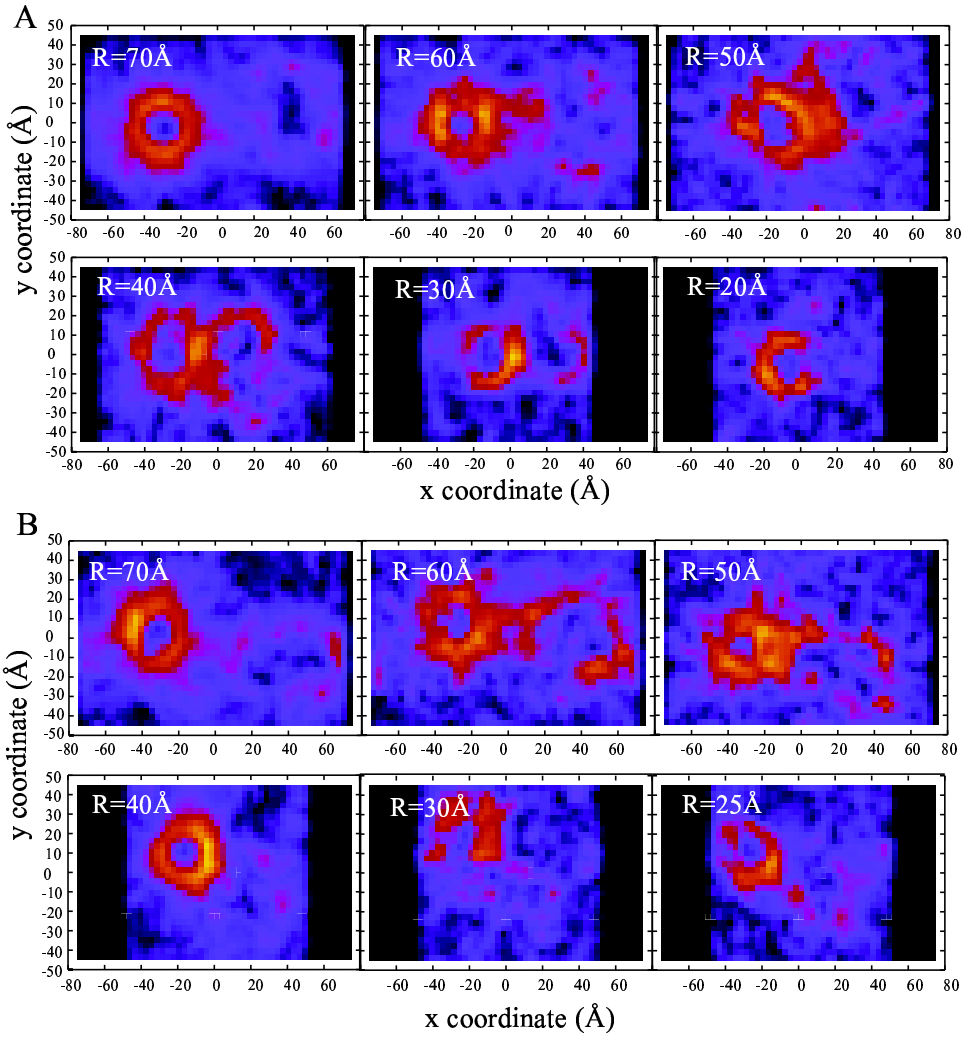

Figure 4.12: Contour plots of solvent-site dipole moments for the all-amine orientation 1 dendrimer (A) and all-amine orientation 2 dendrimer (B). Significant polarization of waters between the dendrimer and DNA are seen at distances of 40 to 60 Å for both all-amine terminated dendrimers.

The DNA has a dramatic effect on the water molecules, as evidenced by a bright

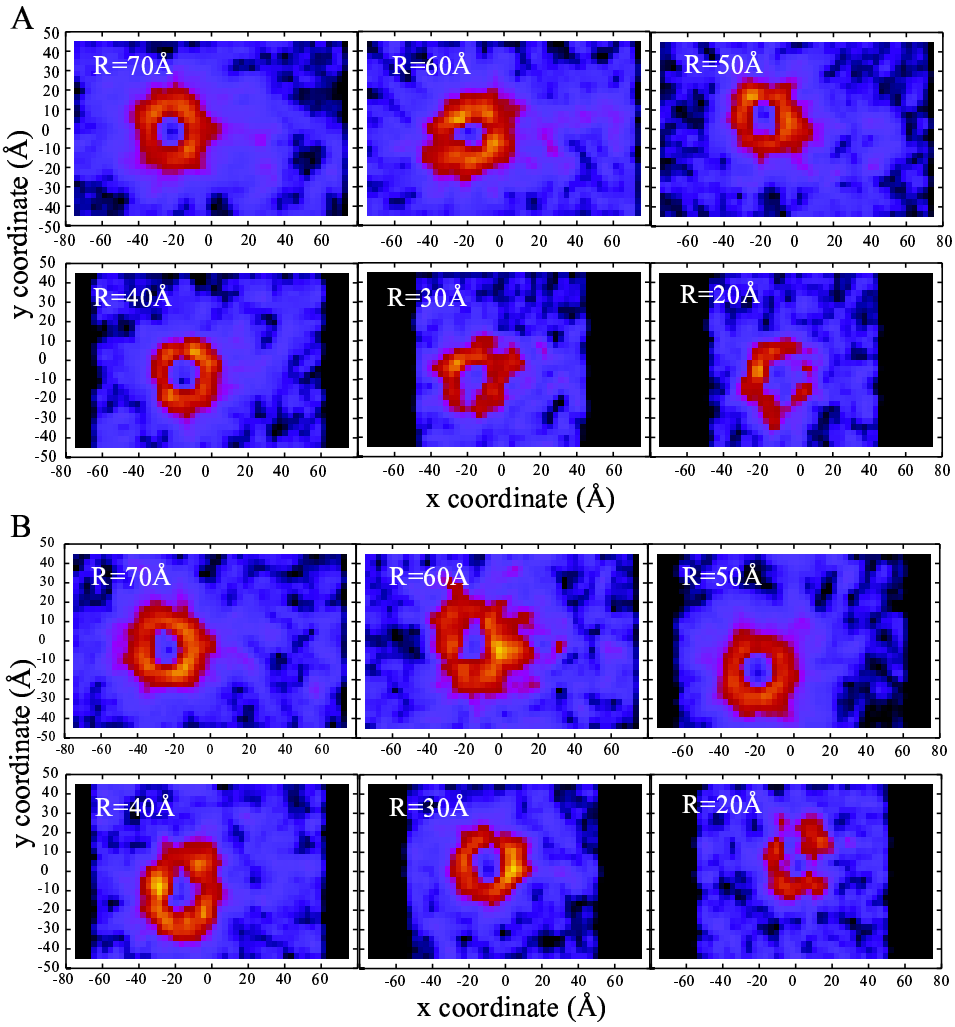

Figure 4.13: Contour plots of solvent-site dipole moments for the amine-acetamide half-and-half dendrimer (A) and the amine-acetamide randomly distributed dendrimer (B). Unlike the all-amine dendrimers there is no bridge of ordered waters formed between the molecules in these systems.

ring of strongly ordered waters within approximately 9.2 Å, or three water layers. Furthermore, the ordering effect appears to be propagated, to a lesser degree, for another 6.2 Å, or approximately two layers of water. The all-amine dendrimer also has an ordering effect on 1-2 water layers, though it is not as strong as the DNA's. The mixed termination dendrimers, on the other hand, show almost no evidence of ordering water. The contour plots show a bridge of strongly ordered waters beginning to form between the all-amine dendrimers and DNA at distances as great as 60 Å.

This effect is particularly strong between 50 and 40 Å for the first orientation and between 60 and 50 Å for the second orientation. There is no such bridge in the amine-acetamide dendrimer systems. Interestingly, however, the ordered water shell around the DNA is disrupted in all four systems in the complexed structure, even on the side facing away from the interaction site.

Figure 4.14 shows a side view of the dipole vectors for solvent sites around the DNA and dendrimer systems over several COM-to-COM distances. Only those sites with a dipole moment greater than 2.0 Debyes are shown. The presence of ordered waters between the branches of the all-amine dendrimer and DNA is clearly visible. There is a bridge of waters between the DNA and the orientation 1 all-amine dendrimer at 50 and 40 Å. The waters between the second orientation also show some ordering, particularly between the upper branches of the dendrimer and a minor groove of the DNA at 50 Å. The mixed termination dendrimers do have a small number of local ordered waters around them but it does not appear that their ordering effect is strong enough to create a bridge between the molecules. Interestingly, the half-and-half dendrimer induces bending of the DNA at 30 Å, a distance at which direct contact has not been made, but at which the ordering of waters between the molecules is significant.

To quantitatively compare the extent of the polarization of waters between the dendrimer and DNA, the average dipole moment per water molecule was calculated for a box between the molecules of volume 30 Å$\times$ 30 Å$\times$ R, where R is the COM-to-COM distance. This average dipole moment is plotted as a function of both COM-to-COM distance and minimum distance in Figure 4.15. The dipole moment is considerably higher for the all-amine orientation 1 dendrimer than for the other 3, although the half-and-half amine acetamide dendrimer also has a high level of

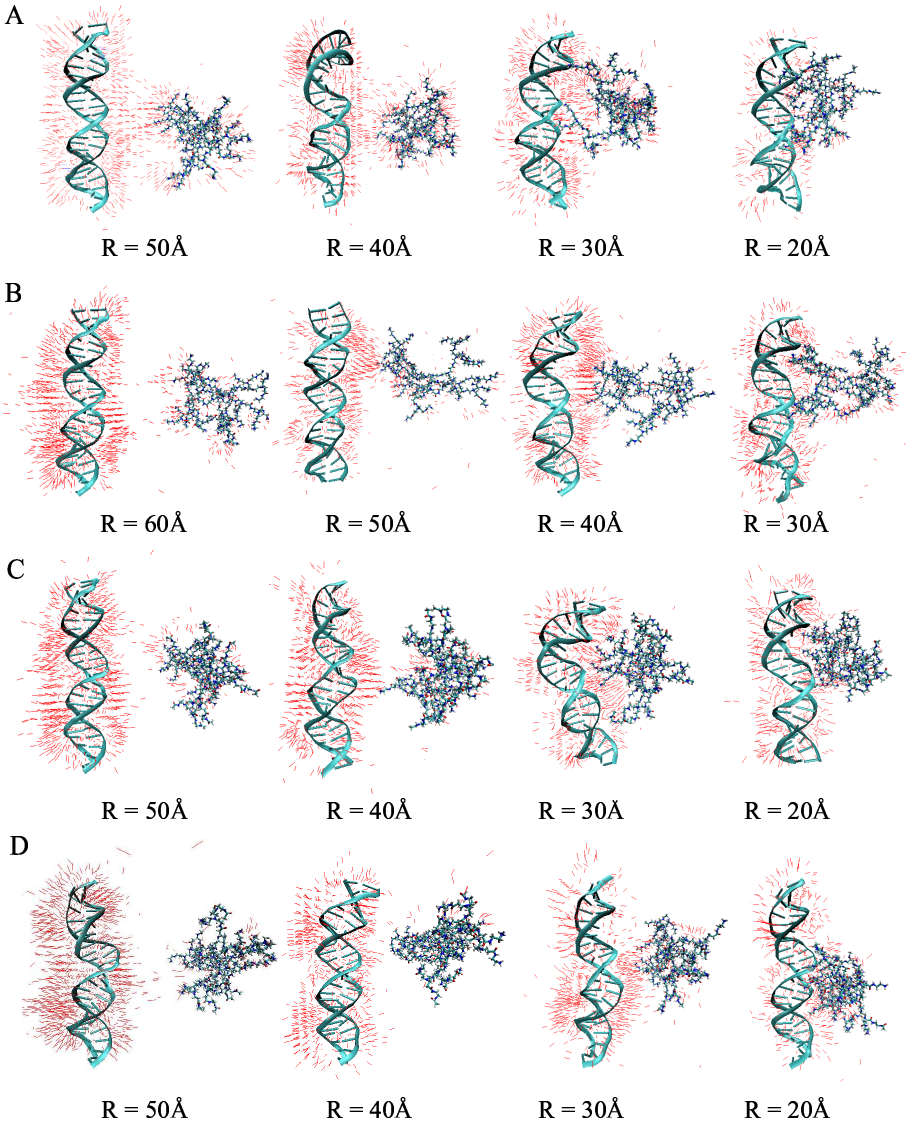

Figure 4.14: Sites of ordered waters are shown in red for the orientation 1 all-amine terminated dendrimer (A) orientation 2 all-amine dendrimer (B) amine-acetamide half-and-half dendrimer (C), and amine-acetamide randomly distributed dendrimer (D).

ordering up to a minimum distance of about 25 Å. The all-amine orientation 1 and the two mixed termination dendrimers show an interesting pattern in which the dipole moment peaks at a minimum distance of 21 and 8 Å, with a decrease in between these distances. The reason for the dip in dipole moment at medium

distances is not obvious, although it may be related to waters being pushed out of the space between the molecules as the dendrimer approaches the DNA.

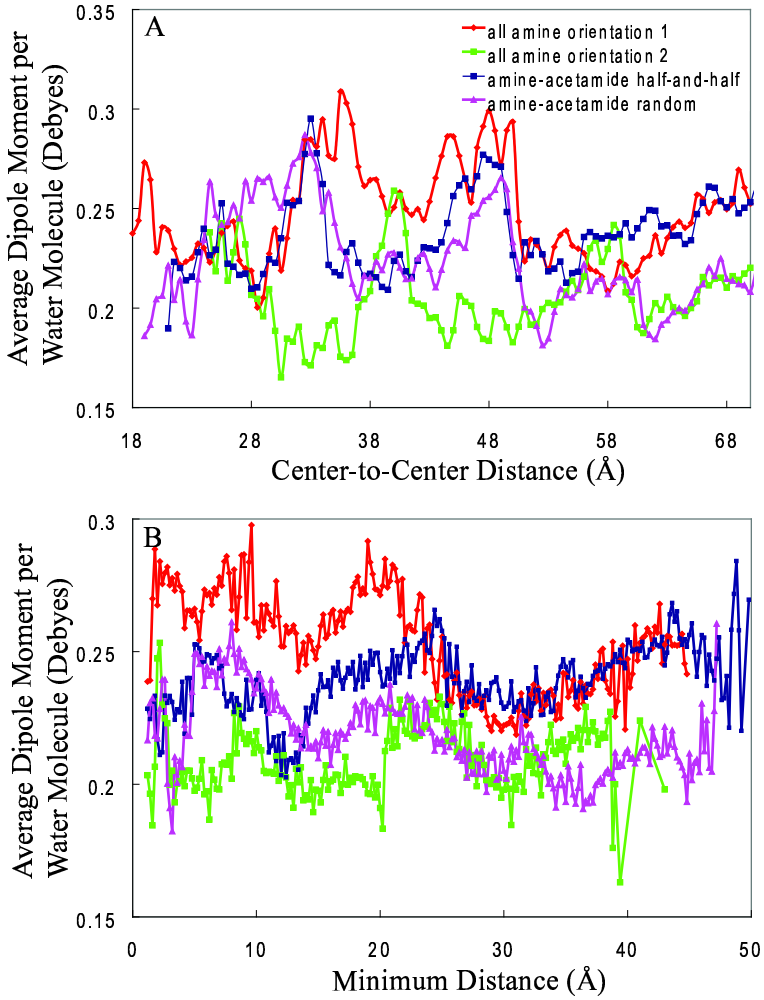

Figure 4.15: Average values for the dipole moment per water molecule, in a box $R_c \times 30 \times 30$ Å, as a function of A. center-to-center distance and B. minimum distance for all four systems.

