Atomistic Simulations Uncover Microscopic Details of Nucleosomal Electrostatics, Energy Landscapes of Proteins and Photovoltaic Polymer Dynamics.

Christopher Kroboth Materese

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry.

Chapel Hill

2010

Approved by:

Advisor: Professor Garegin A. Papoian

Reader: Professor Michael Rubinstein

Reader: Professor Max L. Berkowitz

Reader: Professor Andrew L. Lee

Reader: Professor John M. Papanikolas

Abstract

Christopher Kroboth Materese

Atomistic Simulations Uncover Microscopic Details of Nucleosomal Electrostatics, Energy Landscapes of Proteins and Photovoltaic Polymer Dynamics.

(Under the direction of Garegin A. Papoian)

Molecular dynamics (MD) simulations offer researchers a high resolution window into the atomic world. Though far from perfect, they provide researchers with a direct method for probing molecular scale processes. We have used MD to address a range of questions which span three fundamentally different topics. First, we used MD to elucidate the native state energy landscape of a small globular protein. We are able to identify a subset of direct and water-mediated contacts which may be responsible for sculpting this landscape. Next, we turn to the topic of chromatin, specifically, nucleosomal electrostatics. Using a combination of all-atom simulations and Poisson-Boltzmann calculations we are able to observe and explain counterion condensation levels and distribution patterns around the nucleosome core particle. Additionally, our results reveal the significant solvent accessibility of the core particle. Finally, we use all-atom simulations to examine small, artificial, coiled-coil peptides under development for use in light harvesting antennae. Photosensitive chromophores can be tethered to these peptides at a range of sites, and we observe that the choice of these sites plays an important role in regulating energy transfer. In addition, the flexibility of the tethers imbues the chromophores with a large amount of conformational freedom, making their dynamics important for regulating energy transfer.

Acknowledgements:

Mom and Dad, thank you for your endless support in all facets of my life. To properly thank you both would consume more space than the contents of this dissertation so I will be brief. Thank you for always being there when I needed someone to talk to. Thank you for teaching me the importance of education and inspiring me to try things that were difficult. Thank you for the hours spent helping me proofread my work. Thank you for the financial support you have given me throughout my life. Finally, thank you for always believing in me and making sure I knew it. Professor Papoian, thank you for your inspiration, guidance, and patience throughout the course of my research. You have taught me to be a better scientist and teacher. You have pushed me when I needed to extend myself and helped me build confidence in my work. Pavel and Alex, thank you for all of the help you both have given me through the years of my Ph.D. work. Thanks to my committee members, namely Professor Max Berkowitz, Professor Michael Rubinstein, Professor John Papanikolas, and Professor Andrew Lee for their support and guidance in my research, especially during polymer physics seminars, journal clubs, EFRC meetings, or individual meetings. Thank you to my undergraduate advisor Anne Andrews for encouraging me to pursue graduate work. Thank you to a long list of professors and teachers who have inspired me at different stages of my life. Since this is a chemistry dissertation, I will limit myself to science and math educators that have played an important role in my life. At Penn State, I want to thank Professor Pshemak Maslak for helping me solidify my decision to major in chemistry and I want to thank Professor James Anderson for helping me choose Physical Chemistry. In reverse chronological order, thank you to Stephen Tibenski, Janet Levinson, Eileen Lane, Charles Goodwin, Lynn Bradley, Ken Lyman, Polly Peterson, Dave Barnard, and Judy Wilson. You have all pushed me to work harder and strive for more and you dedicated your time and energy to make learning an exciting journey. Thank you. Finally, thank you to my family and friends, especially my sisters Kate and Beth for encouragement and support.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

1D, 2D, 3D	One-, two-, three-dimensional
1EGL	Eglin c PDB entry
1ID3	Yeast Nucleosome PDB entry
AMBER	Assisted Model Building with Energy Refinement
APBS	Adaptive Poisson-Boltzmann Solver
atm	Atmospheres
B3LYP	Becke 3-Parameter, Lee, Yang and Parr
BALL	Biochemical ALgorithms Library
C1	Chromophore on peptide 1
C2	Chromophore on peptide 2
CHARMM	Chemistry at HARvard Macromolecular Mechanics
CPU	Central Processing Unit
DG	Disconnectivity Graph
DNA	Deoxyribonucleic acid
ε	dielectric
e	Elementary charge

E	Barrier Energy
EFRC	Energy Frontier Research Center
eV	Electronvolts
F10W	Eglin c residue 10 mutation of phenylalanine to tryptophan
ff99SB	An AMBER parameter set
γ	Residual net charge of DNA within 1 nm of chain
GHz, THz	Gigahertz ($(1/s)^9$), Terahertz ($(1/s)^{12}$)
H2A, H2B, H3, H4	Histone proteins
k	Boltzmann constant
K	Kelvin
kcal	Kilocalories
LANL2DZ	Los Alamos National Laboratory 2-Double-Zeta
^{1}MLCT	Singlet metal to ligand charge transfer
^{3}MLCT	Triplet metal to ligand charge transfer
MD	Molecular dynamics
mol	moles
Ru*	Ruthenium complex excited state
mM	millimolar, 10^{-3} mol/L
NAMD	NAnoscale Molecular Dynamics

nm	nanometer, 10^{-9} m
NMA	Normal Mode Analysis
NMR	Nuclear Magnetic Resonance
NPT	Isothermal-Isobaric ensemble
NVE	Constant Volume-Isoenergetic ensemble
$[Os(bpy)_3]^{2+}$	Osmium tris-bipyridine
P-B	Poisson-Boltzmann
PC	Principal Component
PCA, dPCA	Principal Component Analysis, dihedral Principal Component Analysis
PDB	Protein Data Bank
PyMOL	Python based molecular visualization program
$[Ru(bpy)_3]^{2+}$	Ruthenium tris-bipyridine
RDF, g(r)	Radial Distribution Function
RESP	Restricted Electrostatic Potential
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
μs, ns, ps, fs	microsecond (10^{-6} s), nanosecond (10^{-9} s), picosecond (10^{-12} s), femtosecond (10^{-15} s)

S^2, S^2_{axis}	Lipari-Szabo model free order parameter for N-H bonds, 3-fold methyl
T	Temperature
T_f, T_g	Folding temperature, Glass transition temperature
TIP3P	3 point water model
TW	Terawatt

Chapter 1

Introduction

1.1 Energy Landscapes of Proteins

1.1.1 An Introduction to Protein Folding and Features of the Native State

Studies in molecular biology have shown time and again that protein function is intrinsically connected to structure, dynamics, and thermodynamics. Rationalizing protein dynamics and its connection to the protein energy landscape is not only an important biological problem, but also presents extremely interesting physicochemical challenges. In order to provide context and a broader view for our discussion, we will briefly overview the celebrated protein folding problem [1–12] and then shift our focus to the native state.

Proteins are polymer chains consisting of amino acid residues connected by covalent peptide bonds. Many of the principles of polymer physics are directly applicable to the study of proteins. One of the most interesting thermodynamic properties of polymer chains is the "coil – globule" transition [13], which, as with all phase transitions, is directed by the competition of energy and entropy; if the energy of attractive intramolecular interactions overcomes the conformational entropy (as occurs in a poor solvent/low temperature), the chain precipitates into a compact globular state, with the volume fraction of polymer on the order of unity. In the

opposite case (good solvent/high temperature), entropy dominates and a coil state is preferred. This state tends to possess extreme fluctuations and the volume fraction of the polymer tends to zero in the thermodynamic limit ($\rho_p \sim N^{-\alpha}$ where N is the degree of polymerization and $\alpha > 0$).

The "coil – globule" transition in proteins [14, 15] is more interesting than in homopolymers and random heteropolymers and has special features, absent in both. This transition is accompanied or followed by protein folding, during which the protein chain adopts a specific tertiary structure. The tertiary structure differs from the secondary structure in that it dictates the features on the scale of the whole chain, whereas secondary structures like, α-helices and β-strands are locally determined by the sequence. A typical protein sequence is exquisitely built in order to provide a unique native state, that is both lowest in energy and biologically functional [16, 17]. While this is not a single static state in reality [18, 19], for the moment we shall speak in approximate terms because the native ensemble is significantly smaller than that of a homopolymer globule. The sequence also ensures that the native state is kinetically accessible from the unfolded state. These features are essential to meet the demands of the biological role of proteins. If we examine the phase space available to a protein polymer chain (depicted in Figure 1.1 as a schematic representation of the funneled landscape), we can identify several distinct and fundamentally important regions as solvent quality changes. The "coil" region is the most disordered, though still not as random as a typical homopolymer coil [20]. Since there is no clearly defined structure, function should not be possible according to the traditional "structure – function" paradigm. Despite this, some disordered proteins have been shown to be functional [21, 22], though discussion of the non-compact states observed in these proteins is beyond the scope of this dissertation. In the coil state, flickering secondary structure elements reconfigure on the timescale of about 1 ns [23], but to what extent the residual structure exists is still ambiguous [24]. In any case, the coil state is dominated by entropy. Even without considering the solvent, the conformational entropy of the chain is much more significant than

the volume interactions, analogous to gases. In fact, Rouse-Zimm theory gives an appropriate description of the dynamics [25], taking into account viscous friction between the solvent and monomer units and hydrodynamic interactions.