Theoretical Derivation of Attractive Hydration Force According to the order parameter formalism of hydration forces developed by Marcelja and Radic [47, 48] –a phenomenological theory that involves a Landau expansion of the free energy density functional in powers of the order parameter profile and its gradient– the repulsive pressure between two homogeneous surfaces is

$$P_{\text{rep}} = \frac{R}{4\sinh^2(L/\lambda)} \tag{4.2}$$

and the attractive pressure is

$$P_{\text{attr}} = -\frac{A}{4\cosh^2(L/\lambda)} \tag{4.3}$$

where L is the distance between the surfaces, λ is the characteristic decay length for water ordering, and R and A are force coefficients which incorporate the extent of water ordering at the surfaces. For separation distances much greater than λ, these equations become $P_{\text{rep}} = Re^{(-L/\lambda)}$ and $P_{\text{attr}} = -Ae^{(-L/\lambda)}$. While these are adequate for ideally complementary surfaces, for systems of more complex geometry, in the long length limit the equation is exponential with a geometry dependent prefactor. as done in [50, 51, 52, 53, 54].

The PMF of a system is just the distance-dependent integral of the mean force [49]. Therefore, the free energy contribution of these forces can be calculated as

$$\Delta G(L) = -\int_{\infty}^{L} (P_{\text{rep}} + P_{\text{attr}}) dL', \tag{4.4}$$

[52, 55]. Since the surfaces in our system are complementary, i.e. the ordering of water between the molecules is in the same direction at both surfaces, we can assume that the attractive force is significantly larger than the repulsive force [50]. Additionally, the region in which we assume attractive water forces dominate is at a distance of 17-22 Å, which we expect to be much greater than λ. Therefore, we can simplify the formula for the free energy contribution for our system to

$$\Delta G(L) = -\int_{\infty}^{L} Ae^{-(L'/\lambda)} f(L') dL' \tag{4.5}$$

where f(L') is a geometry-dependent function that describes the structures of the two surfaces.

The mean force for the all-amine dendrimer, calculated by taking the numerical derivative of the PMF is shown in Figure 5.4. At the distances observed in our

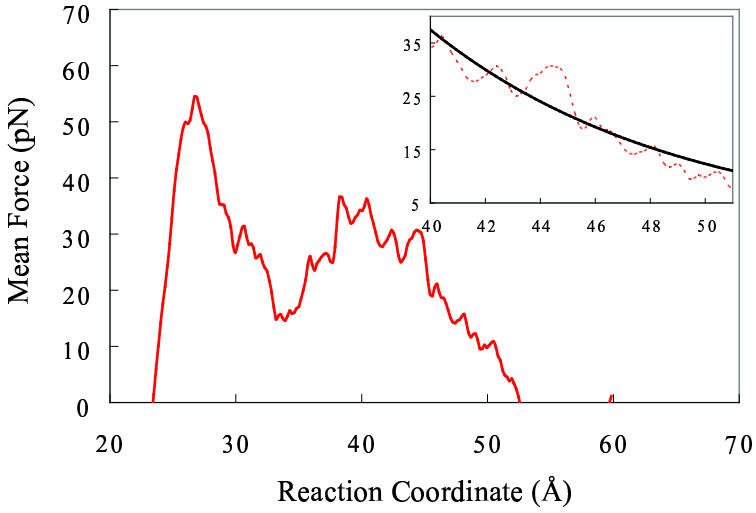

Figure 4.16: Mean Force of the all-amine dendrimer. The region from 40-51 Å, which corresponds to the attractive water forces, is shown in the close-up in the inset. The mean force is region fits an exponential of the form $Ae^{-L/\lambda}$, where L is $R - a_{DNA} - a_{dend}$, $A = 345 \pm 5.78pN$ and $\lambda = 0.9 \pm 0.5\mathring{A}$.

simulation, the effects of geometry appear to be negligible, thus the region which we attribute to attractive hydration forces, 40-51 Åis exponential in shape, as predicted by the order parameter model (Figure 5.4 inset). The exponential fit gives a value of $345 \pm 5.78pN$ for A and $9.0 \pm 0.5\mathring{A}$ for λ. This λ value is approximately twice that calculated from experiments on the interaction between DNA strands condensed by small multivalent cations, which find λ values ranging from 4.0-5.0 Å [50, 52, 51]. This may be due to the fact that we have two surfaces with complimentary charges, which reinforce each other's effect on water structure, as opposed to the experiments cited above, which involve two strands of DNA with some of their charges neutralized by multivalent cations. It is also possible that our simulations overestimate λ due to errors in the water model, which are discussed in the next section.

Measurements of DNA condensation by cations predict a free energy contribution of attractive hydration forces of 1 kcal/mol/bp between DNA strands [50], and 2 kcal/mol/bp in the case where the charges are completely complimentary. Our cal-

culation of a 3.4 - 4.3 kcal/mol contribution from attractive hydration forces seems therefore reasonable given the high charge densities of both the DNA and dendrimer.

Hydration Force and Kinetics It is possible that the way our simulations were

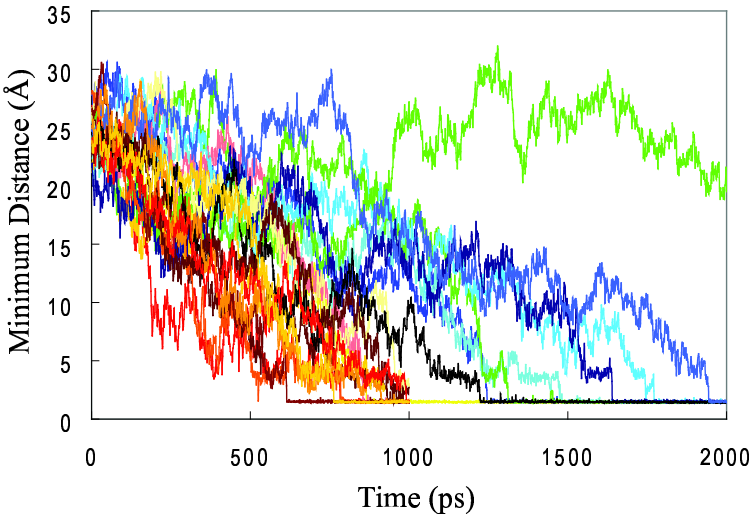

Figure 4.17: The minimum distance between the dendrimer and DNA for the 20 unbiased simulations. The simulations cluster into two groups, "fast" simulations in which contact is made within 1ns, and "slow" simulations in which contact takes over 1ns.

run could contribute to an artificial ordering of waters. While the distance between the dendrimer and DNA is harmonically constrained in each simulation window, the waters and ions are free to move. To test whether artificially holding the molecules apart could lead to the formation of ordered water structures that would not form under normal conditions, we ran a series of 20 constraint-free simulations on the all-amine dendrimer and DNA starting from a center-to-center distance of 50 Å. If the average dipole moment of the waters between the molecules for all these simulations is compared to our umbrella sampling data, it is considerably lower. However, we note that slightly over half of the DNA-dendrimer systems in these simulations complex very quickly, within 1 ns, while the rest take 1.5 ns or longer (Figure 4.17). If we separate the "fast" simulations from the "slow" ones, we find that the faster simulations have a inter-molecule dipole moment that closely matches that of the

umbrella sampling simulations, while the slower ones have a lower, roughly constant dipole moment (Fig. 4.18). The average dipole moment of the "fast" simulations is slightly lower than that for the umbrella simulations for minimum distances from 15-21 Å, however, almost all of the individual simulations have a sharp peak in the dipole moment within this region, although the distance of the peak varies. It is likely that our umbrella sampling simulations represent an ideal case in which the molecules remain oriented with respect to each other in a way that maximizes the ordering of waters between the systems. The difference in water ordering for the fast and slow DNA-dendrimer interactions also implies that these water-bridges have an effect on the rate of interaction between the dendrimer and DNA. It is interesting to note the implications of this in the case of protein-DNA interactions. Unlike dendrimers, which have a uniform charge distribution, many proteins which interact with DNA have solvent-exposed charged groups only in the DNA-binding region. If this water-bridging effect is present in such systems, as simulations by Hamasaki et al. [56] seem to indicate, the presence of such bridges only when the protein is correctly oriented with respect to the DNA could facilitate binding of the protein in the proper orientation.

The behavior of the ions in the all-amine dendrimer system is also worth noting. While the number of ions associated with both the dendrimer and DNA is fairly constant throughout the simulations, the distribution of these ions is not. Figure 5.6 shows a top-down view of the DNA-dendrimer system at different COM-to-COM distances with Cl^- ions in cyan and Na^+ ions in yellow. At approximately 50 Å, the same distance at which the ordered water-bridge becomes apparent, the Cl ions move away from the side of the dendrimer that faces the DNA and reposition themselves on the opposite side. This orientation is maintained until approximately 40 Å, the point

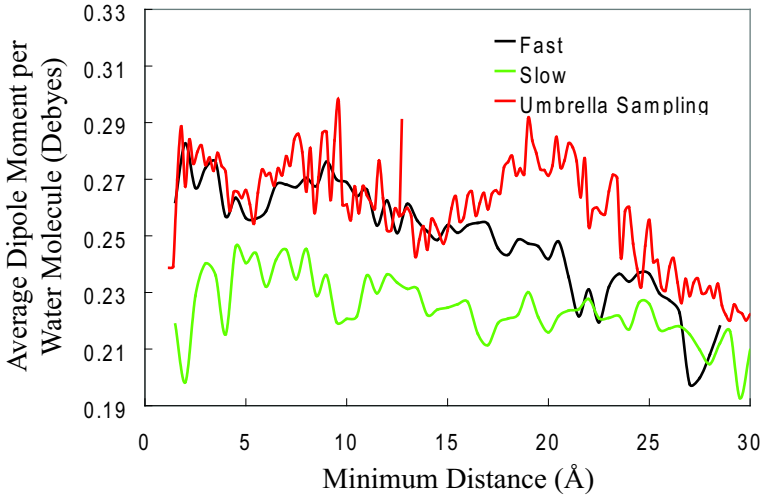

Figure 4.18: Average dipole moments for the water molecules between the all-amine dendrimer and DNA for the umbrella sampling simulations (red) and for the fast (black) and slow (green) unconstrained simulations.

at which the dipole moment of the waters between the molecules dips, and regained at 35 Å when the water dipole moment again increases. Whether this rearrangement of ions is a cause or effect of the ordering of waters is unclear.

4.4 Concluding Discussion

The number of charged terminations and the angle at which the PAMAM dendrimers approach a strand of DNA has a significant affect on both the energy of interaction and the structure of the complex. Generation 3 dendrimers with all amine terminations interact with DNA much more strongly and at greater distances than mixed termination dendrimers, due to the greater number of charges and more spread out structure. In our simulations, all of the dendrimers interacted primarily with the phosphate groups of DNA, although there were a few individual contacts made between dendrimer amines and DNA bases. The acetamide groups on the mixed termination dendrimers appear to have no interaction with any part of the DNA. This is consistent with the nonspecific binding expected between the dendrimer

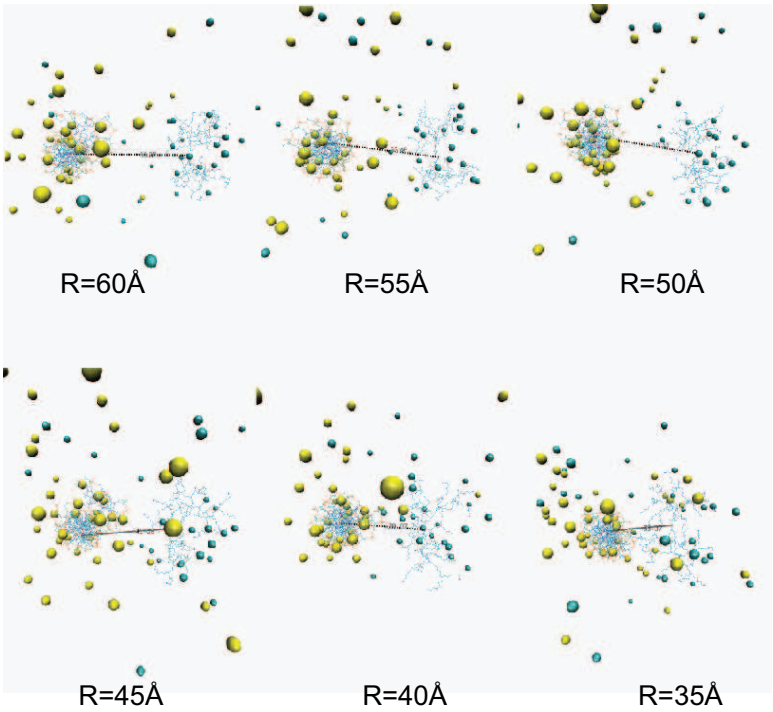

Figure 4.19: Top view of the all amine terminated dendrimer-DNA system with location of ions highlighted.

and DNA. The first orientation of the all-amine terminated dendrimer induced a substantial structural change in the DNA, decreasing the end-to-end distance for the DNA by almost 15 Å and leading to an average bend angle of 51.78°. The dendrimer also deforms significantly during the interaction. This points to a new insight into DNA-dendrimer condensation, whereby not only does DNA condense on the mesoscopic scales comparable to its persistence length when bound to dendrimers (as seen in AFM experiments), but also the dendrimer experiences a local, microscopic contraction on scales comparable to its size, driven by a change in the local electrostatic environment. These results are also in good agreement with the predicted shapes of G3 dendrimers in the columnar DNA phases predicted from SAXS.

The most interesting result of our free energy calculations is the contribution of ordered solvent, at large distances, to the interaction between DNA and the all-amine dendrimer. The high charge density of both molecules causes disruption of the

surrounding water structure, which is propagated through the solvent between the molecules in the case where the dendrimer is oriented to maximize the surface area facing the DNA. The mean force of interaction at the water-mediated distances has an exponential behavior that is in good agreement with the phenomenological order-parameter formalism for water forces developed by Marcelja and Radic [47]. Our simulations also confirm that large, multivalent cations can affect the arrangement of water molecules around DNA when in complex. This has implications for understanding the interaction mechanisms of many systems which involve of molecules with high-charge density, such as DNA condensation by cations, nanoparticle interactions, and DNA-protein interactions. This is especially interesting in the case of DNA binding by proteins, as the results of the unbiased simulations imply that, at least in some cases, water may not only facilitate long-range interactions, but also help to orient the protein appropriately with respect to the DNA. Future work should address whether this type of water behavior is present in DNA-protein systems. Unfortunately, no water model perfectly replicates the behavior of water in experiments. While TIP3P has a dielectric constant close to the experimentally determined one [57], it has an artificially high correlation between adjacent dipoles [58]. This tendency may lead to an exaggeration of the effect of a local electric field on the orientation of waters. While it is therefore likely that our simulations overestimate the contribution of water to the system, we believe that the difference between the all-amine dendrimer and the amine-acetamide dendrimer is too substantial to be a mere artifact of the model.