As the solvent quality decreases (possibly as a result of a temperature decrease), the protein undergoes a transition to *the molten globule region* [26–31](Figure 1.1). In a classical polymer science, this is already a compact state, but in protein folding terms it is an ensemble of rather general intermediate states [32–36] which are still considered unfolded and highly dynamic. Some of them contain a substantial amount of secondary helical structure. Various degrees of helicity were observed [37, 38], as well as some tertiary structure [27, 33, 39–41]. An extension of the analogy above would be to parallel the dynamics of a molten globule to that of liquids [42]. Again the "structure – function" paradigm maintains that there should be no functionality in this state, as there is no well-defined structure. However, exceptions have been found recently, which will be discussed in this paper.

When the solvent quality decreases even further, the protein starts to feel the ruggedness of its own energy landscape. The protein spends more time in the minima of the landscape, rather than in saddle points, and the barriers between the minima become harder to overcome. Such dynamics is similar to that of a supercooled liquid [43]. Thermodynamically, it is still a molten globule, but in a real experiment, some conformational states may remain unexplored despite being thermodynamically favorable, because of the finite run time: the mean first passage time for crossing a barrier of height E is proportional to $\exp(E/kT)$ according to Arrhenius relation.

At some temperature T_g, the barriers between the minima become insurmountable on any feasible experimental timescale. The dynamics of a protein whose temperature has reached T_g, mark the transition to *the region of glassy states* [44, 45]. In this regime, a protein gets trapped in a single, random conformational basin and if these states are not functional, the protein will not be able to perform its required role. Therefore, for many naturally evolved proteins it is unsurprising that T_g is well below room temperature [5, 9, 45, 46].

Typically, proteins undergo folding from a molten globule to what is called *the native state* at folding temperature (T_f) which is well above T_g [3, 5, 9, 47, 48]. To continue our analogy, the native state can be compared to crystalline states [42, 49], as the conformational entropy is much lower and the effective energy of contacts plays the dominant role. However, here the parallel ends, since the native state possesses far more than just vibrational entropy [4–6, 9, 18, 50].

Though the focus of this work deals with systems in the native state, it would be prudent to briefly describe how a protein arrives at this state. In 1969, Levinthal postulated a thought experiment which proved that proteins could not possibly find the thermodynamically stable functional state via random exploration of their phase spaces [1, 9]. Consider a 100 residue protein, where each residue may adopt one of three alternative states, then the total number of possible conformations available to that protein may be roughly estimated as $3^{100} \approx 10^{47}$. If we restrict conformational transitions to occur no faster than once per picosecond (10^{-12} sec), it would take roughly 10^{28} years to complete an exhaustive search, significantly longer than the age of the universe. A protein possessing such an energy landscape could not reach a functional conformation within a biologically relevant timescale, therefore, there must be a defining feature of their energy landscape that circumvents this problem. This realization led to the insight that the energy landscape must be biased in favor of the native state such that folding through an exhaustive search of the protein's phase space does not occur [51]. Unlike random heteropolymers, proteins possess a funneled energy landscape which drives folding to the native state [4–6, 9, 52–54](Figure 1.1). Evolution has carefully selected protein sequences which are able to quickly fold to their native state upon translation. This means in general, organization of a protein's energy landscape is dominated by the native state [55]. However, a free energy bias alone is not enough to guarantee consistent successful folding. If the landscape is funneled but very rugged, trapping still may prevent folding on reasonable timescales. This brings us to another very important, feature of the protein energy landscape. However, in order

to discuss this we must first introduce the concept of frustration.

Frustration, simply put, is the inability simultaneously to achieve favorable interaction energies between all contacts. Frustration leads to an energy landscape with many local minima separated by barriers. Such a situation is also encountered in spin glasses, an elegant example given by Bryngelson *et al* [6]. Consider a system in which magnetic spins are arranged randomly. Interactions between spins may be either ferromagnetic or antiferromagnetic and occur at an equal frequency. Frustration arises out of a competition between these two mutually exclusive tendencies. The contacts between monomer units of a heteropolymer can also be energetically either favorable or unfavorable, and this analogy allows researchers to use methods and models developed for glasses to study proteins.

Proteins have evolved to minimize the frustration in their native states, in order to keep the funneled landscape as smooth as possible to minimize trapping in non-functional conformations [51]. Despite the kinetic need for a minimally frustrated landscape, some residual frustration remains at the bottom of the funnel [56, 57]. The energy landscape of the native basin is degenerate with multiple ways to achieve low effective energy structures. Conceptualizing the native state as an ensemble of states rather than one single state was a significant advancement in protein science. This breakthrough in the description of the native state occurred over three decades ago when Austin *et al.* conducted low-temperature flash photolysis measurements, providing experimental evidence for the existence of substates within the native landscape [58]. In this study, an anomalous myoglobin-ligand binding energy spectrum led the researchers to conclude that myoglobin must possess a multitude of conformational states with differing activation energies. Subsequent experiments by Ansari *et. al.* suggested that the protein energy landscape is organized hierarchically with various states divided into substates which themselves may be subsequently divided further into substates *ad nauseam* [4, 18, 59].

Understanding the dynamics of a protein in the native state is a separate very interesting problem with profound biological applications. Examples include fluorescence intermit-

tency [60] and allosteric regulation [61]. The first molecular dynamics simulation of a protein suggested that protein molecules are more fluid than originally expected [62]. This fluidity provides a protein with the means to transition between conformational states through thermal and solvent fluctuations. There exist diverse viewpoints on the nature of dynamical transitions between the native substates. Some techniques, such as normal mode analysis (NMA), treat proteins as solids possessing vibrations and phonons [63–65]. This approach works well in some cases and has provided significant insight into protein dynamics. However, it has the significant drawback that it is difficult to reconcile with the multitude of conformational minima and single minimum anharmonicities. This technique should be contrasted with Principal Component Analysis (PCA), which will be later discussed in detail, in which anharmonicities are retained [66]. Work has also been done on the related field of instantaneous normal mode analysis to account for the existence of multiple minima in the energy landscape. This technique uses the differing timescales of thermal solvent motion and solute vibrations to assume the separability of the Hamiltonian into individual solvent and solute Hamiltonians in addition to a third Hamiltonian which describe their interaction. This technique has been reviewed by Schmitz et. al. [67], briefly, MD simulations are used to generate a collection of solvent cages which are then frozen for subsequent NMA; anharmonicites can be introduced through quantum corrections. While this is technique can be quite useful in some systems, and has even been applied to peptides [68–70], been used, translation of the work to proteins is complicated by the very large number of local minima in the energy landscape. Another option is to imagine activated hopping between different conformations separated by energy barriers [71, 72]. This can be paralleled to the dynamics of a supercooled liquid. Yet another possibility is that protein's motion is similar to flow of a normal liquid such that the system spends most of the time in saddle points rather than in minima of the energy landscape. There have been indications that activated, or even glassy dynamics, are more relevant to proteins, at least at low temperatures [3, 4, 73].

Knowledge of the existence of a rugged native energy landscape, that is organized in hierarchical tiers, has inspired extensive research both experimentally and theoretically to further our understanding of its nature. Complete characterization of a protein's native landscape is not feasible with modern computer technology, but it is still possible to gain important insight with various techniques. Significant work has been done to characterize the energy landscape for peptides (mainly via models with implicit solvent), using disconnectivity graphs (DG) [72, 74] which provide important insight into finer scale splittings of similar structural clusters and provide information about the transition barrier heights. A different approach to modeling the energy landscape of a protein emerges from graph theory and the analysis of complex systems such as those applied to World-Wide Web [75]. In this technique, molecular dynamics simulations are conducted and conformations are characterized by secondary structural elements. These conformations then become nodes on the network graph while transitions between different conformations become the links. A detailed, hierarchical network emerges which provides significant insight into the nature of the energy landscape. Another approach is to compute a low-dimensional map of the energy landscape. A typical result is a one- or two-dimensional free energy map of the system which can be used to study structural clustering and kinetics of conformational transitions. In the research, discussed in Chapter 2, PCA was the method chosen to characterize the native state.

1.1.2 Kinetic hierarchy elucidated by PCA

As temperature decreases, the *molten globule* folds to the *native state*. The native state, however, is still very dynamic and possesses measurable entropy due to residual frustration. This residual frustration splits the native state into an ensemble of substates, which are organized in hierarchical tiers both kinetically and energetically. Often, the investigation of the *molten globule* is focused more on the thermodynamics, with low-resolution order parameters such as radius of gyration. The *native state*, however, must be studied from a more precise structural

perspective since specific structures are responsible for determining functional activity, such as enzymatic catalysis.