BIBLIOGRAPHY

[1] D. A. Tomalia, H. Baker, J. Dewald, M. Hall, G. Kallos, S. Martin, J. Roeck, J. Ryder, and P. Smith. A new class of polymers - starburst-dendritic macromolecules. *Polymer Journal*, 17:117–132, 1985.

[2] C. Dufes, I. F. Uchegbu, and A. G. Schatzlein. Dendrimers in gene delivery. *Advanced Drug Delivery Reviews*, 57:2177–2202, 2005.

[3] D. A. Tomalia, B. Huang, D. R. Swanson, H. M. Brothers, and J. W. Klimash. Structure control within poly(amidoamine) dendrimers: size, shape and regio-chemical mimicry of globular proteins. *Tetrahedron*, 59:3799–3813, 2003.

[4] A. G. Schatzlein, B. H. Zinselmeyer, A. Elouzi, C. Dufes, Y. T. A. Chim, C. J. Roberts, M. C. Davies, A. Munro, A. I. Gray, and I. F. Uchegbu. Preferential liver gene expression with polypropylenimine dendrimers. *Journal of Controlled Release*, 101:247–258, 2005.

[5] C. Y. Guo, H. Wang, Y. H. Lin, and Q. L. Cai. Application of starburst (TM) PAMAM dendrimers as DNA carriers in vitro. *Progress in Biochemistry and Biophysics*, 31:804–811, 2004.

[6] M. L. Orberg, K. Schillen, and T. Nylander. Dynamic light scattering and fluorescence study of the interaction between double-stranded DNA and poly(amidoamine) dendrimers. *Biomacromolecules*, 8:1557–1563, 2007.

[7] L. Z. Zhou, L. Gan, H. B. Li, and X. L. Yang. Studies on the interactions between DNA and PAMAM with fluorescent probe [Ru(phen)(2)dppz]2(+). *Journal of Pharmaceutical and Biomedical Analysis*, 43:330–334, 2007.

[8] Y. S. Choi, T. S. Cho, J. M. Kim, S. W. Han, and S. K. Kim. Amine terminated G-6 PAMAM dendrimer and its interaction with DNA probed by Hoechst 33258. *Biophysical Chemistry*, 121:142–149, 2006.

[9] A. Nikakhtar, A. Nasehzadeh, and G. A. Mansoori. Formation and stability conditions of DNA-dendrimer nano-clusters. *Journal of Computational and Theoretical Nanoscience*, 4:521–528, 2007.

[10] A. U. Bielinska, J. F. KukowskaLatallo, and J. R. Baker. The interaction of plasmid DNA with polyamidoamine dendrimers: mechanism of complex formation and analysis of alterations induced in nuclease sensitivity and transcriptional activity of the complexed DNA. *Biochimica et Biophysica Acta-Gene Structure and Expression*, 1353:180–190, 1997.

[11] A. U. Bielinska, C. L. Chen, J. Johnson, and J. R. Baker. DNA complexing with polyamidoamine dendrimers: Implications for transfection. *Bioconjugate Chemistry*, 10:843–850, 1999.

[12] Y. S. Choi, T. Thomas, K. Alina, M. T. Islam, and J. R. Baker. Synthesis and functional evaluation of DNA-assembled polyamidoamine (PAMAM) dendrimer clusters for cancer cell specific targeting. *Abstracts of Papers of The American Chemical Society*, 229:U1151–U1151, 2005.

[13] Y. C. Liu, H. L. Chen, C. J. Su, H. K. Lin, W. L. Liu, and U. S. Jeng. Mesomorphic complexes of poly(amidoamine) dendrimer with DNA. *Macromolecules*, 38:9434–9440, 2005.

[14] F. Ritort, S. Mihardja, S. B. Smith, and C. Bustamante. Condensation transition in DNA-polyaminoamide dendrimer fibers studied using optical tweezers. *Physical Review Letters*, 96:118301, 2006.

[15] C. S. Braun, J. A. Vetro, D. A. Tomalia, G. S. Koe, J. G. Koe, and C. R. Middaugh. Structure/function relationships of polyamidoamine/DNA dendrimers as gene delivery vehicles. *Journal of Pharmaceutical Sciences*, 94:423–436, 2005.

[16] H. G. Abdelhady, S. Allen, C. J. Roberts, M. C. Davies, S. J. Tendler, and P. M. Williams. Atomic force microscopy investigations of the effect of DNase I on DNA-pamam dendrimer complexes. *Biophysical Journal*, 84:471A–471A, 2003.

[17] M. F. Ottaviani, B. Sacchi, N. J. Turro, W. Chen, S. Jockusch, and D. A. Tomalia. An EPR study of the interactions between starburst dendrimers and polynucleotides. *Macromolecules*, 32:2275–2282, 1999.

[18] P. K. Maiti and B. Bagchi. Structure and dynamics of DNA-dendrimer complexation: Role of counterions, water, and base pair sequence. *Nano Letters*, 6:2478–2485, 2006.

[19] A. Nikakhtar, A. Nasehzadeh, H. Naghibi-Beidokhti, and G. A. Mansoori. DNA-dendrimer nanocluster electrostatics prediction with the nonlinear poisson-boltzmann equation. *Journal of Computational and Theoretical Nanoscience*, 2:378–384, 2005.

[20] C. J. Su, H. L. Chen, M. C. Wei, S. F. Peng, H. W. Sung, and V. A. Ivanov. Columnar mesophases of the complexes of in DNA with low-generation poly(amidoamine) dendrimers. *Biomacromolecules*, 10:773–783, 2009.

[21] M. L. Ainalem, A. M. Carnerup, J. Janiak, V. Alfredsson, T. Nylander, and K. Schillen. Condensing DNA with poly(amidoamine) dendrimers of different generations: means of controlling aggregate morphology. *Soft Matter*, 5:2310–2320, 2009.

[22] 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.

[23] 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.

[24] W. F. van Gunsteren and M. Karplus. A method for constrained energy minimization of macromolecules. *Journal of Computational Chemistry*, 1:266–274, 1980.

[25] 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.

[26] 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.

[27] P. R. Hardwidge, D. K. Lee, T. P. Prakash, B. Iglesias, R. B. Den, C. Switzer, and L. J. Maher. DNA bending by asymmetrically tethered cations: influence of tether flexibility. *Chemistry & Biology*, 8:967–980, 2001.

[28] K. M. Kosikov, A. A. Gorin, X. J. Lu, W. K. Olson, and G. S. Manning. Bending of DNA by asymmetric charge neutralization: All-atom energy simulations. *Journal of The American Chemical Society*, 124:4838–4847, 2002.

[29] A. Topp, B. J. Bauer, D. A. Tomalia, and E. J. Amis. Effect of solvent quality on the molecular dimensions of PAMAM dendrimers. *Macromolecules*, 32:7232–7237, 1999.

[30] P. K. Maiti and W. A. Goddard. Solvent quality changes the structure of G8 PAMAM dendrimer, a disagreement with some experimental interpretations. *Journal of Physical Chemistry B*, 110:25628–25632, 2006.

[31] F. Thoma and T. Koller. Unravelled nucleosomes, nucleosome beads and higher-order structures of chromatin - influence of non-histone components and histone H-1. *Journal of Molecular Biology*, 149:709–733, 1981.

[32] T. J. Richmond and C. A. Davey. The structure of DNA in the nucleosome core. *Nature*, 423:145–150, 2003.

[33] J. B. Pesavento, J. A. Lawton, M. K. Estes, and B. V. V. Prasad. The reversible condensation and expansion of the rotavirus genome. *Proceedings of The National Academy of Sciences of The United States of America*, 98:1381–1386, 2001.

[34] M. H. Werner, A. M. Gronenborn, and G. M. Clore. Intercalation, DNA kinking, and the control of transcription. *Science*, 271:778–784, 1996.

[35] J. Perezmartin, F. Rojo, and V. Delorenzo. Promoters responsive to DNA bending - a common theme in prokaryotic gene-expression. *Microbiological Reviews*, 58:268–290, 1994.

[36] P. C. van der Vliet and C. P. Verrijzer. Bending of DNA by transcription factors. *Bioessays*, 15:25–32, 1993.

[37] R. Grosschedl. Higher-order nucleoprotein complexes in transcription - analogies with site-specific recombination. *Current Opinion in Cell Biology*, 7:362–370, 1995.

[38] D. M. . Crothers and T. A. . Steitz. Transcriptional activation by Escherichia coli CAP protein. *Cold Spring Harbor Monograph Series; Transcriptional regulation*, pages 501–534, 1992.

[39] F. L. Lim, A. Hayes, A. G. West, A. Pic-Taylor, Z. Darieva, B. A. Morgan, S. G. Oliver, and A. D. Sharrocks. Mcm1p-induced DNA bending regulates the formation of ternary transcription factor complexes. *Molecular and Cellular Biology*, 23:450–461, 2003.

[40] S. Ryu, S. Garges, and S. Adhya. An arcane role of DNA in transcription activation. *Proceedings of The National Academy of Sciences of The United States of America*, 91:8582–8586, 1994.

[41] M. Thomas and A. M. Klibanov. Non-viral gene therapy: polycation-mediated DNA delivery. *Applied Microbiology and Biotechnology*, 62:27–34, 2003.

[42] R. W. Friddle, J. E. Klare, S. S. Martin, M. Corzett, R. Balhorn, E. P. Baldwin, R. J. Baskin, and A. Noy. Mechanism of DNA compaction by yeast mitochondrial protein abf2p. *Biophysical Journal*, 86:1632–1639, 2004.

[43] A. N. Kapanidis, Y. W. Ebright, R. D. Ludescher, S. Chan, and R. H. Ebright. Mean DNA bend angle and distribution of DNA bend angles in the CAP-DNA complex in solution. *Journal of Molecular Biology*, 312:453–468, 2001.

[44] I. T. Weber, G. L. Gilliland, J. G. Harman, and A. Peterkofsky. Crystal-structure of a cyclic AMP-independent mutant of catabolite gene activator protein. *Journal of Biological Chemistry*, 262:5630–5636, 1987.

[45] E. Gallicchio, M. Andrec, A. K. Felts, and R. M. Levy. Temperature weighted histogram analysis method, replica exchange, and transition paths. *Journal of Physical Chemistry B*, 109:6722–6731, 2005.

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

[47] N. Radic and S. Marcelja. Solvent contribution to Debye screening length. *Chemical Physics Letters*, 55:377–379, 1978.

[48] S. Leikin, V. A. Parsegian, D. C. Rau, and R. P. Rand. Hydration forces. *Annual Review of Physical Chemistry*, 44:369–395, 1993.

[49] J.G. Kirkwood. Statistical mechanics of fluid mixtures. *Journal Chemical Physics*, 3:300–313, 1935.

[50] D. C. Rau and V. A. Parsegian. Direct measurement of the intermolecular forces between counterion-condensed DNA double helices - evidence for long-range attractive hydration forces. *Biophysical Journal*, 61:246–259, 1992.

[51] B. A. Todd, V. A. Parsegian, A. Shirahata, T. J. Thomas, and D. C. Rau. Attractive forces between cation condensed DNA double helices. *Biophysical Journal*, 94:4775–4782, 2008.

[52] R. P. Rand, N. Fuller, V. A. Parsegian, and D. C. Rau. Variation in hydration forces between neutral phospholipid-bilayers - evidence for hydration attraction. *Biochemistry*, 27:7711–7722, 1988.

[53] Z. Arsov, M. Rappolt, and J. Grdadolnik. Weakened hydrogen bonds in water confined between lipid bilayers: The existence of a long-range attractive hydration force. *Chemphyschem*, 10:1438–1441, 2009.

[54] S. Leikin, D. C. Rau, and V. A. Parsegian. Temperature-favored assembly of collagen is driven by hydrophilic not hydrophobic interactions. *Nature Structural Biology*, 2:205–210, 1995.

[55] N. A. M. Besseling. Theory of hydration forces between surfaces. *Langmuir*, 13:2113–2122, 1997.

[56] N. Hamasaki, H. Miyagawa, D. Mitomo, A. Yamagisi, and J. Higo. DNA-protein binding mediated by a solvent site-dipole field. *Chemical Physics Letters*, 431:160–163, 2006.

[57] T. Simonson. Accurate calculation of the dielectric constant of water from simulations of a microscopic droplet in vacuum. *Chemical Physics Letters*, 250:450–454, 1996.

[58] P. Mark and L. Nilsson. Structure and dynamics of liquid water with different long-range interaction truncation and temperature control methods in molecular dynamics simulations. *Journal Of Computational Chemistry*, 23:1211–1219, 2002.

CHAPTER V

Monte Carlo Calculations of Force Extension Curves for DNA Condensation by PAMAM Dendrimers

5.1 Introduction

The basic process that makes dendrimers capable of transfecting DNA is the very substantial compaction of the DNA molecule that takes place upon cooperative dendrimer binding. For example, a μm-long DNA in extended form, can, in the presence of dendrimers, be condensed to a size of tens of nm; this results in a dramatic increase in the density of DNA segments by several orders of magnitude. Such condensation of DNA is know to be effected by several multivalent polyamine cations, including spermidine^{3+} [1, 2], spermine^{4+} [1], and cobalt hexamine^{3+} [3, 4, 5]. The mechanism of this cation induced compaction has been widely studied in force manipulation experiments [6, 7, 8, 9, 10, 11]. In these experiments, the elastic response of cation-condensed DNA is studied by pulling one end of the condensed DNA molecule at constant force or velocity until the molecule reaches its full extension followed by a relaxation stage where the pulling force is reversed and the DNA is allowed to return to the condensed state. The force at the ends of the DNA as a function of the end-to-end distance, the so called force-extension curves (FEC), are measured. These experiments usually reveal one of two characteristic force curves, either a stick-release pattern, which has a sawtooth shape, or a plateau of constant force.

This same technique has been applied to dendrimer condensed DNA in a single molecule manipulation study by Ritort et al. [12]. In that study, single DNA molecules (a 7.2 μm-long λ-phage DNA fragment was used) condensed by dendrimers of a particular generation are pulled from their ends using optical tweezers; the experiment is repeated with three different generations of dendrimer (G5, G6, and G8) and in various ionic conditions. The FECs for the dendrimer-DNA systems reveal characteristic force plateaus and hysteresis between pulling and relaxation (see Fig. 5.5), indicating the existence of a first-order transition between and an extended and a condensed state of the DNA (states which are confirmed, in the same study, by AFM visualization). The optical tweezer manipulations do show that DNA condenses around dendrimers and are able to measure characteristic forces, but whether the dendrimers induce structural changes in the DNA or interact with multiple segments or cannot be determined from the experiments [12]. Also lacking is a microscopic understanding of the decondensation transition and the hysteresis observed in the macroscopic (or, more accurately, mesoscopic) pulling data. For quantitative comparison of the computational study described in Chapter 4, we use the actual experimental single molecule data reported in Ref. [12]; via a Monte Carlo model, we scale up the microscopic observables derived from our atomistic MD simulation to generate macroscopic force-extension theoretical curves that can be directly compared with the measured force-extension curves obtained by optical tweezer pulling.