PCA is a powerful linear orthogonal technique used to aid in the comprehension of complex multidimensional systems, such as a protein, by reducing the phase space while retaining essential degrees of freedom [66] and accounting for anharmonicities. PCA operates through the diagonalization of a covariance matrix of the coordinate fluctuations of the system. The eigenvectors obtained by this diagonalization are the principal components which are sorted by decreasing eigenvalues. The eigenvalues represent the variance of the data along each eigenvector, meaning that the first principal component retains the greatest variance of the data followed by the second and so on. It has been shown that the majority of the degrees of freedom in a full multidimensional hyper-space of a protein's dynamics are uninteresting and contain no essential information [66]. Therefore, in order to simplify our problem we identified the essential degrees of freedom by projecting our MD trajectory into PC space, and histogramming that data. Data projected into a PC which only characterizes fluctuations within a single state acts like a harmonic oscillator and will appear as a single Gaussian distribution [66]. On the other hand, data projected onto one of the PCs belonging to the essential subspace would appear as a multi-peaked distribution with different peaks representing the states accessible within that PC. Thus, the essential degrees of freedom were identified by isolating the Non-Gaussian forming PCs and eliminating Gaussian forming PCs.

Much work has been accomplished toward the application of Principal Component Analysis to protein systems. Early work by Amadei et.al. laid the foundation [66] and subsequently, significant work in toward characterizing molecular dynamics trajectories of peptides has been performed. García et. al. used PCA to create free energy surface maps of a small peptide as a function of temperature using the first two PCs [76]. Becker evaluated the effectiveness of PCA at capturing peptide dynamics and also used it to create energy surfaces [77]. In a later work, Levy and Becker used principal coordinate analysis, a PCA variant in conjunction

with disconnectivity graph analysis, to illustrate the effects of conformational constraints on the peptide energy landscape [78]. A paper by Altis *et. al.* directly addresses the choice of the original coordinate system upon which PCA is performed and illustrated that dihedral angles provide significant advantages over Cartesian coordinates in the construction of energy landscapes [79]. Recently, Hegger examined the complexity of peptide folding and compared the dimensionality of an energy landscape obtained by PCA to the dimensionality of the dynamics obtained through Lyapunov analysis [80]. Our work can be seen as an extension of these earlier efforts in that we used PCA to characterize the MD trajectory of a small protein, revealed the latent kinetic hierarchy and studied the divergence in inter-residue contacts which sculpted this hierarchy in addition to examining the role of water-mediated interactions.

In Chapter 2, the results of a study of the small globular protein *eglin c* are presented. As mentioned, the goal was to observe the nature of the hierarchical landscape defining the dynamics of *eglin c* in an explicit solvent simulation. Additionally, we wanted to determine the role played by inter-residue direct and water-mediated contacts in sculpting this landscape. To achieve these goals, we conducted a long 336 ns MD simulation of the protein in explicit solvent and employed dihedral angle principal component analysis of to deconvolute essential degrees of freedom and reduce the dimensionality of the system. Our results showed that the first four PCs revealed highly Non-Gaussian and the next seven revealed slightly Non-Gaussian distributions of trajectory data, suggesting eglin c is sufficiently described in at most an eleven dimensional manifold or roughly 3% of the characterized degrees of freedom. This result is important because it reassures us that it is indeed possible to describe protein's native basin with a manageable number of collective degrees of freedom.

After establishing limits on the essential phase space, we proceeded with a detailed characterization of the energy landscape. Using a two-dimensional histogram of the trajectory projected in the first two principal components we observed the appearance of several densely populated regions. These regions were considered basins in the energy landscape as there is

a spontaneous tendency for the system to adopt the structural conformations associated with those basins. While the totality of the protein dynamics was not restricted to reside solely within these basins, they did represent a very significant portion of the native landscape. With our knowledge of the energy landscape from the first two degrees of freedom, it was possible to extract further detail by examining each basin individually. Trajectory data residing within a single basin was then plotted as a two-dimensional histogram in the second and third PCs (thus, we zoomed into the basin using the next set of PCA dimensions). By doing this we further identified basins in the energy landscape which were eclipsed when viewed from the first two PCs as shown in Figure 1.2. This procedure may be repeated in an iterative fashion for all subsequent essential degrees of freedom. To the best of our knowledge, this technique has not been used prior to this work. The result of this work is the identification of a hierarchy of states (Figure 1.3) in the energy landscape whose structural significances may be determined by sampling the individual conformations residing within each basin.

Structural sampling from the observed basins provided valuable information about the protein's native dynamics. At a coarse visual level, we used the collection of conformations in each sampling to create an average structure, displaying the characteristic features of that basin. The most significant large scale structural rearrangements occurred between basins obtained from the histogram of the first two principal components. Differentiation observed from the splitting of subsequent PCs became increasingly subtle. For a more substantive description of the significance of the landscape hierarchy, we examined the inter-residue contacts present in each basin. We found that basin splitting is accompanied by a change in intramolecular contacts formed within the protein as depicted in Figure 1.4. The observed landscape hierarchy was thus characterized by a divergence of intermolecular contacts between basins. Subordinate basins retained the defining features of their superiors while further differentiating amongst themselves, showing ultrametric features characteristic of spin glasses [81, 82]. Interestingly, we extended our search for divergent contacts to include water-mediated interactions and found

that these interactions played a similarly important role in basin definition. This observation is consistent with prior works on the way water-mediated interactions guide protein folding and stabilize native state architectures [83–86].

The general techniques used in this work are useful for investigating other protein systems. In summary, our results showed that competition between direct and water-mediated inter-residue contacts sculpts the landscape and dynamics of the native state and exemplifies the residual frustration of proteins' native ensembles.

1.1.3 Role of Water

Water plays an extremely important role in both protein folding and native dynamics [83, 87–90]. The hydrophobic effect has been long known as the main driving force of protein folding [91], and early heteropolymer models introduced it as a pairwise attraction between the hydrophobic monomer units [92–95]. As we mentioned in the introduction this tradition still has an impact on contemporary terminology, like "energy" landscape, which includes phenomena of entropic origin. As described below, our results indicate that the true thermodynamic energy landscape is much more rugged, than the landscapes studied in protein folding theories. Long-range hydrophilic contacts are also mediated by water [85] – introducing a second well into the pairwise polar – polar potential improves protein structure prediction. This second well corresponds to a water-mediated contact, a common occurrence in the native state [85,96]. Overall, water plays a significant role in smoothing the folding funnel and guiding the folding of a protein to its native state [85]. The dynamics of the protein and its hydration shell are also interconnected. In some cases protein motion is though to be slaved to solvent motion [58,97], as solvent effectively cages the protein. The collective modes of these coupled protein and solvent dynamics can be probed by THz spectroscopy. A recent study suggest that the so-called protein dynamical transition – the rapid increase of the dynamics at ~ 200 K – is originated in the motions of water and interaction between water and side-chains [98]. THz spectroscopy

has also been used to study the effect of mutations on hydration shells of the proteins, showing a case where psedo-wild-type has a much more pronounced effect on long-distance solvation water than a point mutant [99].

In Chapter 2 we show that the presence of water-mediated contacts is correlated with hierarchical organization of the PCA derived basins, which in turn is relevant to hierarchical organization of the energy landscape. Water bridging was found to involve both side chain and backbone chemical groups. These bridges were found both between residues close enough to be considered in direct contact, and interactions at longer range. The mediation of direct contacts often included residues of like charge. We suggested that hydrogen bonding, created by favorable water interactions diminishes repulsion between the like charged residues, allowing them to develop such a contact. We observed changes in water-mediated contacts between different basins in our hierarchy, suggesting that water plays a role in shaping said hierarchy. Proteins have evolved in aqueous environments and water can play important, specific, microscopic roles as shown in this work. The importance of water at specific sites in the protein structure underscores the importance of explicit solvent in fine scale structure determination. Furthermore, recent experimental work points to the importance of water in sculpting specific protein structure and function. In Szep *et. al.*, the authors perform crystallographic studies on wild type and mutant FKBP12 binding protein and provide evidence that a specific crystallographic water, distal to the binding pocket, plays a role in sculpting the shape of that pocket [86]. The authors suggest that despite the fact that the binding pocket is >8 Å away, this specific water plays a key role on a network of interactions which shape it.

1.2 Chromatin

1.2.1 Fitting Meters of DNA into a Micrometer Scale Nucleus

Complex eukaryotic organisms typically require and possess very large genomes. For example, a typical human genome possesses approximately 3 billion base pairs of DNA in duplicate, which, if fully extended, would stretch to over 2 m in length [100]. This DNA must be made to fit inside a cell nucleus on the micron scale. In addition to the compaction, the DNA must be arranged in such a way that some sections are quickly and easily accessible for processes such as replication, transcription or translation. To further complicate the matter, DNA is higly charged and possesses $2e^-$ charges per base pair, making self-repulsion a significant hurdle to overcome in the process of compaction. In order to deal with this problem, eukaryotic cells have evolved a specialized protein and DNA superstructure known as chromatin fiber [101]. In general, the most basic repeat units of this fiber consist of a protein-DNA complex known as a nucleosome and a segment of DNA known as linker which connect to the next nucleosome [102]. The nucleosome itself consists of DNA wrapped in $\sim$ 1.7 turns around a set of highly specialized proteins known as the histone core. The histone core is an octomer containing 4 separate histone proteins in duplicate (H2A, H2B, H3, and H4). The histones themselves are highly conserved with nearly identical sequences and structures between organisms as diverse as humans and yeast. Histone proteins possess a large net positive charge which serves as a partial counter to the surrounding negatively charged DNA [103]. In an extended state, nucleosomes and linker DNA are likened to beads on a string, also known as 10 nm fiber. At the first level of compaction, nucleosomes and linker DNA arrange themselves into what is known as 30 nm fiber, the structure of which is remains controversial. Subsequently, chromatin arranges itself into structures known as chromosomes, structures $\sim$10,000 times shorter than the contour length of the fiber [100].

Schiessel likens chromatin to a library with the following analogy [100]. If you consider the

text contained in books to be analogous to the information contained in the genome, an individual book may contain $\sim$ 10 km of text (if each word was written in a single line) and a whole library may contain $\sim$10,000 books leading to a total of $\sim$ 100,000 km of text. If unorganized, finding a needed line of information hundreds of thousands of kilometers of text would be a daunting task. Thankfully, the text in the library is arranged hierarchically in lines, paragraphs, pages, books and shelves. In an analogous fashion, chromatin is also hierarchically organized although many questions about the mechanisms for information access still exist. Thus, the hierarchical arrangement of the genome into nucleosomes and then increasingly complex levels of chromatin fiber achieves two goals. It efficiently packages the DNA fiber into a small space and regulates access to the genome for purposes of replication, transcription or translation.

1.2.2 Electrostatics of the Nucleosome

Chromatin folding is strongly dependent on the concentration of mobile ions which mitigate the electrostatic repulsion, and alterations to the surrounding ionic solution can have substantial consequences for the formation of higher-order chromatin structures [104]. Thus, how mobile ions interact with both the nucleosomal DNA and the protein histone core is an important issue to be addressed. Prior work was done in this field to examine ion condensation around strands of free DNA, which can be seen as an approximation of the linker DNA segments in chromatin fiber [105–108]. The next logical extension of this research was to study the nucleosome because a precise description of site specific electrostatics of this structure is likely important for an accurate description of chromatin folding. It has been suggested, for example, that histone tails, which are flexible, highly charged, and unstructured portions of histone proteins that protrude outward from the core and play an important role in chromatin compaction, may actually bind to a specific acidic patch on the histone core of neighboring nucleosomes [103, 109–111]. In Chapter 3, our recent study of nucleosomal electrostatics is discussed in detail. As a preview, some of the basic findings are discussed here. First, the DNA surrounding the nucleosome has

a charge of -292, therefore according to counterion condensation theory, we would expect $\sim$ 219 of these charges to be neutralized by counterions. The histone core provides a net charge of +76 within 1 nm of the DNA, therefore, a release of 76 mobile counterions is anticipated if we assume a 1 to 1 replacement. However, we observed a net +174 charge from counterions within 1 nm, indicating that only 45 counterions were released, instead of expected 76. In order to explain this phenomenon, we built up the system with a series of P-B calculation in a manner akin to the Aufbau principle. This allowed us to provide a detailed analysis of the causes of the excess positive charge: First, the close wrapping of DNA leads to an additive effect of the electric fields produced by the DNA and causes an increase in sodium condensation relative to that normally observed in free DNA. Second, the difference in dielectric constants between the water and the histone core which the DNA encircles creates a mirroring effect that enhances sodium condensation. Third and finally, immobilizing some counter charges near the DNA leads to an overall enhancement of the charge neutralization. Thus we find that the histone core, in conjunction with counterions, is better able to neutralize the charge of the encircling DNA than would be possible with monovalent counterions alone.

2-D RDFs and 3-D ion density plots revealed the presence of highly localized sodium condensation on the largely positively charged histone core at the acidic patch which is suggested to play a significant role as a potential internucleosomal interaction site for histone tail binding [103, 111]. This finding suggests that histone tails binding to the acidic patch may be entropically favorable because of counterions release. The last bit of information gleaned from this study related to hydration patterns around the nucleosome. Water molecules were found to permeate the nucleosome interior. This is important because it leads to an overall increase in the average interior dielectric which also allows mobile ions to penetrate to the interior of the core and diminished electrostatics repulsion between individual histone monomers.

1.3 Photovoltaic Polymer Dynamics

According to the Department of Energy, in 2009, renewable energy sources only account for 7.3% of total US energy production [112]. Of these renewables, there are a wide range of source types (Figure 1.7). The largest slice of the current renewable pie comes from fuels derived from biomass (53%) [112]. The advantages of these sources are that they are carbon neutral since they only emit carbon which was absorbed from the atmosphere in the first place, they are domestic, and they make use of the same or similar infrastructure to that used by fossil fuels. The main drawback of this resource is that its scale is limited by the amount of arable land available putting it in direct competition with land used for the food supply. Another troubling aspect of using biomass for energy is the possibility that it actually consumes more energy to grow the plants than it produces [113]. The next largest slice of the pie comes from hydroelectric sources (34%) [112]. Hydroelectric power is an excellent source of clean renewable energy and is cost competitive with fossil fuels. Although hydroelectric power is carbon neutral, domestic, efficient, and cost effective, it also suffers from a number of limitations and drawbacks. These include limitations on location, initial cost, safety issues, environmental damage from flooding, changes in water quality and damage to fish populations. Next on the list, though far below hydroelectric power, is wind power (7%) [112]. Wind produces energy at a fairly cost competitive rate and is carbon neutral. Wind does, however, suffer from two large drawbacks. First, with current technology, even if wind turbines were installed wherever it was economically feasible they would only generate 2 to 6 TW of power (current consumption is near 13 TW with an expected 30 TW of addition power needed by 2050) [114]. On the bright side, newer windmill technology designed to harvest wind 80 meters above the ground is projected to up the production to 72 TW, though these estimates may be overly optimistic [114]. The second significant drawback of wind power is the inconsistency of the wind itself. In general, wind tends to be strongest in the early morning and grows weaker during the day. This presents a problem because peak energy usage is during the middle of the day. To deal with

this problem, and similar problems with other renewables including solar power, new methods of energy storage must be developed to ensure a consistent level of available power. Geothermal energy takes up 5% of the renewable pie. This method also requires a suitable location for efficient implementation. The best locations tend to be around volcanoes and fault lines which itself poses potential problems. Additionally, the initial cost for the creation of these plants tends to be large. Finally, we arrive at direct use of solar energy. Sunlight is a promising form of renewable energy and has the greatest potential of all carbon-neutral sources. If it were possible to capture and store all of the sunlight that strikes the earth over the course of one hour, it would be sufficient to meet the energy needs of the entire planet for more than a year. Direct use of solar energy currently accounts for a mere 1% of the renewable energy produced in the United States.

There are two primary methods of converting sunlight into usable energy. First, it can be converted directly into electrical current through the photovoltaic effect in which electrons in the material are promoted from the valence to the conduction band. This method is used in traditional solar cells. Alternatively, sunlight can be used to power chemical reactions including those which can split water into H_2 and O_2 or reduce CO_2 to methanol or methane. Presently, neither of these options is cost competitive with fossil fuels, in part because of the need for rare metals and expensive fabrication techniques [115]. To address this issue, the Department of Energy awarded UNC with a grant to form a highly collaborative and interdisciplinary group which has been dubbed the UNC Energy Frontier Research Center (EFRC). This group contains scientists from a wide range of backgrounds from synthetic organic chemists to polymer physicists who all share the common goal of the development of new solar fuel technologies.

In Chapter 4, current, ongoing research for the EFRC is discussed. One of the projects in the EFRC involves the development of light harvesting antennae using polymer scaffolds which support Ruthenium based chromophores. One of the most important features of these antennae is the ability to perform efficient energy transfer along its length and a regular, controlled

positioning of chromophores makes their behavior more predictable. Marcey Waters' group is developing such scaffolds using small coiled-coil peptides. Coiled-coils may be useful because they may have the desired predictable structure. To better understand these systems, we are using all-atom MD simulations to provide atomic level detail for energy migration studies by explicitly modeling relevant conformational changes of the peptide scaffold. Currently, we are running four all-atom, explicit solvent simulations on different peptide scaffolds developed by the Waters group. Specifically, we use the AMBER force field [116] with the ff99SB parameter set [117]. Since this system contains non-biological amino acids and transition metal complexes, the parameter set must be supplemented with additional data such as partial charges and force constants. Force constants, equilibrium bond distances and Van der Waals parameters for the ruthenium complex were obtained from published literature. Partial charges of unparameterized atoms were obtained using B3LYP [118] density functional theory calculations with the LANL2DZ basis set [119–121]. Mulliken charges were used in lieu of the traditionally preferred RESP charges [122] because the RESP technique was incorrectly predicting partial charges of ruthenium and surrounding nitrogen atoms. Since no crystal structure was available for the system, the initial structure was generated using PyMOL [123], a molecular visualization tool capable of constructing simple peptides. The peptides were initially generated independently using PyMOL's helical parameters and were then manually aligned with care taken to avoid steric clashes and satisfy hydrophobic the strip of the coiled coil interface.