Potentials of mean force along the approach coordinate, calculated from all-atom molecular dynamics simulations in combination with umbrella sampling (see Ch. 4 for simulation details) were used to produce force extension curves of DNA condensation by a Generation 3 dendrimer, for comparison to the experiments presented

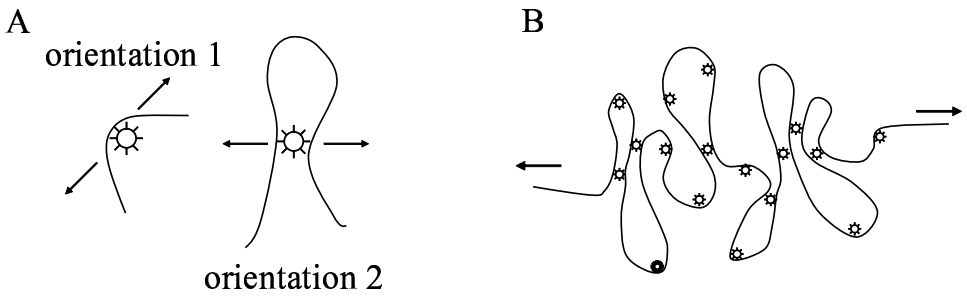

Figure 5.1: Cartoon representation of microscopic (A) and mesoscopic (B) interactions between dendrimers and DNA. The arrows indicate the direction o the pulling force. The relative detachment force felt by the dendrimer will depend on the structure or the DNA-dendrimer complex.

in Ref [12]. For these calculations, only the results for dendrimers with all amine terminations were used. Given our atomistic, MD-calculated PMF (and its two associated geometries) for the separation of a *single* DNA-dendrimer contact, we can extrapolate to the experimentally-relevant mesoscopic length (μm) of an *entire* DNA molecule with N dendrimers bound by using a stochastic model for the attachment/detachment kinetics (see Figure 5.1). We employ an elastically coupled two-level system of contacts, used previously to describe the unfolding of proteins and carbohydrates [13], that evolves according to a Monte Carlo (MC) procedure. A similar stochastic model was used by Ritort et al. [12] to interpret the experimental data in terms of a dynamical equilibrium between an extended and a condensed state. The resulting FEC from our combined MD/MC simulations accurately reproduce the experimental data reported Ref. [12], kindly provided by Professor F. Ritort.

5.2 Methods

The Monte Carlo model used to calculate force extension curves works as follows. For DNA, the FEC, i.e., the dependence of the force F on the extension x is given

by an elastic version of the worm-like chain model (EWLC),

$$F(x,L) = \frac{k_B T}{p}\left(\frac{1}{4(1+\frac{F}{\gamma}-\frac{x}{L})^2} - \frac{1}{4} + \frac{x}{L}\right) \tag{5.1}$$

where p is the persistence length (in our case 20 nm), γ is Young's modulus (2000 pN) and L is the effective contour length, i.e. the length of DNA free to extend without breaking a contact (7.2 μm for the fully extended DNA molecule used in the experiment, with an initial value of 2.7 μm to account for the amount of slack existent in the fully condensed DNA). We note that the persistence length chosen for our model is lower than the standard value of 50 nm. It has been shown that the presence of multivalent cations significantly decreases the persistence length of DNA [14, 5]. Concentrations of cobalt hexamine just below the threshold required to condense DNA reduce the persistence length to 20nm [15], and higher concentrations of this trivalent cation result in a persistence length as low as 15 nm [16]. It is reasonable to assume that highly charged molecules such as dendrimers would also have this effect, and so the value of 20 nm is in fact more suitable for our system than 50 nm would be.

Unlike in typical applications of the EWLC, here the dependence on L of F is made explicit because L changes upon each dendrimer binding/detachment event. Eq. (5.1) describes accurately the naked DNA FEC, i.e., the force for a given length L of DNA without any dendrimer bound. However, it fails to capture the essential features (plateaus and hysteresis) of dendrimer-condensed DNA FECs. To model these, the effect of the pulling force on the rates of contact breaking/formation ($\alpha(F), \beta(F)$) need to be incorporated. In the simplest approach, they are given by the Bell model

[17],

$$\alpha(F) = \omega e^{-(\Delta G_E - F x_E^*)/k_B T)} \tag{5.2}$$

$$\beta(F) = \omega e^{-(\Delta G_C + F x_C^*)/k_B T)}, \tag{5.3}$$

where ΔG_E and ΔG_C are the free energy barriers to break or form a contact, i.e., to extend or to compact, respectively, x_E^* and x_C^* are characteristic widths of the free energy wells, and $\omega = 5 \cdot 10^4$ s is the reciprocal of the diffusive relaxation time. For our model, the variables determining the force-dependency of the rates were derived directly from the PMFs for the two all-amine dendrimers from the umbrella sampling simulations described in Chapter 4. The free energy changes for the two orientations of the all-amine dendrimer were averaged to obtain $\Delta G_E = 12.2$ kcal/mol, $\Delta G_C = 0$ and $x_E^* = 3.1$ nm; to account for the elastic linkage that creates a transient capture well for rebinding (see Evans [18] for details), x_C was set to $\delta L = 15$ nm, the average increase of contour length upon each detachment event (see below). This is expected to be a lower estimate because diffusion away from the DNA binding site limits by the ability of the dendrimer to recombine; even so, for forces beyond 10 pN contact breaking probability dominates and reformation of contacts becomes negligibly small. Moreover, while the recombination at zero force is barrierless in our PMF (as it should be for a short DNA duplex, a possible refinement of our model may include of a recombination barrier in orientation 2 to account for long-scale DNA looping [19]. While this will not change the barrier to break a contact (both "reactant" and "transition state" will be raised by the same amount) and hence will not modify the upper FEC plateaus, it may improve somewhat the fit for the lower, equilibrium curve.

The extension advances linearly with time, $x = vt$, with v the velocity; in practice,

a discretization $t_i = i\delta t$ is used. The evolution of contact counts is given by first order kinetics,

$$dN_C/dt = -dN_E/dt = -\alpha(F)N_C + \beta(F)N_E, \tag{5.4}$$

where N_C is the number of intact contacts and N_E is the number of broken (extended) contacts, with their sum a constant, $N_C + N_E = 300$, a value consistent with the experimental data [12]. At each time step, a Monte Carlo acceptance-rejection scheme is used to determine if a given contact changes its state, according to the probabilities $\alpha(F)\delta t$ and $\beta(F)\delta t$ of a contact being either broken or, respectively, reformed between t_i and $t_i + \delta t$. If a contact is broken, then the contour length, L in Eq. (5.1), is increased by $\delta L = 15$ nm, i.e., by the average contour length per contact; the extension x in the same Eq., modeling the cantilever motion, continues to increase linearly without a jump. Conversely, if a contact is formed, the contour length is decreased by the same amount. In either case, with the new values of x and L, the force $F(x, L)$ for the next step is recalculated using Eq. (5.1). For each contact, the angle of the DNA bend (for orientation 1) or the geometry of the straddling complex (for orientation 2), both of which modulate the direction of the force felt by the dendrimer, can be different. The effect of the resulting disorder (or structural noise) on the force along the pulling direction, averaged over all contacts, was modeled by multiplying the force in Eqs. (2) and (3) by the cosine of a Gaussian-distributed angle with zero mean and standard deviation of 30°. This structural noise could alternatively be calculated by using a distribution of values for the free energy barrier instead of the pulling angle, as was done in the original experimental fit. Ritort et al. [12] used an exponential distribution of the free energy with a mean of approximately 11.79 kcal/mol and variance 5.95 kcal/mol to fit the data, which is in good agreement with our two "extreme" free energy values.

5.3 Results

Structure of DNA-dendrimer complex

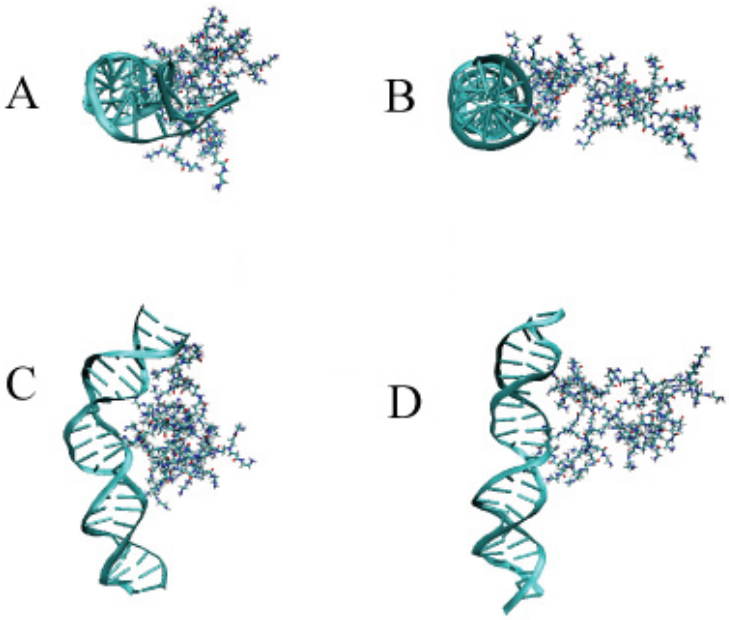

Figure 5.2: Dendrimer structure of the core and first branching generation (center) and four atomic snapshots of the two orientations simulated with MD: Top and side views of the complexed structure for orientation 1 (panels A,C) and orientation 2 (panel B,D).

Snapshots from the all-amine dendrimers from the simulations described in the previous chapter (Figure 5.2) present the structural effect of complexation for the two orientations. The bound structures show that orientation 1 has the greatest effect on the structure of both molecules in complex. In this case, the dendrimer stretches out to cover as much of the DNA as possible, bending the DNA in the process by an average value of 51.78°. Orientation 2 of the dendrimer does not bend the helical axis significantly. Most likely, this is because the second orientation is characteristic of a lateral binding mode in which each of the two lobes binds a distinct helical fragment (i.e., the dendrimer straddles two duplex-DNA segments that are close in space but distant in sequence), while orientation 1 may corresponds to a dendrimer bound to a single, bent DNA segment. These two modes are sketched in Fig. 5.1 A. For

both orientation 1 and 2, the dendrimer-binding contour length for DNA from the simulation (between 10-18 base pairs) is in accord with contour lengths derived from ethidium bromide fluorescence titration experiments on low-generation dendrimer-DNA complexes [20]. Moreover, similarly to what was observed for other cationic polyamines, namely, spermidine^{3+} and spermine^{4+} [1], our simulation shows that G3 binding does not perturb base pairing; this is of crucial importance if low-generation PAMAM dendrimers are to be used for gene compaction pre-delivery.

Effect of Force on Free Energy Profiles In the limit of infinite sampling, the potential of mean force we have computed is the free energy of the system as a function of the reaction coordinate [21, 22], which, by definition, involves integration over all other degrees of freedom. For a spherically asymmetrical molecule (as is the dendrimer), all the relative rotational orientations cannot be sampled within reasonable computation time. While the PMFs reported here and in the previous chapter are for a given interaction geometry without rotational averaging, they report on the contribution of each respective configurational orientation that resulted upon an approach that enabled overall rotation. This approximates well the free energy for an interaction in which the molecules would be free to rotate, i.e., would yield an ergodic average. An important point to make here however involves a discussion of non-ergodicity, which can manifest not only in the simulation when perpendicular degrees of freedom are not sampled, but also in the actual experiments when pulling is fast. Of particular relevance is the hysteresis in the single molecule study [12] for finite pulling rates; a similar hysteresis was observed for a different polycation-compacted DNA conglomerate [19]. The hysteresis is indicative of the coexistence of an extended and a condensed phase. This involves not only metastability through trapping in average energy minima along the pulling coordinate, but also, for indi-

vidual single-molecules, trapping in different dendrimer-DNA orientations that may not be averaged over in the conglomerate. The latter, the perpendicular noise, is the source of heterogeneity in the different single-molecule traces and was modeled by random force directions in our Monte Carlo model (see below).

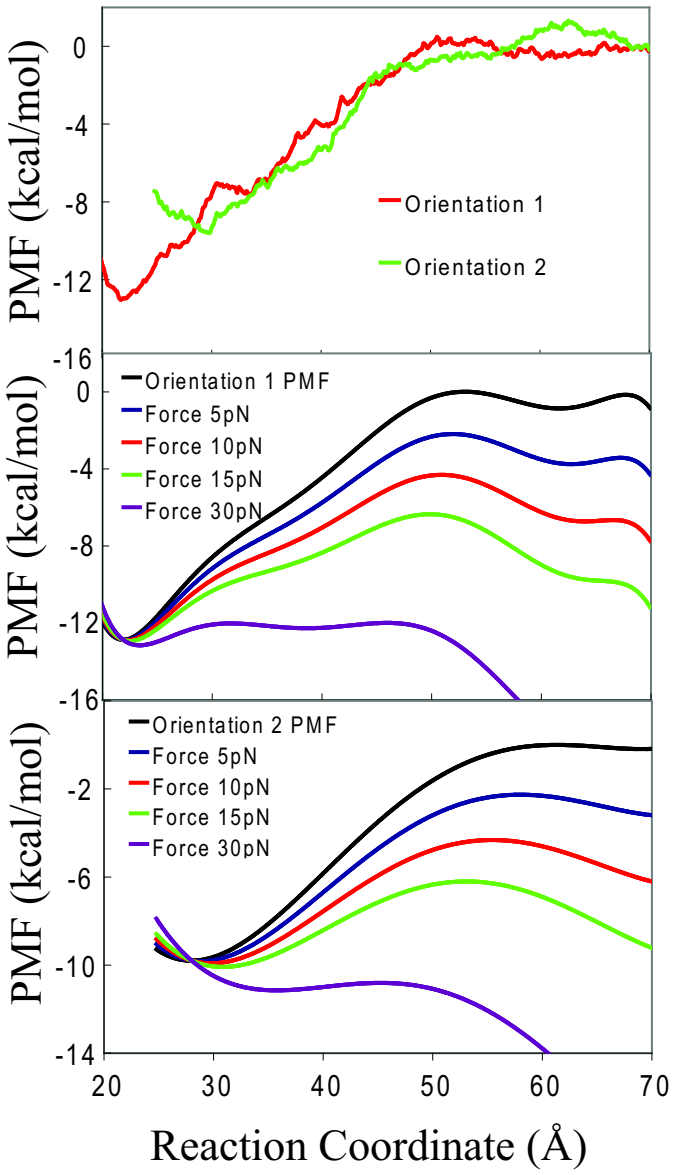

Figure 5.3: A. Potentials of mean force for the two orientations of dendrimer. B. Relative PMFs under various pulling forces for orientation 1. C. Relative PMFs for orientation 2.