Within a window of 150 ns ($7.5x10^7$ time steps, $\sim$30,000 CPU hours) we notice an unexpected and significant association of the chromophores and their tethers with both the main chain of the peptide and with the opposing chromophore and tether. The chromophores and their tethers are also able to dissociate and rebind. We hypothesize that this phenomenon is driven by the hydrophobic effect. The association and dissociation events may reveal a set of preferred conformations which possess highly non-Gaussian chromophore- chromophore distance distributions. Since these events occur on the timescale of experimental energy relax-

ation studies, we expect that peptide dynamics may play an important role in energy transfer processes. We characterize the major differences in these chrmophore-chromophore distance distributions between systems with differing tethering points, and will use our results to interpret experimental energy relaxation studies conducted by John Papanikolas' group.

1.4 General Techniques

1.4.1 Molecular Dynamics

Over half a century ago, Molecular dynamics (MD) was born following the advent of computers. The ergodic hypothesis, which states that given sufficient exploration of the phase space, the time average of a system property will equal the statistical average, provides the theoretical underpinnings to the application of MD simulations to study real systems. A paper published by Alder and Wainwright [124] outlined a method by which it would be possible "to calculate exactly the behavior of several hundred interacting classical particles." The method was developed for very simple molecules and could not handle systems larger than a few hundred atoms because of the limitations of the computers at the time of its inception. In order to best mimic macroscopic systems with such a small number of atoms, the authors suggest the use of periodic boundary conditions. Periodic boundary conditions also had the advantage of producing accurate virial coefficients. The method was simple: First, the force on each particle in the system would be calculated from its interactions with neighbors. The system would then be permitted to evolve with these constant forces for a short period of time. After the move generated in the previous time interval, the forces could be recalculated and the system allowed to evolve again. This initial work used a simple square well potential and an (N,V,E) ensemble to avoid approximations and allow comparison to analytical results. Though simplistic, these simulations were used to study phenomena such as energy redistribution and sample diffusion. In the following years, the use of more complex interaction potentials became com-

monplace and molecular dynamics was used to investigate structural and dynamical properties of liquids [125].

Nearly 20 years after its initial inception, the first molecular dynamics simulation on a protein was performed by McCammon and Karplus [62]. The simulation was short by modern standards; only 9.2 ps, used a crude force-field, omitted hydrogen atoms, and was conducted in a vacuum. Despite these flaws, this simulation provided a whole new perspective on the world of proteins and the importance of their dynamics. Although some theory and experimental data pointed to the contrary, at the time, proteins were thought to possess relatively rigid structures, thanks in part to X-ray crystallography. This simulation showed that at room temperature, the protein readily explores the energy surface surrounding the crystal structure and that the internal motions of the protein were much more "fluid-like" than previously thought. Since that time, the field of molecular dynamics simulations has greatly matured and all-atom simulations with tens to hundreds thousands of atoms with up to a microsecond duration are not uncommon.

There is a rather wide selection of modern MD force-fields with CHARMM [126] and AMBER [127] being two of the most popular for biologically relevent applications. The functional forms of CHARMM and AMBER are very similar, so for simplicity, only the latter will be discussed in detail (Eq (1.1)).

$$\begin{aligned} V(r) = & \sum_{bonds} \frac{1}{2} k_b (x - x_0)^2 \quad + \quad \sum_{angles} k_a (\theta - \theta_0)^2 \quad + \\ & \sum_{dihedrals} \frac{V_n}{2} [1 + cos(n\omega - \gamma)] \quad + \\ & \sum_{nonbonded} \left(\frac{A_{ij}}{r_{ij}^{12}} \right) - \left(\frac{B_{ij}}{r_{ij}^{6}} \right) + \left(\frac{q_i q_j}{r_{ij}} \right) \end{aligned} \tag{1.1}$$

k_b is the harmonic bond force constant for covalently bonded atoms, k_a is the harmonic angular bending force constant, V_n is the barrier height of the torsion, n is the periodicity of the torsion, γ

is determines the phase of the torsion, and A and B are Van der Waals parameters. The harmonic approximations made in this potential are satisfactory for states which do not significantly deviate from their equilibrium values. For the sake of computational efficiency, several other approximations are typically made. For example, a cutoff distance is set for Van der Waals interactions because they decay at a relatively fast rate. Coulomb interactions, on the other hand do not decay rapidly and a simple truncation is not acceptable. This is an important problem for systems with periodic boundary conditions because the coulombic potential would diverge. To deal with this problem, Ewald summation was introduced. The concept of Ewald summation [128] is as follows: Each charge is neutralized by superimposing upon it a spherical Gaussian cloud of the opposite charge. The interaction of these charges can now be easily solved in real space. Next, a second set of Gaussians is introduced, opposite in charge to those previously added. The potential arising from these Gaussians may be solved in reciprocal space with Fourier series. By splitting the problem in this manner, the sum becomes tractable for periodic systems.

1.4.2 Poisson-Boltzmann Theory

Electrostatic interactions are important in molecular modeling, especially for in biomolecular systems. On average, over 20% of all amino acids in globular proteins are ionized in vivo and an additional 25% are polar [129]. For DNA, each base pair carries a charge of $2e^{-}$, making it a highly charged system, heavily regulated by electrostatics. Ideally, electrostatics would be modeled explicitly for each charged or polarizable component in a system; however, for many systems, such precision is either unneeded or intractable because of limited computational resources. For these systems, use of mean field techniques such as Poisson-Boltzmann theory (P-B) can provide useful qualitative insight and often quantitatively accurate results. P-B theory is a powerful method which uses the Poisson equation of classical electrostatics to determine the potential around an object and uses a Boltzmann weighting to describe the distribution of

mobile counterions surrounding that object. A good review of the history of the development of P-B theory was was published by Fogolari *et. al.* [130]. Briefly, P-B theory was independently developed by both Gouy and Chapman in the early 1910s [131, 132]. The equation was later generalized by Debye and Hückel in the 1920s who applied its linearized form to the to the study of ionic solutions [133]. Onsager studied the statistical mechanical basis of the equation cast in terms of potential of mean force [134] and Kirkwood subsequently noted that P-B theory made the assumption that it was possible to replace the potential of mean force with the mean electrostatic potential [135]. Onsager also conceived of counterion condensation, which was later popularized by Manning [136], using concepts from the Debye-Hückel method. The method was gradually extended to systems beyond simple ionic solutions, but for the most part, the work was still based on simplistic geometries. DNA, for example, could be approximated as a linear charged rod. It was not until the 1980s that methods were developed to solve the P-B equation for an arbitrary shape. At the present, P-B theory is commonly used to study the electrostatics of large irregularly shaped biomolecules and can be used as a comparison with MD simulation results.

Starting from the Poisson equation:

$$-\varepsilon\nabla^2\phi(\mathbf{x}) = \rho(\mathbf{x}) \tag{1.2}$$

The divergence of the gradient of the electrostatic potential field is equal to the charge density. The charge density itself can be written as:

$$\rho(\mathbf{x}) = \sum_i Z_i e c_i(\mathbf{x}) + \rho_{ext}(\mathbf{x}) \tag{1.3}$$

In this expression, the first term corresponds to mobile ions and the second term corresponds to charges provided by a fixed or external source. If we now assume that the mobile charges are distributed by Boltzmann weight, we arrive at the expression for the P-B equation:

$$-\varepsilon\nabla^2\phi(\mathbf{x}) = \sum_i Z_i e c_{0,i} e^{\frac{-Z_i e\phi(\mathbf{x})}{k_b T}}(\mathbf{x}) + \rho_{ext}(\mathbf{x}) \tag{1.4}$$

To represent a system in this way we must make several initial assumptions: First, only direct charge-charge interactions are considered, permanent and induced dipole effects are ignored. Next, all counterions are considered point charges and there are no short range force effects. Finally, water is a continuous dielectric medium that can be simply described with a dielectric constant. Of course real atoms and molecules are polarizable and their electron density will distort in the presence of an external electric field, and they have a finite size. Additionally, at short range water retains some structure. These assumptions are all reasonable under dilute conditions. In general, this means that highly charged systems with many counterions such as DNA are not well suited for treatment with P-B calculations. It is important, however, to remember these limitations as they may play an important role in interpreting results. For example, in Chapter 3, the study of the electrostatics of the nucleosome, we find that molecular dynamics simulations predict three distinct peaks in the plot of $g(r)$ for Na+ while P-B theory only predicts one single peak. This is because P-B theory neglects the finite size of Na+ ions.