The free energy profiles (the PMFs) along the approach coordinates for the two orientations are shown in Fig. 5.3. The first orientation, with its largest side fac-

ing the DNA, has the most favorable change in free energy, -13.5 kcal/mol. The other orientation of the all-amine dendrimer has a total free energy change, of -10.9 kcal/mol. The overall shape of the PMFs reveals the typical profile of long range attraction, a minimum, and a steep repulsion at short distances. The slight differences in the location of the minima have to do with the structural changes of the dendrimers upon interaction with the DNA in the two geometries.

The interaction free energy per cationic charge computed for the 32+ charged G3 dendrimer was 0.028 kT/bp and 0.024 kT/bp for the two orientations. This compares favorably with data from two independent single molecule experiments on DNA condensed by the polycation spermidine^{3+}, reporting an intramolecular attraction per charge of 0.0277 kT/bp [5] and 0.02 kT/bp [2], data also consistent with osmotic stress measurements in bulk condensed DNA. Additionally, our computed free energies falls within the range of yet other experimental free energy per charge estimates of 0.0175-0.036 kT/bp inferred for spermine^{4+} and spermidine^{3+}, respectively [10, 23].

Forces of Interaction The negative gradient of the PMF is, by definition, the mean force [21]. Forces of interaction of the dendrimer-DNA systems (Fig. 5.5), were calculated from the free energy plots by taking the numerical derivative of the potential of mean force with respect to the reaction coordinate. This sets the so-called adiabatic limit for the detachment force, i.e., this is the equilibrium force needed to maintain a certain DNA dendrimer separation when the other degrees of freedom have had time to relax (i.e., upon pulling slowly). This limit can be however exceeded when pulling faster than diffusive relaxation [18]. As seen in Fig. 5.5, the PMFs reveal a distribution of forces in accord to the corresponding experimental study [12]. For example, the approx 20-60 pN force range computed

from our PMFs (Fig. 5.3) is in accord with the force range to break dendrimer DNA interactions, from the easily-detachable to the more resilient ones, observed in the pulling experiments.

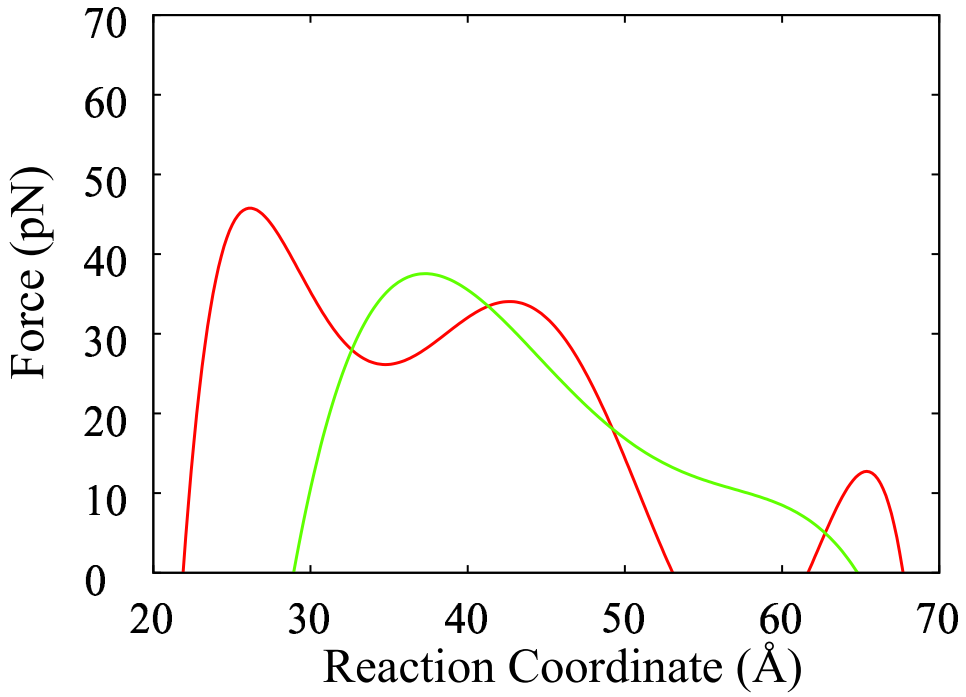

Figure 5.4: Mean forces vs. separation for the two orientations derived from the numerical derivative of the PMFs computed for orientation 1 (red) and 2 (green). Insets (in both panel A and B) depict the two possible mechanisms of DNA condensation.

While we do see bending of the DNA, it is likely that its mesoscopic collapse is also the result of lateral interactions involving two DNA duplexes joined by a dendrimer, and not just of elastic buckling; similar lateral interactions were inferred from experiments on DNA collapse by trivalent cations [5]). As such, we projected the mean forces computed from the PMF along the direction of DNA dendrimer separation as schematized in Fig. 5.4. A force applied to pull the DNA-dendrimer system will lower the free energy of the extended states (see Fig. 5.3, C and D) and will tilt the computed PMF for each dendrimer-DNA contact by $-Fx^*$, where F is the magnitude of the force in the direction of pulling and $x*$ a characteristic length

scale (which may be, in principle, F-dependent) ascribed to the distance from the "reactant region" to the transition state. Each dendrimer-DNA contact can thus be thought of as a two-state system, a detached (extended) state, and a collapsed state. In the presence of dendrimers and a relatively high external force, the detached DNA dendrimer state is favored, while upon decreasing the magnitude of the force the stretched polymer will collapse when the attractive free energy will be lower than the free energy of the detached state. Because there are many contacts per DNA molecule, there will exist a coexistence curve with both detached and collapsed contacts, indicative of a first-order phase transition. The FEC observed in the single molecule data (Fig. 5.5) clearly follow this predicted behavior. In the language of bifurcation theory, the tilting effect of the force on the energy landscape is akin to the effect of changing a parameter in a "fold" catastrophe that can be used to describe microscopically the transition from extended to collapsed states.

The upper FEC plateau of the curve was calculated with v =0.2 μm/s, the actual velocity used in the experiment. For the lower curve, v was decreased to 0.01 μm/s, under an assumption of equilibrium in which all contacts were given the chance of breaking at each time step (akin to sending v to zero). As the velocity approaches zero, the system approaches equilibrium, the upper plateau approximates the condensation curve, and the hysteresis disappears. Fig. 5.5 shows the remarkable similarity between the such calculated FECs and the experimental single-molecule FECs. The fact that we get directly from the umbrella sampling free energy calculations, i.e., without fitting, values for ΔG's and x^*'s that, when translated in to FECs, yield curves comparable to the ones derived by the experiments of Ritort et al., shows that our simulations validate the assumption of the two state, extended-collapsed system used in the interpretation of the experimental FECs.

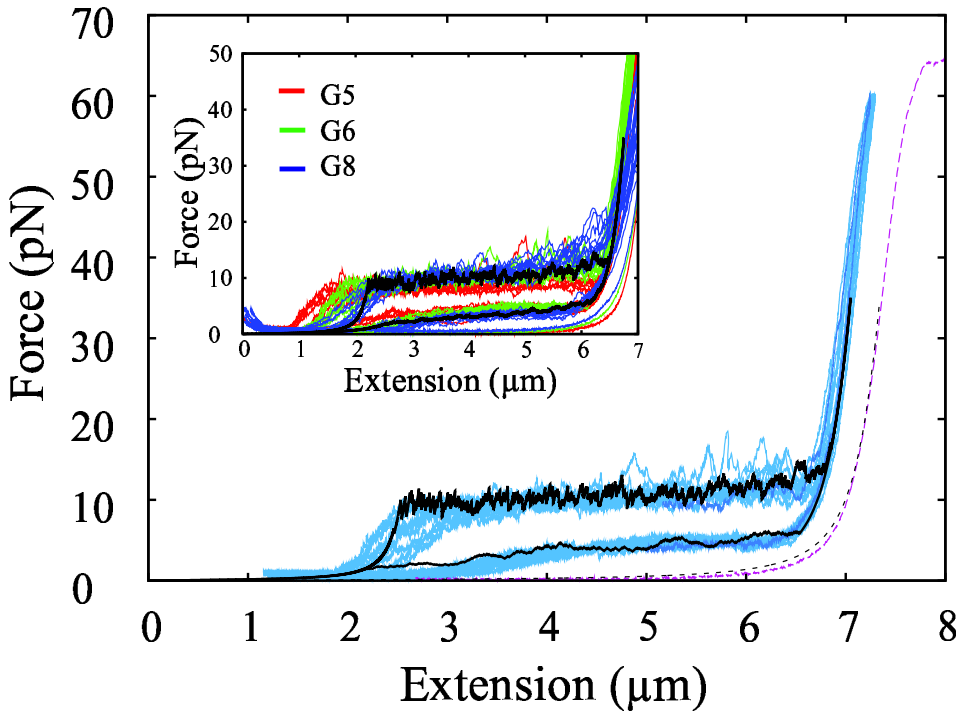

Figure 5.5: Experimental single-molecule force-extension curves (blue, data from Ritort et al.) and theoretical curve (black) computed for an entire DNA molecule with $N = 300$ contacts, using the parameters derived from the umbrella sampling free energy calculations for an individual dendrimer DNA contact (see text). Experimental curve for naked DNA and a fitted EWLC (see Eq. 5.1) are shown in purple and black dotted curves, respectively. The theoretical curve is also shown compared to forces for the other generations of dendrimer studied in the experiments in the inset. The plateau height is similar for all sizes of dendrimer.

It is of interest to consider the effect of solution ions on the FECs in connection to the single molecule data measured at various ionic concentrations. Fig. 5.6 shows the our calculation of the binding free energy for G5 at various ion concentrations of NaCl. For these calculations we do not have estimates for ΔG_E from simulations. Instead we varied the value of ΔG_E so that results of the corresponding MC runs using the elastically-coupled two state model yield FECs that fit the experimental curves. For these calculations we assumed *a priori* that the salt concentration would have no effect on the length to the top of the free energy barrier, although in reality this may not be exactly the case, and used the same value as in the previous fit. We find that the relationship between ion concentration and free energy is linear for concentrations between 50 mM and 500 mM, with increasing ion concentration corresponding to lower free energy of interaction. The linear relationship breaks down at the lowest ion concentration, 10 mM. This relationship is in agreement with experiments on the effect of ion concentration on the free energy of interaction for DNA and a nonspecific DNA-binding protein [24], which also found a linear relationship between salt concentration and binding free energy for all but very low salt concentrations. This makes intuitive sense if the interaction is primarily driven by electrostatics, as one would expect the ions to screen molecular charges. The amine-acetamide mixed-termination dendrimers described in Ch. 4 had ΔG_E values of 3.6 kcal/mol for a dendrimer with randomly distributed charges and 4.6 kcal/mol for a dendrimer with amine charges on one "lobe" and acetamide on the other. These values are below even the energy value fit to the highest ion concentration (7 kcal/mol at 500 mM) based; on that we would predict that the lower-charge dendrimers would not be able to condense DNA, since at 500 mM the DNA acts like the naked DNA curve.

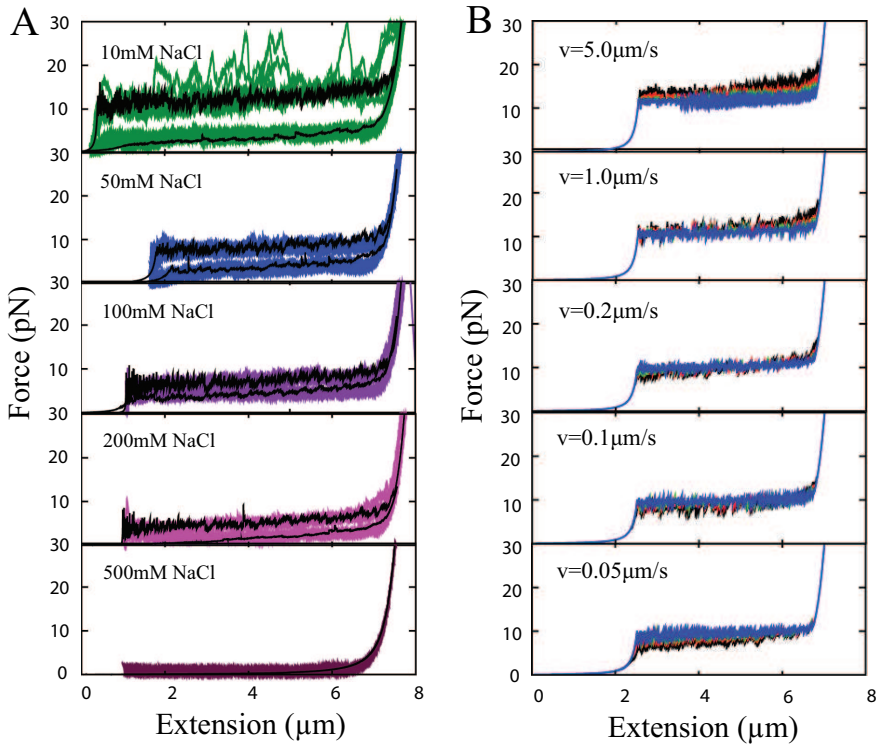

Figure 5.6: A. Experimental force curves for G5 at various ion concentrations with theoretical curves in black. Theoretical curves were calculated using ΔG_E 13.6 kcal/mol (10mM NaCl), 11.5 kcal/mol (50mM NaCl), 11 kcal/mol (100mM NaCl), 10 kcal/mol (200mM NaCl), and 7 kcal/mol (500mM NaCl). B. The effect of velocity of pulling on the force plateau for DNA-condensation by dendrimers. At the velocity used in experiments, $v = 0.2\mu m/s$, the several ΔG_E and x_E^* values yield virtually indistinguishable plateaus, but at different velocities the plateaus begin to separate. The curves shown are for $(\Delta G_E, x_E^*) = (20\ k_BT,\ 3\ \text{nm})$ black, $(25\ k_BT,\ 5\ \text{nm})$ red, $(30\ k_BT,\ 7\ \text{nm})$ green, and $(35\ k_BT,\ 9\ \text{nm})$ blue.

An intriguing finding stemming from the experiments and our microscopic simulations collectively, is that the upper plateau of the hysteretic curves for G5, G6 and G7 studied experimentally, as well as the G3 dendrimer in the MD-based MC modeling all seem, remarkably, to superimpose, despite their difference in size and overall charge (see Fig. 5.5 and 5.6 herein and in Fig. 2 in Ref. [12]). The only variables that affect the height of the plateau, as opposed to the length or EWLC fit in Eq. [5.1], are ΔG_E (plateau goes up as ΔG_E increases), x_E^* (plateau goes

down as x_E^* increases), and velocity v. This may be explained that the effect on the force-dependent rates of the fact that ΔG_E increases with larger dendrimer size is offset by the concomitant increase in x_E^*. To explore if this seeming "universality" is preserved across various pulling speeds, we used our MC model to gauge the effect of v on the force curve for DNA-condensation by dendrimers. As seen Fig. 5.6, at the velocity used in the actual experiments, one can fit several values of ΔG and x_E^* to give virtually indistinguishable FECs, but at different velocities the curves begin to separate out. Experiments done on the different dendrimer sizes at varying speeds could help elucidate the degree to which PMFs are similar (as implied by the single-molecule study) or the extent of ΔG_E and x_E^* increase with dendrimer size.

5.4 Concluding Discussion

Using all-atom molecular dynamics simulations and free energy calculations, we have performed an atomistic study of aspects concerning the structural changes and driving forces involved in DNA interacting with generation 3 PAMAM amine-terminated dendrimers from which we derived a mesoscopic model for the extensibility of dendrimer-condensed DNA. A free energy profile along the interaction coordinate was calculated, as well as the mean forces as a function of DNA-dendrimer separation. Using energy, force and geometry parameters computed at the atomic level, a Monte Carlo model for a mesoscopic force-extension curve was constructed that produced a force extension curve that reproduced to a high degree of accuracy the experimental single-molecule curves on DNA-dendrimer conglomerates.

Of the two orientations we studied, only the first orientation of the dendrimer induced a substantial structural deformation of the DNA locally, decreasing the end-to-end distance for the DNA by almost 15 Å and leading to an average bend angle

of 51.8°. On the other hand, the second orientation most likely serves as a bridge between adjacent DNA molecules. This would lead to what is usually referred to as DNA looping in the context of protein-DNA interactions. The sawtooth-shaped peaks visible in both the experimental and simulated force extension curves are similar in character to the sawtooth patterns observed experimentally in single-molecule force-extension measurements with known DNA looping proteins [25, 26] and serve as potential evidence for such looping in the DNA-dendrimer system.

Due to their similarity in chemistry and size to histone proteins, it has been suggested that DNA may actually wrap around high generation dendrimers. The results of the force pulling experiments and our ability to match them using data for a lower generation dendrimers imply that this is not in fact the case. A mathematical model for the interaction between charged flexible rods and charged spheres [27] predicts that the force extension curve for unwinding a single chain wrapped around a sphere will have a slope proportional to the inverse of the sphere radius, before reaching a maximum force which is also dependent on the sphere size. This model has recently been applied to the interactions between G4 dendrimers and DNA [28]. For the case of a large strand of DNA wrapped around multiple discrete sphere-like molecules, this model would predict a force curve with a sawtooth pattern and constantly increasing force. This is exactly what is seen in FECs of DNA condensed by histone proteins [29, 30, 31, 32, 33], which show a sharp sawtooth pattern with a constantly increasing force. The force extension curves provided by Ritort, however, have a relatively constant, flat plateau even for very high generation dendrimers. This type of FEC is in much better agreement with models in which DNA condensates have a toroidal structure. A theoretical model of this type of interaction predicts a nearly constant force plateau [34]. This is also in agreement with small-angle

x-ray scattering experiments on DNA-dendrimer complexation [35] and cryo-TEM studies [36, 37] which show DNA forming toroidal structures in the presence of low generation dendrimers, although cryo-TEM studies indicate that the morphology of the DNA-dendrimer complex is more globular for generation 5 and higher dendrimers [36]. The similarity of the experimental FECs to DNA-looping proteins rather than DNA-wrapping proteins would seem to indicate that these structures are due to the formation of loops rather than wrapping of the DNA molecule around dendrimers for larger generation dendrimers as well.

We have used several simplifications in our determination of the force-extension curve for this system. In principle, they may affect the results, so we discuss their validity. Recent work by Dudko et al. [38] have shown that the Bell model, which we use in these calculations, is not the most accurate description of the relationship between free energy and force. Their work shows that Kramers theory of diffusive barrier crossing can be used to model a more accurate relationship between free energy and force than the relatively simpler Bell model. The improvement is especially evident in free energy surfaces that resemble the cusp model. However, for linear-cubic free energy profiles and low to intermediate pulling rates the results of the Bell model are close to those of the more robust Kramers theory model. As the experimental velocity was in the intermediate range and our free energy surface is roughly cubic in shape, the Bell model should still give reasonable results for our system. For a free energy surface that is cusp-like, these types of calculations are better done with the Kramers theory method described in Dudko et al. Another potential problem with our calculations using this simple model is the assumption that while the applied force affects the barrier height, ΔG_E, it does not affect the barrier length, x. Close inspection of the plots in Figure 5.3 reveals that this is in

fact not true, as increasing force is applied to the system, the barrier of the free energy surface drifts to the left. This change, however, is minor, amounting to about 1-3 Å for the forces used in the experiment, and we therefore assume that our use of a force-independent x_E is a reasonable approximation that should not substantially affect our results. A more accurate calculation would take into account the effect of force on x_E or an analysis as in Dudko et al.

By changing the parameters in the MC model we revealed the microscopic origin of the hysteresis observed in the first-order phase transition between the extended and compacted DNA forms. Moreover, the broad range of ionic and pulling parameters that were sampled with the model can be used to offer suggestions for windows of conditions that can probe new single-molecule behavior in future experiments.

This work has been accepted for publication in Biophysical Journal

BIBLIOGRAPHY

[1] H. Deng, V. A. Bloomfield, J. M. Benevides, and G. J. Thomas. Structural basis of polyamine-DNA recognition: spermidine and spermine interactions with genomic B-DNAs of different GC content probed by raman spectroscopy. *Nucleic Acids Research*, 28:3379–3385, 2000.

[2] Y. Murayama, Y. Sakamaki, and M. Sano. Elastic response of single DNA molecules exhibits a reentrant collapsing transition. *Physical Review Letters*, 90:018102, 2003.

[3] P. G. Arscott, C. L. Ma, J. R. Wenner, and V. A. Bloomfield. DNA condensation by cobalt hexaammine(III) in alcohol-water mixtures - dielectric-constant and other solvent effects. *Biopolymers*, 36:345–364, 1995.

[4] H. Deng and V. A. Bloomfield. Structural effects of cobalt-amine compounds on DNA condensation. *Biophysical Journal*, 77:1556–1561, 1999.

[5] C. G. Baumann, V. A. Bloomfield, S. B. Smith, C. Bustamante, M. D. Wang, and S. M. Block. Stretching of single collapsed DNA molecules. *Biophysical Journal*, 78:1965–1978, 2000.

[6] P. G. Arscott, A. Z. Li, and V. A. Bloomfield. Condensation of DNA by trivalent cations .1. effects of DNA length and topology on the size and shape of condensed particles. *Biopolymers*, 30:619–630, 1990.

[7] Bloomfield V A. Theory of DNA condensation by multivalent cations. *Journal of Biomolecular Structure and Dynamics*, 8:A20, 1991.

[8] V. A. Bloomfield. Condensation of DNA by multivalent cations - considerations on mechanism. *Biopolymers*, 31:1471–1481, 1991.

[9] V. A. Bloomfield, C. L. Ma, and P. G. Arscott. Role of multivalent cations in condensation of DNA. *Macro-Ion Characterization*, 548:195–209, 1994.

[10] T. T. Nguyen, I. Rouzina, and B. I. Shklovskii. Reentrant condensation of DNA induced by multivalent counterions. *Journal of Chemical Physics*, 112:2562–2568, 2000.

[11] S. Q. He, P. G. Arscott, and V. A. Bloomfield. Condensation of DNA by multivalent cations: Experiment studies of condensation kinetics. *Biopolymers*, 53:329–341, 2000.

[12] F. Ritort, S. Mihardja, S. B. Smith, and C. Bustamante. Condensation transition in DNA-polyaminoamide dendrimer fibers studied using optical tweezers. *Physical Review Letters*, 96:118301, 2006.

[13] M. Rief, J. M. Fernandez, and H. E. Gaub. Elastically coupled two-level systems as a model for biopolymer extensibility. *Physical Review Letters*, 81:4764–4767, 1998.

[14] G. S. Manning. The persistence length of DNA is reached from the persistence length of its null isomer through an internal electrostatic stretching force. *Biophysical Journal*, 91:3607–3616, 2006.

[15] D. Porschke. Intramolecular collapse of DNA structure and dynamics. *Sarma, R. H. And M. H. Sarma (Ed.). Biomolecular Stereodynamics, Vol. 4; Fourth Conversation in The Discipline Biomolecular Stereodynamics, Albany, N.Y., Usa, June 4-8, 1985. 324p. Adenine Press: Guilderland, N.Y., Usa. Illus*, pages 11–20, 1986.

[16] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante. Ionic effects on the elasticity of single DNA molecules. *Proceedings of The National Academy of Sciences of The United States of America*, 94:6185–6190, 1997.

[17] G. I. Bell. Models for specific adhesion of cells to cells. *Science*, 200:618–627, 1978.

[18] E. Evans. Probing the relation between force - lifetime - and chemistry in single molecular bonds. *Annual Review of Biophysics and Biomolecular Structure*, 30:105–128, 2001.

[19] K. Besteman, S. Hage, N. H. Dekker, and S. G. Lemay. Role of tension and twist in single-molecule DNA condensation. *Physical Review Letters*, 98:058103, 2007.

[20] W. Chen, N. J. Turro, and D. A. Tomalia. Using ethidium bromide to probe the interactions between DNA and dendrimers. *Langmuir*, 16:15–19, 2000.

[21] J.G. Kirkwood. Statistical mechanics of fluid mixtures. *Journal of Chemical Physics*, 3:300–313, 1935.

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

[23] R. Zhang and B. I. Shklovskii. The pulling force of a single DNA molecule condensed by spermidine. *Physica A-Statistical Mechanics and Its Applications*, 349:563–570, 2005.

[24] T. Lundback and T. Hard. Salt dependence of the free energy, enthalpy, and entropy of nonsequence specific DNA binding. *Journal of Physical Chemistry*, 100:17690–17695, 1996.

[25] D. N. Fuller, G. J. Gemmen, J. P. Rickgauer, A. Dupont, R. Millin, P. Recouvreux, and D. E. Smith. A general method for manipulating DNA sequences from any organism with optical tweezers. *Nucleic Acids Research*, 34:e15, 2006.

[26] G. J. Gemmen, R. Millin, and D. E. Smith. Dynamics of single DNA looping and cleavage by sau3ai and effect of tension applied to the DNA. *Biophysical Journal*, 91:4154–4165, 2006.

[27] H. Schiessel, R. F. Bruinsma, and W. M. Gelbart. Electrostatic complexation of spheres and chains under elastic stress. *Journal of Chemical Physics*, 115:7245–7252, 2001.

[28] K. Qamhieh, T. Nylander, and M. L. Ainalem. Analytical model study of dendrimer/DNA complexes. *Biomacromolecules*, 10:1720–1726, 2009.

[29] B. D. Brower-Toland, C. L. Smith, R. C. Yeh, J. T. Lis, C. L. Peterson, and M. D. Wang. Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA. *Proceedings of The National Academy of Sciences of The United States of America*, 99:1960–1965, 2002.

[30] G. J. Gemmen, R. Sim, K. A. Haushalter, P. C. Ke, J. T. Kadonaga, and D. E. Smith. Forced unraveling of nucleosomes assembled on heterogeneous DNA using core, histones, NAP-1, and ACF. *Journal of Molecular Biology*, 351:89–99, 2005.

[31] G. V. Soni, L. Brar, F. M. Hameed, A. K. Raychaudhuri, and G. V. Shivashankar. Distinct levels in the nanoscale organization of DNA-histone complex revealed by its mechanical unfolding. *Applied Physics Letters*, 90:163904, 2007.

[32] S. Y. Ran, X. L. Wang, W. B. Fu, W. C. Wang, and M. Li. Single molecule DNA compaction by purified histones. *Chinese Science Bulletin*, 53:836–841, 2008.

[33] Y. Cui and C. Bustamante. Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure. *Proceedings of The National Academy of Sciences of The United States of America*, 97:127–132, 2000.

[34] C. Battle, B. van den Broek, M. C. Noom, J. van Mameren, G. J. L. Wuite, and F. C. MacKintosh. Unraveling DNA tori under tension. *Physical Review E*, 80:031917, 2009.

[35] C. J. Su, H. L. Chen, M. C. Wei, S. F. Peng, H. W. Sung, and V. A. Ivanov. Columnar mesophases of the complexes of in DNA with low-generation poly(amidoamine) dendrimers. *Biomacromolecules*, 10:773–783, 2009.

[36] M. L. Ainalem, A. M. Carnerup, J. Janiak, V. Alfredsson, T. Nylander, and K. Schillen. Condensing DNA with poly(amidoamine) dendrimers of different generations: means of controlling aggregate morphology. *Soft Matter*, 5:2310–2320, 2009.

[37] A. M. Carnerup, M. L. Ainalem, V. Alfredsson, and T. Nylander. Watching DNA condensation induced by poly(amidoamine) dendrimers with time-resolved cryo-TEM. *Langmuir*, 25:12466–12470, 2009.

[38] O. K. Dudko, G. Hummer, and A. Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Physical Review Letters*, 96:108101, 2006.

CHAPTER VI

Conclusions

The studies presented here demonstrate several ways in which experiments and simulations can be used effectively to complement and enhance each other. Knowledge of experimental results was used to improve both the efficiency and accuracy of simulations by using experimental potentials of mean force to guide umbrella sampling simulations and by adapting steered molecular dynamics simulations for a system in which the high velocities used in simulations was expected to affect the pathway. In the latter case, the simulations revealed the source of heterogeneity that could not be explained with the experimental results alone. Additionally, an umbrella sampling study of DNA complexation with PAMAM dendrimers was able to provide chemical, structural, and energetic information of the interaction between the molecules at a level of detail far beyond that which is accessible experimentally. The results of these simulations were then used in a Monte Carlo calculation that exploited the relationship between force and free energy to calculate force extension curves analogous to experiment.

In Chapter 2, a method was described for using potentials of mean force from experiments to improve umbrella sampling simulations. Knowing the PMF *a priori* allows one to choose optimal restraining potentials. Comparison of guided umbrella

sampling simulations, to unguided, equal-constraint umbrella sampling simulations for two different systems showed that the guided method is more effective and requires fewer windows than the unguided one. For the pentapeptide case, guided umbrella sampling not only reproduced the PMF of the restrained reaction coordinate, but also resulted in accurate PMFs of secondary reaction coordinates. Similarly, guided umbrella sampling on the forced unfolding of titin I27 not only produced a free energy profile in good agreement with experimental data, but also reproduced the unfolding pathway seen in previous simulations and predicted by experiments.

In cases where an experimental PMF is difficult to obtain, it is nevertheless possible to use experimental information when designing simulations. Forced unfolding experiments on the protein domain described in Chapter 3 indicated that the system potentially unfolding via three pathways, the distribution of which depended on the pulling rate. Steered MD simulations were run at an elevated temperature in order to bypass this velocity effect. The results of these high-temperature simulations showed that the domain does indeed follow three unfolding pathways. The differences between these pathways was found to be due to hydrophobic interactions within the core of the domain, indicating that hydrophobic effects may have a more substantial effect on the stability of mechanical proteins than previously thought. Additionally, multidimensional potentials of mean force were calculated to reveal the free energy surface of coordinates relevant to the different pathways.

Free energy calculations of a different type of system, the interaction between DNA and a charged nanoparticle, a polyamidoamine dendrimer, are detailed in Chapter 4. The total free energies of the interaction between DNA and generation 3 dendrimers with different charges and orientations, as well as individual contributions to the free energy, were calculated from umbrella sampling simulations. Most interestingly, it

was found that ordered waters contributed significantly to the free energy for a dendrimer with 32 amine terminations and its widest side facing the DNA. These water effects are described using a phenomenological model that has been used to explain experimental observations of DNA condensation by multivalent cations. As well as the energetic analysis, structural changes in both the DNA strand and dendrimer upon complexation were characterized and compared to experiment, and the nature of the hydrogen bonding interactions was found to be dominated by interactions between the DNA phosphates and dendrimer amines.