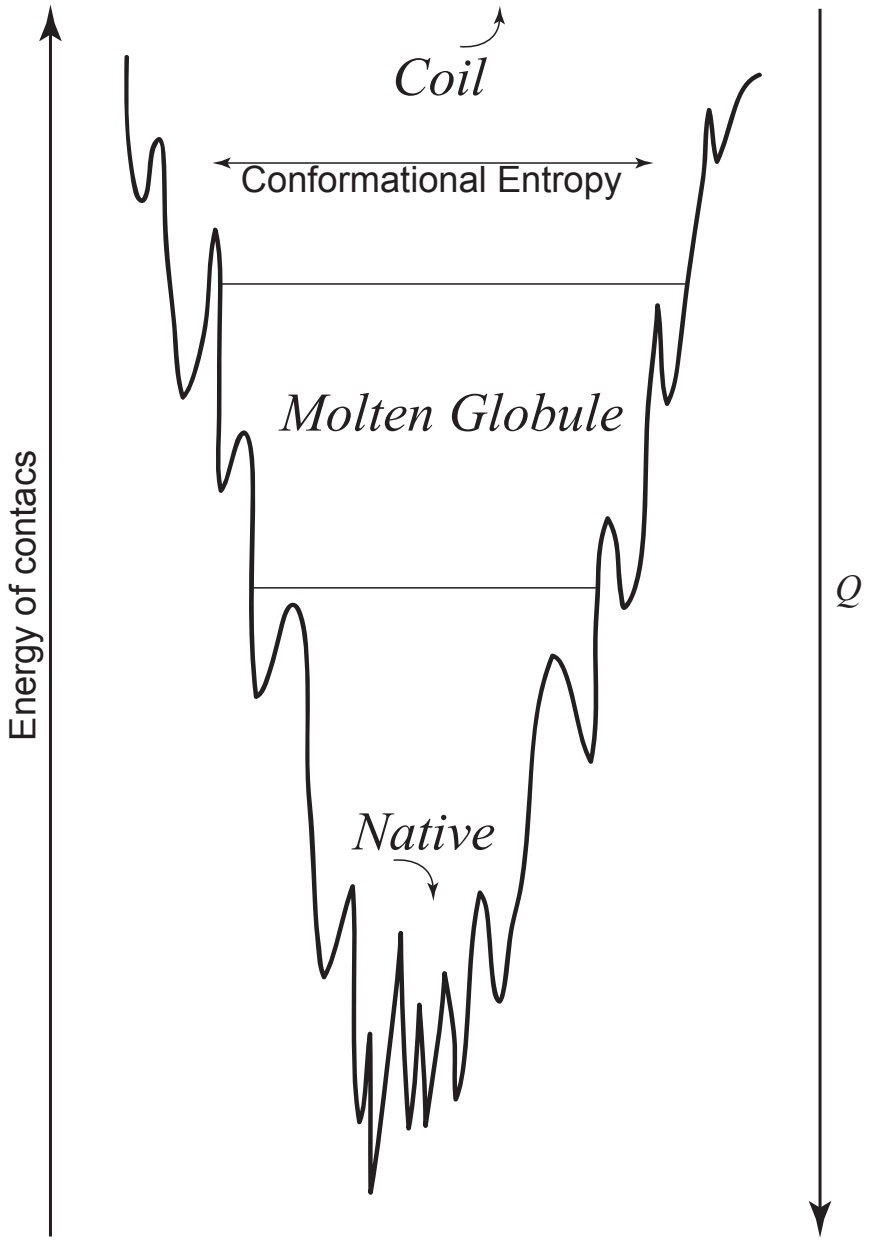


Figure 1.1: A funneled energy landscape of a typical protein depicting the major structural and dynamical regimes. The order parameter Q, characterizing similarity of the given conformation to the native structure, increases from 0 to 1 as the protein descends into the native state of the funnel.

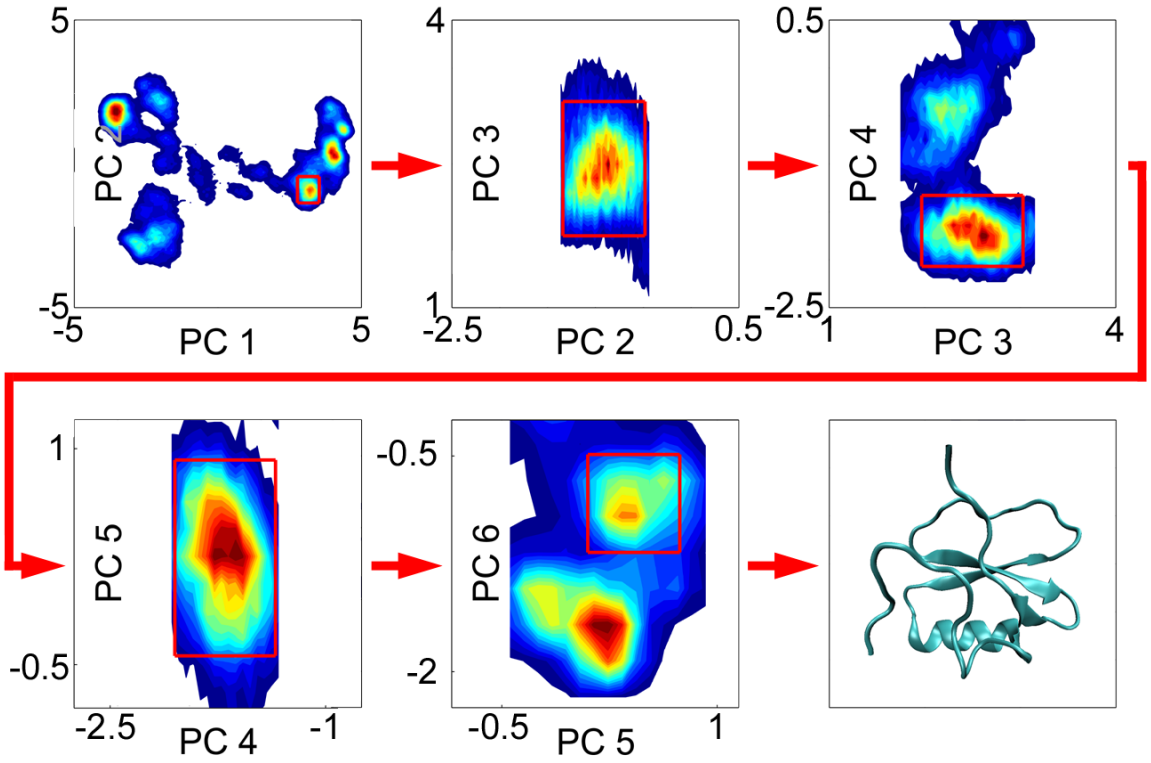


Figure 1.2: We isolated each basin in our 2-D PC landscapes and projected them into the next set of PC's. This had the effect of further separating the data into more basins which themselves were each isolated and projected in subsequent PCs. At the end, we arrived at a basin whose structure can be well characterized.

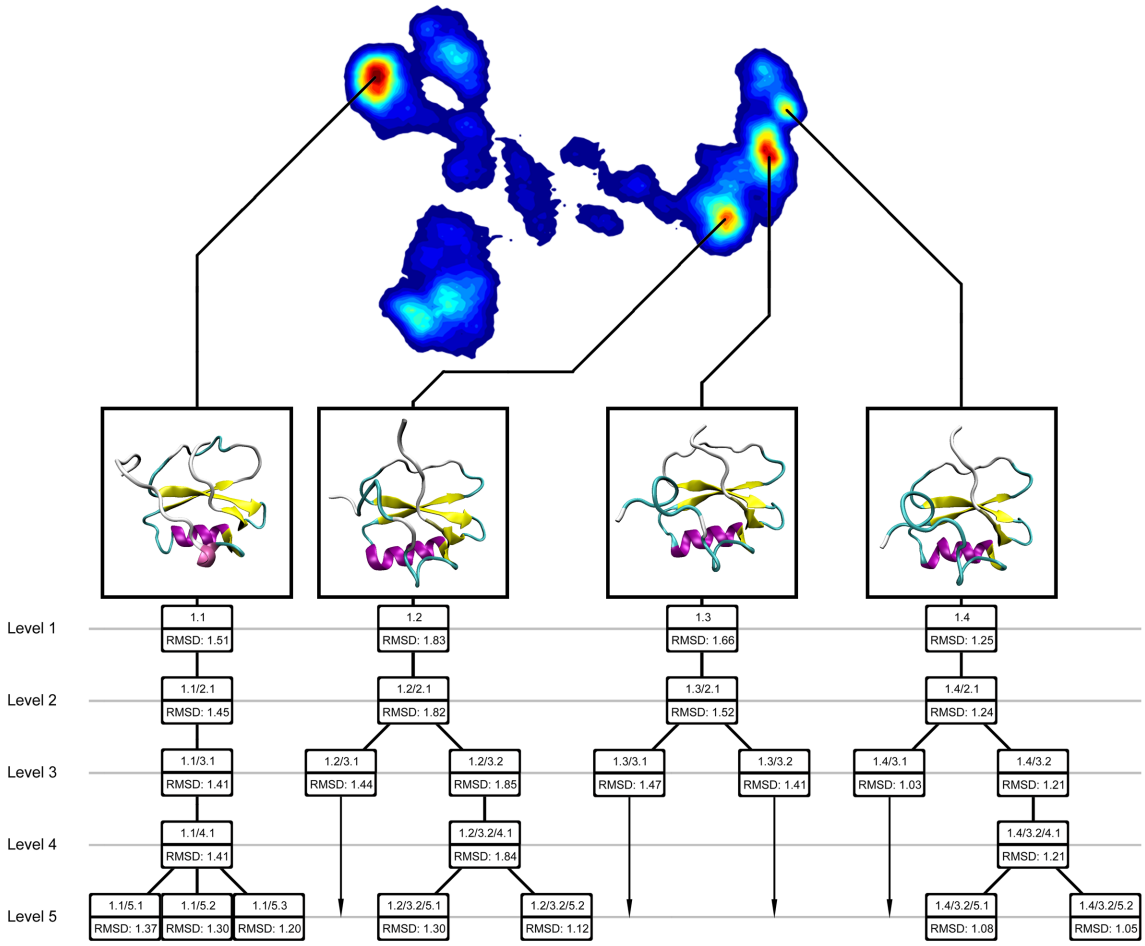


Figure 1.3: Tree of basin hierarchy is shown. Branches terminated before the 5th level could be continued, however, no further separation of those branches was found within the range of PCs investigated.