The results of the simulations described in Chapter 4 were used to replicate experimental data using a Monte Carlo simulation of force in Chapter 5. Force pulling experiments on DNA condensed by dendrimers provided force extension curves which could be directly compared to our simulations. The free energy data obtained for the interaction between DNA and amine-terminated G3 dendrimers was used with the Bell-Evans model to calculate force extension curves which closely matched the experimental results.

Neither simulations nor experiments can be expected to provide both detailed and accurate results when used by themselves. The combination of experimental and theoretical techniques is more powerful than either method alone. In designing and interpreting experiments of any type, it is important to consider and correct for limitations in the chosen method. The work in this dissertation has aimed to demonstrate ways in which experimental evidence can be incorporated into molecular dynamics simulations, as well as to show how simulations can augment the information revealed experimentally.

APPENDICES

APPENDIX A

Fortran Codes for Calculating Potentials of Mean Force from Force Extension Curves

To calculate PMFs from force vs time and extension data, the following codes were used to implement the equations described in chapter 3, equation 3.2. First the work was calculated from a file containing time, extension, and force. In the case of multidimensional PMFS, a second file with the time and second reaction coordinate is needed. Once the work is caluclated, the work files are combined and the data is binned as a function of time. For the multimdimensional case, the data in each time bin is also binned by the secondary reaction coordinate to give a two-dimensional array of work values. The Jarzynski equation is then used to determine the PMF.

The implementation of the Jarzynski equality uses the method described by Hummer and Szabo in PNAS [1]. This method uses an iterative process to improve the unbiasing of the probability distribution as in WHAM [2]. An initial calculation of the PMF is made using the formula

$$G_0(x) = -\beta ln \frac{\sum_t \frac{\langle \delta(x-x_t) e^{-\beta W_t} \rangle}{\langle e^{-BW_t} \rangle}}{\sum_t \frac{e^{[-\beta U(x,t)]}}{\langle e^{-\beta W_t} \rangle}} \tag{A.1}$$

.

The value $\langle e^{-\beta W_t} \rangle$ is replaced with $e^{-\beta[U(x,t)+G_0(x)} dx / e^{-\beta G_0(x)}$. The equations are then solved iteratively unitl they reach convergence.

A.1 Code for calculation of work as a function of time and position

file name work.f

```
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
! !
! Calculate Work from force extension curves by intergrating over force !
! will combine this data for all values, bin, and calculate pmf !
! W=int(F*dx) !
! correction for deflection, needed for PMF calcs: U(x,t)=k/2(x-vt)^2!
! DeltaW=W(t)-U(x,t) !
! !
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
parameter ( n=10000000)
parameter (m=4500)
DIMENSION ncount(n)
double precision work(n),U(n),U0,x_skew(n)
real time(n),a(n),x(n),force(n),x2(n)
real vel,dt,k,kB,Temp,B,f
Character mdfile*30,filein*30,filein2*30,fileout*30
Integer I,J,nhigh,nfile
! open metadata file with list of force files, output files, temp, k
write(*,*)'input mdfile'
read(*,*)mdfile
open(unit=98,file=mdfile,status='old',form='formatted')
5 read(98,*,END=400)filein,fileout,Temp,k
```

```
! open force and work files, get names from mdfile
OPEN(UNIT=2,FILE=filein,STATUS='OLD',FORM='FORMATTED')
OPEN(UNIT=3,FILE=fileout,STATUS='new')
nhigh = 0
! read in data
15 READ(2,*,END=200)xtemp,ytemp,ztemp,ftemp
if (xtemp .eq. 0.0000) goto 15
nhigh=nhigh+1
x_skew(nhigh) = xtemp
force(nhigh)=ftemp
goto 15
200 Close(2)
do i=1,nhigh
x(i)=x_skew(i) - x_skew(1)!shiftxpositionosx = xskew - x_equilib.
time(i)=(i-1)*0.2 ! calculate time in ps
enddo
! set constants
! Temperature and spring constant are taken from mdfile
kB=0.01986
B=1./(kB*Temp)
vel=0.1 !A/ps
dt=0.2
! calculate work
f=force(1)
work(1)=0.000
```

```
do i=2,nhigh

f=f+force(i)*(x(i)-x(i-1)) ! SUM(F*delx)

work(i)=f/69.475 ! 69.745 pN/A=1kcal/mol

enddo

! calculate pulling potential U(x,t)

do i=1,nhigh

x2(i)=(x(i)-vel*time(i))*(x(i)-vel*time(i))

U(i)=0.5*k*x2(i)

enddo

! write out work

DO 40 J=1,nhigh

write(3,50)time(j),x(j),work(j),U(j)

40 CONTINUE

50 FORMAT(2F10.3,2d25.15)

CLOSE(3)

goto 5 !return to mdfile for next set of data

400 close(98)

END
```

A.2 Code for calculation of work as a function of time, position, and a secondary reaction coordinate

file name work2d.f

```
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

! !

! Calculate Work from force extension curves !

! will combine this data for all values, bin, and calculate pmf !
```

```
! W=int(F*dx) !
! correct for deflection, U(x,t)=k/2(x-vt)^2!
! DeltaW=W(t)-U(x,t) !
! !
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
parameter ( n=10000000)
parameter (m=4500)
DIMENSION ncount(n)
double precision work(n),U(n),U0
real time(n),a(n),x(n),force(n),x2(n),sasa(n)
real vel,dt,k,kB,Temp,B,f,xinit(n),forceinit(n)
Character mdfile*30,filein*30,filein2*30,fileout*30
Integer I,J,nhigh,ninit
! open metadata file with list of force files, output files, temp, k
write(*,*)'input mdfile'
read(*,*)mdfile
open(unit=98,file=mdfile,status='old',form='formatted')
5 read(98,*,END=400)filein,filein2,fileout,Temp,k
OPEN(UNIT=2,FILE=filein,STATUS='OLD',FORM='FORMATTED')
open(Unit=4,FILE=filein2,STATUS='OLD',FORM='FORMATTED')
OPEN(UNIT=3,FILE=fileout,STATUS='new')
nhigh = 0
ninit = 0
! read in data
15 READ(2,*,END=200)xtemp,ytemp,ztemp,ftemp
```

```
if (xtemp .eq. 0.0000) goto 15 !skip over misplaced zeros
if (ninit .gt. 22300) goto 200 !ma file length, otherwise will crash
ninit=ninit+1
xinit(ninit)=xtemp
forceinit(ninit)=ftemp
goto 15
200 Close(2)
do i=1,ninit,5
nhigh=nhigh+1
enddo
do j=1,nhigh
x(j)=xinit(j+4*(j-1))-xinit(1) !skip every 5 data points, since sasa files
force(j)=forceinit(j+4*(j-1)) !are every ps, rather than .2ps
enddo
do i=1,nhigh
read(4,*)temp,sasa(i) ! read sasa file
enddo
do i=1,nhigh
time(i)=(i-1)*0.2*5 ! calculate time in ps
enddo
! set constants
kB=0.01986
B=1./(kB*Temp)
vel=0.1 !A/ps
dt=0.2
```

```
! calculate work
f=force(1)
work(1)=0.000
do i=2,nhigh
f=f+force(i)*(x(i)-x(i-1))
work(i)=f/69.475 ! 69.745 pN/A=1kcal/mol
enddo
! calculate pulling potential
do i=1,nhigh
x2(i)=(x(i)-vel*time(i))*(x(i)-vel*time(i))
U(i)=0.5*k*x2(i)
enddo
! write out work
DO 40 J=1,nhigh
write(3,50)time(j),x(j),sasa(j),work(j),U(j)
40 CONTINUE
50 FORMAT(3F10.3,2d25.15)
CLOSE(3)
goto 5
400 close(98)
END
```

A.3 Code for implementation of Jarzynski equality

file name jarz.f

```
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
```

```
!
! calculate free energy from Force extension curves
! using Jarzynski equlity and 2nd order calculation of work
! as described in Park and Schulten, J. Chem. Phys. 120, p5946-5961
!
! equations used:
!
!  PMF(t)=¡W(t)¿-B/2*(¡W(t)^2) > - < W(t) >^2)fromexpansionoflog < e^( -
BW) >
!
! ¡W(t)¿=INT(v¡F¿dt)-k/2(¡x¿-vt)^2
!
! input files for this program are generated from SMD output using work.f
!
parameter ( n=10000000)
parameter (m=4500)
DIMENSION ncount(n)
double precision rcd(n),work_avg(n), w2(n), PMF(n)
double precision work2(n),wbin(m),w2bin(m),bin(m)
double precision time(n),a(m),x(n),force(n)
double precision vel,k,kB,Temp,B,xbin(m),fbin(m)
double precision expw(n),expbin(m),altpmf(m),x2(m)
Character filein*25,filedat*25,fileout*25
Integer I,J,nhigh
write(*,*)'input filein name and fileout name'
```

```
read(*,*)filein,fileout
write(*,*)'input temperature'
read(*,*)Temp
OPEN(UNIT=2,FILE=filein,STATUS='OLD',FORM='FORMATTED')
OPEN(UNIT=3,FILE=fileout,STATUS='new')
nhigh = 0
! read in data
15 READ(2,*,END=200)ttemp,xtemp,wtemp,utemp
nhigh=nhigh+1
time(nhigh)=ttemp
x(nhigh)=xtemp !rxn coordinate
! force(nhigh)=ftemp
work2(nhigh)=wtemp*wtemp !(Work-potential of spring)^2
expw(nhigh)=exp(-0.84051*wtemp/10000.) !exp are too small for memory, division by 100 is to bypass this problem, will correct this value in final calculation of pmf. This is a very crude way of doing this, but these values are for comparison only.
goto 15
200 Close(2)
! set up intial values for bins
do j=1,m
a(j)=0 !no. of values in bin
bin(j)=0.0 !time bin
xbin(j)=0.0 !rxn coordinate bin
fbin(j)=0.0 !force bin
w2bin(j)=0.0 !work^2bin
```

```
expbin(j)=0.0 !e^(-BW)bin
enddo
! set constants
! Temp=350.
kB=0.0019829 !Boltzmann constant, kcal/mol/K
B=1./(kB*Temp) !Beta=1/kBT
vel=0.1 !A/ps
k=5.95 !kcal/mol/A^2
dt=0.2 !ps
! bin data based on time
bin(1)=0.000
do i=2,m
bin(i)=bin(i-1.)+0.1000
enddo
Do 30 J=1,nhigh
Do i=1,m
if (abs(x(J)-bin(i)) .le. 0.04) then
! xbin(i)=xbin(i)+x(j)
fbin(i)=fbin(i)+force(j)
w2bin(i)=w2bin(i)+work2(j)
expbin(i)=expbin(i)+expw(j)
a(i)=a(i)+1.00
endif
enddo
30 continue
```

! normalize bins

do i=1,m

!xbin(i)=xbin(i)/a(i)

fbin(i)=fbin(i)/a(i)

w2bin(i)=w2bin(i)/a(i)

expbin(i)=expbin(i)/a(i)

enddo

! calc ¡W(t)¿ and ¡W(t)^2 > − < $W(t)$ >^{2}

f=0.0

do i=1,m

f=f+fbin(i) !calculate average force

work$_a vg(i) = (f * vel * dt * 5)/69.479!69.5 is to convert from pN to kcal/mol, * 5 to compensate for s$

$if bin size is changed this must be too$

x2(i)=(xbin(i)-vel*bin(i))*(xbin(i)-vel*bin(i))

work$_a vg(i) = work_a vg(i) - 0.5 * k * x2(i)!W - U$

w2(i)=w2bin(i)-work$_a vg(i) * work_a vg(i)$

! write(*,*)work$_a vg(i) * work_a vg(i)$

enddo

! calc PMF

do i=1,m

pmf(i)=work$_a vg(i) - (B/2.) * w2(i)$

enddo

! alternative method calculate log¡e$^(-BW) > directly - for comparison$

do i=1,m

altpmf(i)=(-10000./B)*log(expbin(i)) !*100 to compensate for earlier step

```
enddo
! write out PMF - time rxn coord altpmf pmf
DO 40 J=1,m
write(3,50)bin(j),work_avg(j),altpmf(j)
40 CONTINUE
50 FORMAT(2F10.3,1d25.15)
CLOSE(3)
END
```

A.4 Code for implementation of Jarzynski equality for 2D PMF

jarz2d.f

```
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
! !
! calculate free energy from Force extension curves !
! using Jarzynski equlity and modification of WHAM method from !
! Hummer and Szabo PNAS 98, p 3658-3661 !
! Modification for 2Dimensional Free energy surfaces !
! !
! !
! equations used: !
! !
! SUM_t[< d(x - xt)e^-BW(phi|x,t) > / < e^-BW >]!
! PMF(phi—x)=G(phi—x)=-1/B*ln(


                                              )!
! SUM_t[e - BU(phi|x,t)/ < e^-BW >]!
```

```
! !
! !
! this equation is made self-consistent as in WHAM by replacing ¡e^{-BW}>!
! with e-B*deltaG(t)=INT(e^{[}-B(U(phi|x,t)-G(phi|x))]dx!
! !
! !
! e^{-}B*W(phi|x,t)iscalculatedbyfirstbinninge^{-}BWbyx,thenforeach!
! x, binning e^{-}BWalongphitogenerateahistgramofvaluesforeachxbin!
! !
! !
! input files for this program are generated from SMD output using work2d.f !
! !
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
parameter (n=10000000)
parameter (m=250)
parameter (m2=300)
Character mdfile*25,filein*25,filedat*25,fileout*25
double precision Temp,bmin1,bmin2,bmax1,bmax2,delbin1
double precision delbin2,nbin1,nbin2
DIMENSION ncount(n)
double precision G(m,m2),Gnew(m,m2)
double precision expw(n),expu(n),phi(n),bin(m)
double precision time(n),a(m),x(n),U(n),Uz(m,m)
double precision hzp(m,m,m2),uzp(m,m,m2),azp(m,m,m2)
double precision vel,k0,kB,B,xbin(m)
```

```
double precision h(m,m),nt(m),ntnew(m),a2(m,m)
double precision ut(m,m2),ht(m,m2),hnorm(m,m,m2),unorm(m,m,m2)
double precision ntop(m),nbot(m),pbin(m2),G1(m),ag(m),minf
Integer I,J,L,K,nhigh
! input file with parameters mdfile should contain one line with:
! Name of work file, Name of output file, Temperature, minimum value, maximum value, number of bins for the 1st rxn coord, then min value, max value, number of bins for the second reaction coordinate
write(*,*)'input filein name'
read(*,*) mdfile
open(unit=98,File=mdfile,status='old',Form='formatted')
read(98,*)filein,fileout,Temp,bmin1,bmax1,nbin1,bmin2,bmax2,nbin2
delbin1=(bmax1-bmin1)/nbin1
delbin2=(bmax2-bmin2)/nbin2
OPEN(UNIT=2,FILE=filein,STATUS='OLD',FORM='FORMATTED')
OPEN(UNIT=3,FILE=fileout,STATUS='new')
nhigh = 0
! set constants
kB=0.0019829 !Boltzmann constant, kcal/mol/K
B=1./(kB*Temp) !Beta=1/kBT
vel=0.1 !A/ps
k0=5.95 !kcal/mol/A^2
dt=0.2 !ps
! read in data
15 READ(2,*,END=200)ttemp,xtemp,phitemp,wtemp,utemp
```

```
nhigh=nhigh+1
time(nhigh)=ttemp
x(nhigh)=xtemp !rxn coordinate
phi(nhigh)=phitemp
expw(nhigh)=exp(-1.*B*wtemp/1000000.) ! work(t)
expu(nhigh)=exp(-1.*B*Utemp/1000000.) !U(x,t)
goto 15
200 Close(2)
! set up intial values for bins
do j=1,nbin1
a(j)=0.00 !no. of values in bin
bin(j)=0.0 !time bin xbin(j)=0.0 !rxn coordinate bin
nt(j)=0.0 !e^( - BW)bin
enddo
! bin data based on time
bin(1)=bmin1
do i=2,nbin1
bin(i)=bin(i-1.)+ delbin1
enddo
xbin(1)=bmin1/10.
do i=2,nbin1
xbin(i)=xbin(i-1)+delbin1*vel
enddo
! calc ¡e^-BW >
Do 30 J=1,nhigh
```

```
Do i=1,nbin1
if (abs(time(J)-bin(i)) .le. (delbin1/2.)) then
nt(i)=nt(i)+expw(j)
a(i)=a(i)+1.00
endif
enddo
30 continue
! normalize bins
do i=1,nbin1
nt(i)=nt(i)/a(i) !¡e^-BWt >
enddo
! define U(z,t),e^W(z,t)
! set variables
do i=1,nbin1
do j=1,nbin1
uz(i,j)=(0.0,0.0)
h(i,j)=(0.0,0.0)
a2(i,j)=(0.0,0.0)
enddo
enddo
do i=1,nbin1
do j=1,nbin1
do k=1,nbin2
uzp(i,j,k)=(0.0,0.0) !0.0)
hzp(i,j,k)=(0.0,0.0) !0.0)
```

```
azp(i,j,k)=(0.0,0.0) ! 0.0)
enddo
enddo
enddo
pbin(1)=bmin2
do i=2,nbin2
pbin(i)=pbin(i-1)+delbin2
enddo
do i=1,nhigh
do j=1,nbin1
if (abs(time(i)-bin(j)) .le. (delbin1/2.)) then
do k=1,nbin1
if(abs(x(i)-xbin(j)) .le. (delbin1*vel/2))then
Uz(j,k)=Uz(j,k)+expu(i)
h(j,k)=h(j,k)+expw(i)
a2(j,k)=a2(j,k)+1.000
endif
do l=1,nbin2
if(abs(phi(i)-pbin(l)) .le. (delbin2/2.)) then
uzp(j,k,l)=uzp(j,k,l) + expu(i)
hzp(j,k,l)=hzp(j,k,l) + expw(i)
azp(j,k,l)=azp(j,k,l) + 1.000
endif
enddo
enddo
```

```
endif
enddo
enddo
! h(l)
do i=1,nbin1
do j=1,nbin1
Uz(i,j)=Uz(i,j)/a2(i,j)
h(i,j)=h(i,j)/a2(i,j)
enddo
enddo
do i=1,nbin1
do j=1,nbin1
do k=1,nbin2
uzp(i,j,k)=uzp(i,j,k)/azp(i,j,k)
hzp(i,j,k)=hzp(i,j,k)/azp(i,j,k)
enddo
enddo
enddo
! now divide by ¡e^-BW(t) >
do i=1,nbin1
do j=1,nbin1
do k=1,nbin2
hnorm(i,j,k)=hzp(i,j,k)/nt(i) ! ¡d(x-xt)e^-BW(z,t) > / < e^-BW(t) >
unorm(i,j,k)=uzp(i,j,k)/nt(i) ! [e^-B * u(z,t)]/ < e^-BW(t) >
enddo
```

```
enddo

enddo

! sum over time slices

do i=1,nbin1

do j=1,nbin2

ht(i,j)=(0.000,0.00)

ut(i,j)=(0.000,0.00)

enddo

enddo

do i=1,nbin1

do j=1,nbin1

do k=1,nbin2

ht(i,k)=ht(i,k)+hnorm(i,j,k)

ut(i,k)=ut(i,k)+unorm(i,j,k)

enddo

enddo

enddo

!calc initial PMF

do i=1,nbin1

do j=1,nbin2

G(i,j)=(-1/B)*1000000*log(ht(i,j)/ut(i,j))

enddo

enddo

! start iterative loop

! replace ¡e^W(t) > withe^-B(u + G)
```

```
40 do i=1,nbin1
G1(i)=0.000000
ag(i)=0.000000
ntop(i)=0.000000
nbot(i)=0.000000
enddo
do i=1,nbin1
do 41 j=1,nbin2
if(G(i,j) .ne. G(i,j)) then ! first determine if G(i,j) is nan
goto 41 ! if NAN, G will not equal itself
endif ! if not then sum over j
G1(i)=G1(i)+G(i,j)
ag(i)=ag(i)+1.00
!write(*,*)G1(i) 41 continue
enddo
do i=1,nbin1
G1(i)=G1(i)/ag(i)
enddo
do i=1,nbin1
do j=1,nbin1
ntop(i)=ntop(i)+exp(-1.*B*(Uz(i,j)+G1(i))/1000000.)
nbot(i)=nbot(i)+exp(-1.*B*G1(i)/1000000.)
!write(*,*)ntop(i),nbot(i)
enddo
enddo
```

```
do i=1,nbin1
ntnew(i)=ntop(i)/nbot(i)
!write(*,*)ntnew(i)
enddo
! recalc h and u w/new nt
do i=1,nbin1
do j=1,nbin1
do k=1,nbin2
hnorm(i,j,k)=hzp(i,j,k)/ntnew(i) ! ¡d(x-xt)e^-BW(z,t) > / < e^-BW(t) >
unorm(i,j,k)=uzp(i,j,k)/ntnew(i) ! [e^-B * u(z,t)]/ < e^-BW(t) >
!write(*,*)hnorm(i,j,k),unorm(i,j,k)
enddo
enddo
enddo
! sum over time slices
do i=1,nbin1 do j=1,nbin2
ht(i,j)=(0.00,0.00)
ut(i,j)=(0.00,0.00)
enddo
enddo
do i=1,nbin1
do j=1,nbin1
do k=1,nbin2
ht(i,k)=ht(i,k)+hnorm(i,j,k)
ut(i,k)=ut(i,k)+unorm(i,j,k)
```

```
enddo
enddo
enddo
! calculate new G(x)
do i=1,nbin1
do j=1,nbin2
Gnew(i,j)=(-1/B)*1000000*log(ht(i,j)/ut(i,j))
enddo
enddo
! test for convergence
do i=1,nbin1
do 42 j=1,nbin2
if(Gnew(i,j) .ne. Gnew(i,j)) then ! can't compare NANs so skip them
goto 42
endif
if(G(i,j) .ne. G(i,j)) then
goto 42
endif
if(abs(Gnew(i,j)-G(i,j)) .gt. 0.1) then
do k=1,nbin1
do l=1,nbin2
G(k,l)=Gnew(k,l) ! replace old G with new one
enddo
enddo
goto 40 ! go to recalc
```

```
endif
42 continue
enddo
minf=0.000
do i=1,nbin1
do j=1,nbin2
if (Gnew(i,j) .lt. minf) then
minf=Gnew(i,j)
endif
enddo
enddo
! write out PMF - time rxn coord altpmf pmf
DO 45 J=1,nbin1
write(3,50)
do k=1,nbin2
write(3,50)xbin(j),pbin(k),Gnew(j,k)-minf
enddo
45 CONTINUE
50 FORMAT(2F10.3,2d25.15)
CLOSE(3)
END
```

BIBLIOGRAPHY

[1] G. Hummer and A. Szabo. Free energy surfaces from single-molecule force spectroscopy. *Accounts of Chemical Research*, 38:504–513, 2005.

[2] 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 .1. the method. *Journal of Computational Chemistry*, 13:1011–1021, 1992.

APPENDIX B

Monte Carlo Fortran Code for calculating FECs from free energy data

The monte carlo simulations of Force extension curves from free energy data usign the Bell-Evans model were calculated using the following fortran code. The barrier height and length of the free energy curve are used to calculate the rates of contact formation and breaking, which is then used to determine the probabilities. An acceptance/rejection scheme based on these probabilities and a random number generator is enacted at each time step, and a new force is calculated based on the results. The force itself is calulated using the extensible worm-like chain model. The force is solved iteratively until convergence at each time step, with an initial guess force based on the previous step.

B.1 MC code

force$_m c.f$

! Monte Carlo Simulation of FECs from Free Energy Barrier height and length ! using the Bell-Evans model !

program force calculation

character file1*50,file2*50,fileout*50,temp*15

parameter (m=35300)

```
double precision f(m),force(m),ktn(m)
real kbT,x(m),N,Nf(m),Nu(m),A,d,dT,dx,r(m),ddx(m),test(m)
real alpha, beta,aF(m),bF(m),t(m),v,Pu(m),Pf(m),L(m),r2(m)
real theta(m)
integer j,i,idum
write(*,*)'input fileout name'
read(*,*)fileout
open(UNIT=3,FILE=fileout,STATUS='NEW')
t(1)=0.00000
x(1)=0.00000
do i=2,m
t(i)=t(i-1)+0.00100
enddo
L(1)=2.7 ! initial slack value
dT=0.00100 !timestep
dx=0.2000*dT !dx = velocity*timestep
Nf(1)=300. ! number of intact contacts
Nu(1)=0. ! number of broken contacts
! set constants
kbT=0.00411
A=kbT/0.020
d=2000.
alpha=exp(-0.0850/kbT) ! rate=exp(deltaG_contact_formation/kbT)
beta=exp(-0.08/kbT) ! rate=exp(deltaG_contact_break/kbT)
idum=1
```

```
do i=1,m
theta(i)=0.+0.07*expdev(idum) ! randomize angle of force for noise
enddo ! write(*,*) alpha,beta
! calculate force from ewlc
! initialize
force(1)=A*(1/(4*(1-x(1)/L(1))**2)-0.25+x(1)/L(1))
f(1)=force(1)
do i=2,m
Nf(i)=Nf(i-1)
Nu(i)=Nu(i-1)
x(i)=x(i-1) + dx
L(i)=L(i-1)
do j=1,100
f(j+1)=A*(1/(4*(1+f(j)/d-x(i)/L(i))**2)-0.25+x(i)/L(i)) !EWLC
! write(*,*) f(j+1)
if (abs(f(j+1)-f(j)) .lt. 0.005) then ! solv iteratively
force(i)=f(j+1)
goto 30
endif
30 enddo
! calculate alpha(force) and beta(force)
aF(i)=alpha*exp(cos(theta(i))*force(i)*0.00315/kbT)
bF(i)=beta*exp(-1*force(i)*0.003/kbT)
! write(*,*) aF(i)
! calculate probabilities
```

```
Pu(i)=Nf(i)*aF(i)*dT*35300. ! prob of contact breaking
Pf(i)=Nu(i)*bF(i)*dT*35300. ! prob of contact forming
! write(*,*) Pf(i),Pu(i)
! generate random number to determine whether break was made
! if contact is broken, contour length is increased: L(i)=L(i)+0.015
r(i)= ran3(idum)
if (r(i) .lt. Pu(i)) then
L(i)=L(i)+0.015
! write(*,*)x(i+1)
Nf(i)=Nf(i)-1
Nu(i)=Nu(i)+1
! write(*,*)Nf(i),Nu(i)
endif
r2(i)=ran3(idum)
if (r(i) .lt. Pf(i)) then
L(i)=L(i)-0.15
Nf(i)=Nf(i)+1
Nu(i)=Nu(i)-1
endif
35 enddo
do 40 j=1,m
write(3,50)x(j),force(j)
40 continue
50 format(2F15.5)
close(3)
```

```
    END
    ! Functions expdev, ran1, and ran3 are (C) Copr. 1986-1992 Numerical Recipes
Software
    FUNCTION expdev(idum)
    integer idum
    real expdev
    c uses ran1
    Real dum,ran1
    1 dum=ran1(idum)
    if(dum .eq. 0.) goto 1
    expdev=-log(dum*0.07)
    return
    END
    FUNCTION ran1(idum)
    integer idum,ia,im,iq,ir,ntab,ndiv
    real ran1,am,eps,rnmx
    parameter (ia=16807,im=2147483647,am=1./im,iq=127773)
    parameter (ir=28365,ntab=32,ndiv=1+(im-1)/ntab)
    parameter (eps=1.2e-7,rnmx=1.-eps)
    integer j,k,iv(ntab),iy
    save iv,iy
    data iv /ntab*0/, iy /0/
    if (idum.le.0 .or. iy .eq. 0) then
    idum=max(-idum,1)
    do j=ntab+8,1,-1
```

```
k=idum/iq
idum=ia*(idum-k*iq)-ir*k
if (idum .lt. 0) idum=idum+im
if (j .le. ntab) iv(j)=idum
enddo
iy=iv(1)
endif
k=idum/iq
idum=ia*(idum-k*iq)-ir*k
if (idum .lt. 0) idum=idum+im
j=1+iy/ndiv
iy=iv(j)
iv(j)=idum
ran1=min(am*iy,rnmx)
return
end
FUNCTION ran3(idum)
INTEGER idum
INTEGER MBIG,MSEED,MZ
C REAL MBIG,MSEED,MZ
REAL ran3,FAC
PARAMETER (MBIG=1000000000,MSEED=161803398,MZ=0,FAC=1./MBIG)
c PARAMETER (MBIG=4000000,MSEED=1618033,MZ=0,FAC=1/MBIG)
INTEGER i,iff,ii,inext,inextp,k
INTEGER mj,mk,ma(55)
```

```
C REAL mj,mk,ma(55)
SAVE iff,inext,inextp,ma
DATA iff /0/
if(idum.lt.0.or.iff.eq.0)then
iff=1
mj=MSEED-iabs(idum)
mj=mod(mj,MBIG)
ma(55)=mj
mk=1
do 11 i=1,54
ii=mod(21*i,55)
ma(ii)=mk
mk=mj-mk
if(mk.lt.MZ)mk=mk+MBIG
mj=ma(ii)
11 continue
do 13 k=1,4
do 12 i=1,55
ma(i)=ma(i)-ma(1+mod(i+30,55))
if(ma(i).lt.MZ)ma(i)=ma(i)+MBIG
12 continue
13 continue
inext=0
inextp=31
idum=1
```

```
endif
inext=inext+1
if(inext.eq.56)inext=1
inextp=inextp+1
if(inextp.eq.56)inextp=1
mj=ma(inext)-ma(inextp)
if(mj.lt.MZ)mj=mj+MBIG
ma(inext)=mj
ran3=mj*FAC
return
END
```