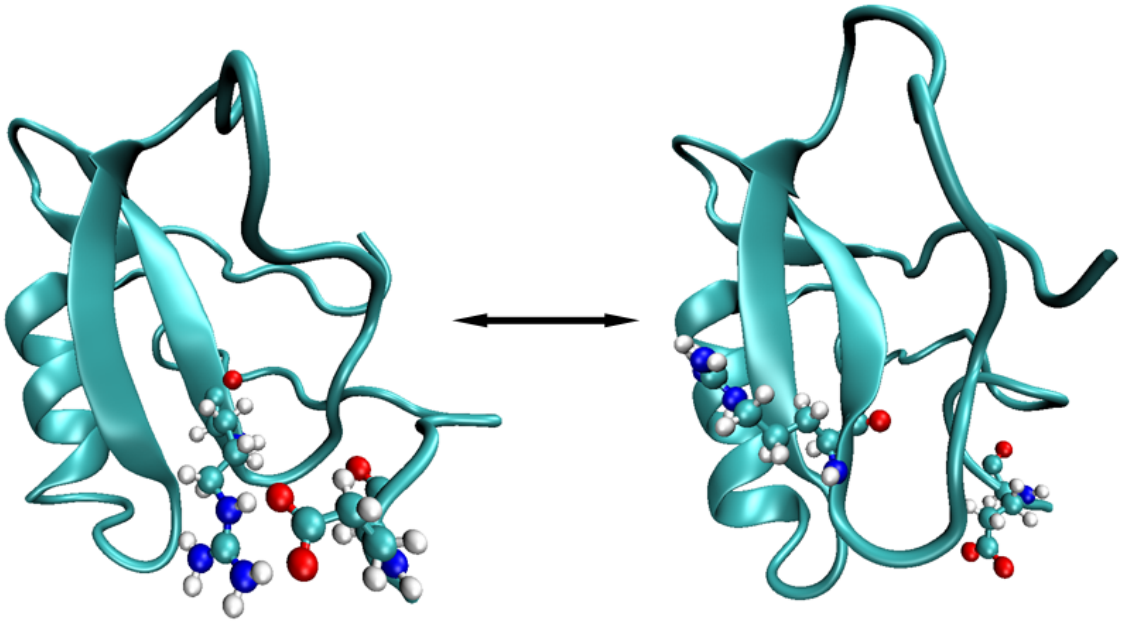


Figure 1.4: Upon basin splitting, significant changes in inter-residue contacts would typically occur. This figure shows a representative divergence of contacts between two basins. Structures from basins 1.1 and 1.4, shown here, display an Arginine - Glutamic acid salt bridge broken.

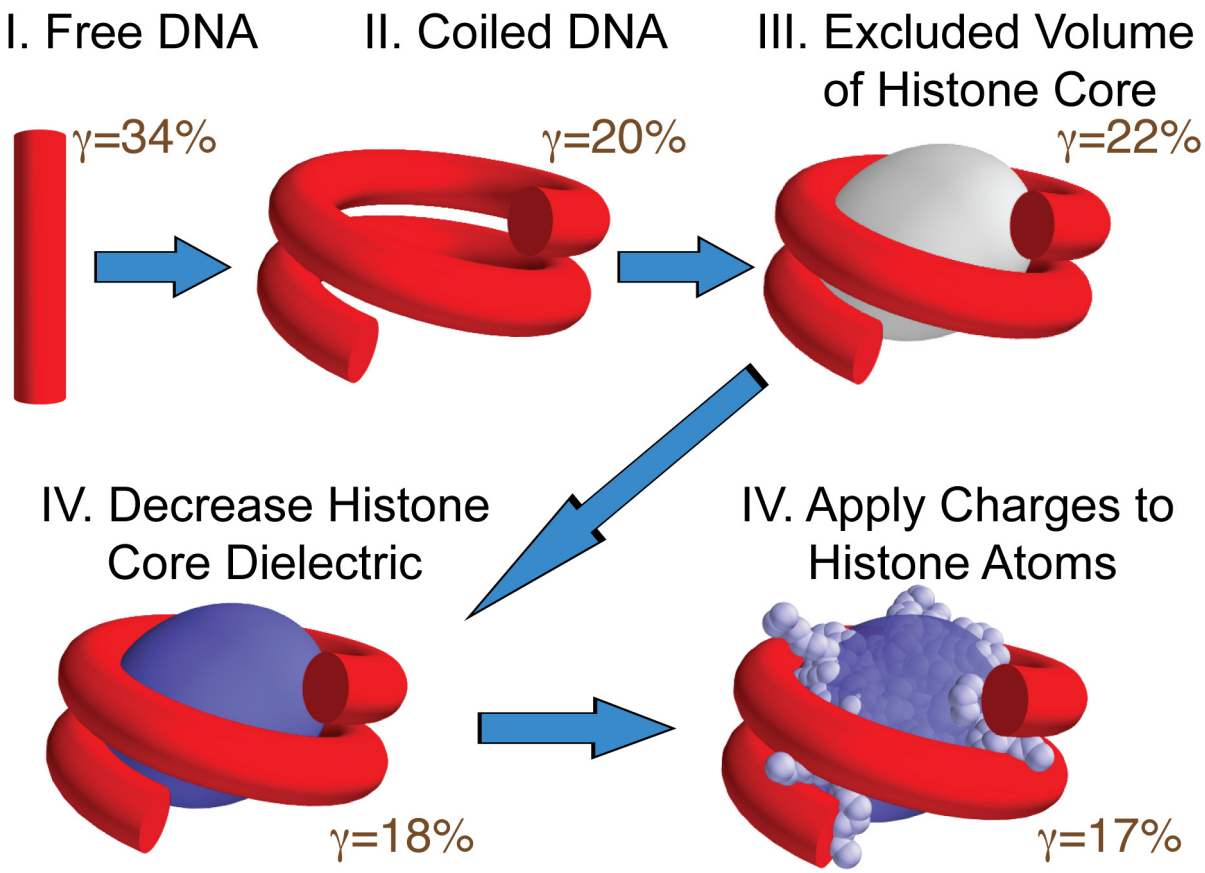


Figure 1.5: This figure shows a conceptual step by step nucleosomal assembly pathway for understanding the effects of nucleosome formation on counterion condensation.

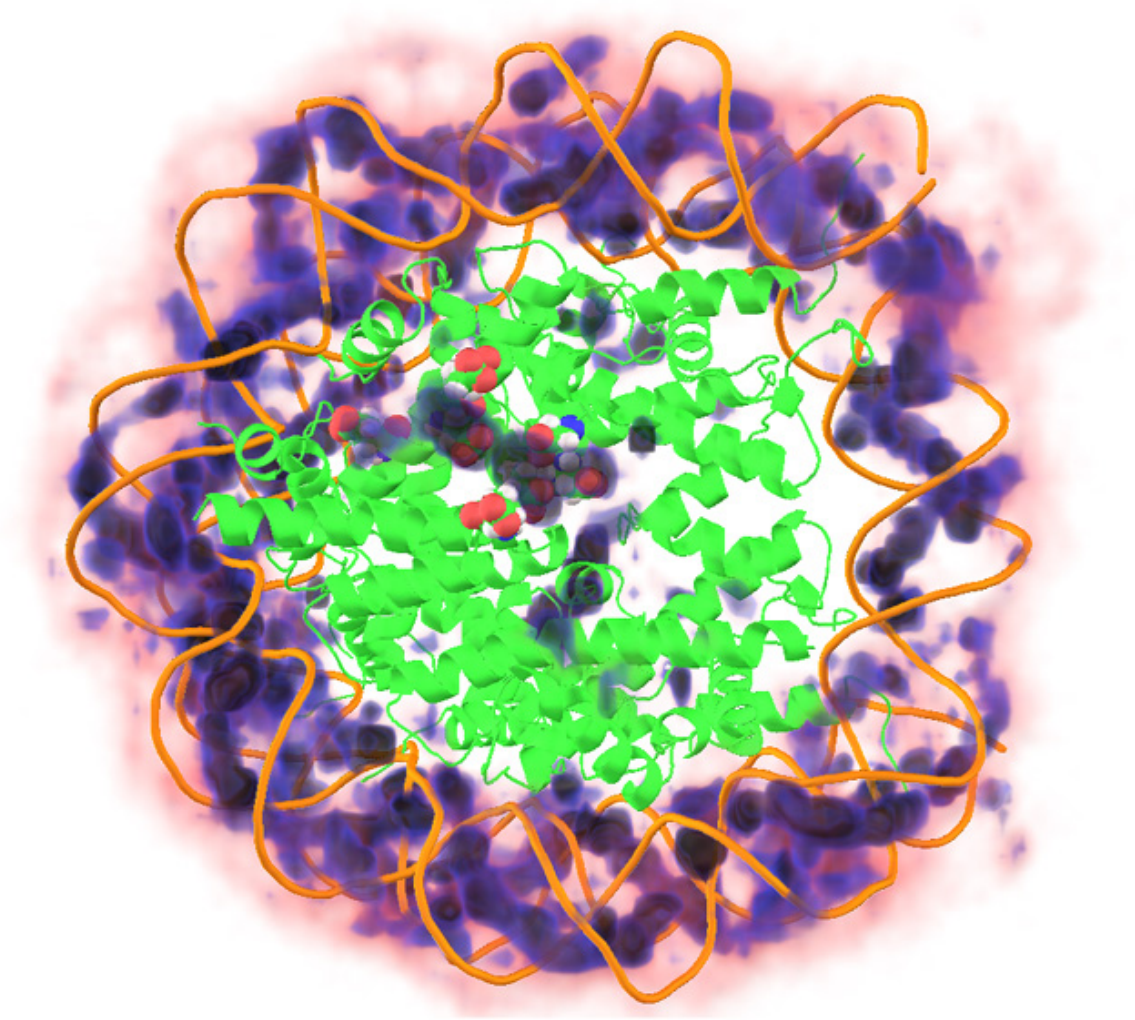


Figure 1.6: A 3-dimensional distribution of sodium around the nucleosome. The crystallographically identified acidic patch has been highlighted as spheres on the surface of the histone core and a high level of sodium condensation is observed around these residues.

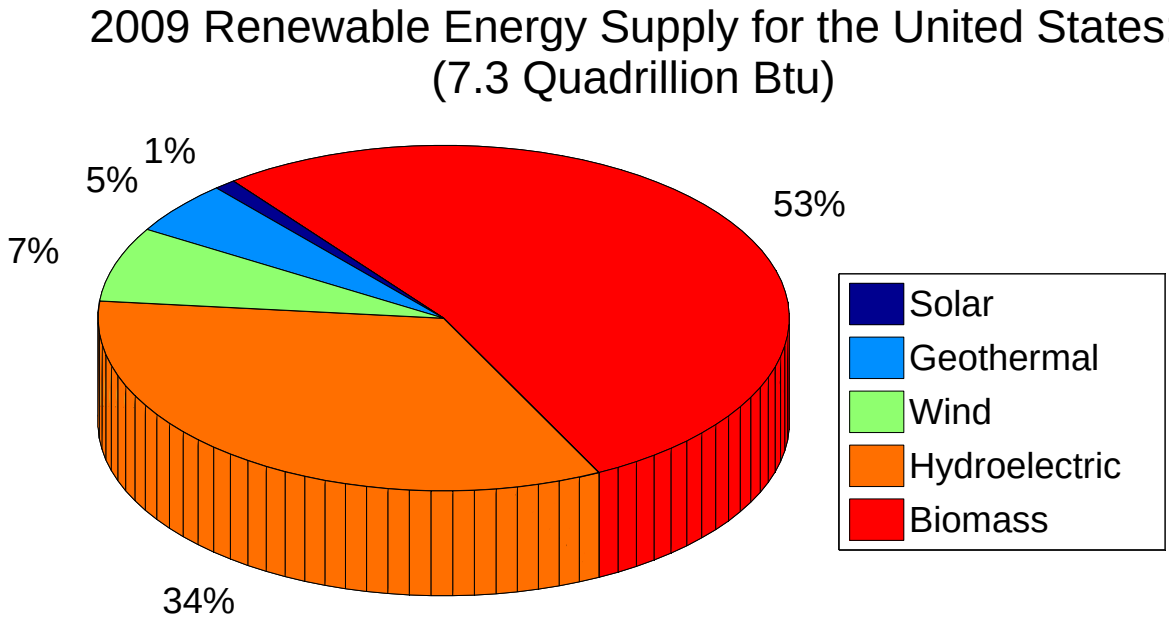


Figure 1.7: Summary of the renewable energy resources of the United states in 2009.

Chapter 2

Hierarchical Organization of Eglin C Native State Dynamics Is Shaped by Competing Direct and Water-Mediated Interactions

Abstract

The native state dynamics of the small globular serine protease inhibitor eglin c has been studied in a long 336 ns computer simulation in explicit solvent. We have elucidated the energy landscape explored during the course of the simulation by utilizing Principal Component Analysis. We observe several basins in the energy landscape in which the system lingers for extended periods of time. Through an iterative process we have generated a tree like hierarchy of states describing the observed dynamics. We observe a range of divergent contact types including salt bridges, hydrogen bonds, hydrophilic interactions and hydrophobic interactions, pointing to the frustration between competing interactions. Additionally, we find evidence of competing water-mediated interactions. Divergence in water-mediated interactions may be

found to supplement existing direct contacts but they are also found to be independent of such changes. Water-mediated contacts facilitate interactions between residues of like charge as observed in the simulation. Our results provide insight into the complexity of the dynamic native state of a globular protein and directly probe the residual frustration in the native state.

2.1 Introduction

Globular proteins are characterized by a funnel-like energy landscape with a deep minimum associated with the native state [3, 6, 137–140]. The native state is somewhat degenerate and possesses a rugged energy landscape as a result of residual frustration between subtle competing structural conformations. [3, 6, 56, 139, 140]. This frustration may arise in part from the competition of possible inter-residue contacts which are explored during the protein's dynamics and can be paralleled with residual entropy in spin glasses [3,139,140]. Under physiological conditions, much of the native landscape is thermally accessible, resulting in incessant fluctuations between available states [3, 6, 138–140]. Hence, understanding the organization of the local minima in the native state energy landscape is vital to our understanding of protein function, such as in allosteric proteins and enzymes [61,141,142]. The dynamic nature of the native state is created by a complex interplay of protein and solvent degrees of freedom [143]. The complexity is simplified in part because a folded protein possesses a core set of stable contacts which are responsible for maintaining the native structure thereby reducing the number of possible degrees of freedom. The dynamical fluctuations within the remainder of the protein leads to the formation of competing transient contacts, which in turn, leads to frustration [56].

The importance of the solvent on native state stability and dynamics is widely appreciated. It is well known that water is essential for the stability of many globular proteins, since hydrophobic collapse is considered to be one of the primary driving forces in both the process of creating and maintenance of the folded structure [87, 143–145]. Additionally, water is known to play important site specific structural roles and some tightly-bound waters have been detected through crystallography [143, 146–149]. It has been shown that water plays a vital role in mediating hydrophilic contacts and inclusion of these effects in force fields has led to improved protein folding structural predictions [85]. Direct interaction of hydrophilic residues may result in a large desolvation penalty, consequently, an interaction through a water bridge may become preferable [22, 85]. Although it has long been appreciated that direct and water-

mediated interactions sculpt the protein's native state hierarchy, how this occurs in practice is often unclear. The central goal of this work is to address this question.

In this work we study the long timescale dynamics of the protein eglin c using molecular dynamics simulation. Eglin c is a small 70-residue serine protease inhibitor found naturally in the leech *Hirudo medicinalis* whose dynamics have been studied extensively by NMR relaxation experiments [141, 150–154]. Revealing the organization of the energy landscape of eglin c is the first goal to be accomplished in this work. The native state energy landscape is best thought of in terms of a hierarchy of similar conformational states, arranged in a tiered fashion [3, 137–140, 155]. Significant work has been done to characterize the energy landscape for peptides, using, for example, disconnectivity graphs (DG) [72, 74]. DGs provide important insight into finer scale splittings of like structural clusters and provide valuable information about the energetic barriers of the system. Constructing these graphs for a protein solvated in explicit water is less straightforward, though a low-resolution DG has been created by Tarus *et al.* through use of a coarse grained contact clustering algorithm [156]. In the spirit of DG, though without the detailed energetics, we have characterized the energy landscape of the native state using principal component analysis (PCA). We have employed PCA to examine the effective dimensionality of the energy landscape and obtain a reduced space in which to study our system. In prior works, low dimensional reaction coordinates have been shown to be effective in describing protein folding processes [157–159]. The obtained PC space was used to isolate highly preferred protein conformations generated from our simulation. The preference for the system to adopt certain configurations is an indication of local basins in the energy landscape around these configurations. In this way, we have used PCA to unmask a hierarchy of states adopted by the protein within the scope of our simulation. Once a topology of the energy landscape has been obtained, it is possible to search for the sources of the attractors within individual basins. We will show that the observed hierarchical splittings can be characterized by divergent sets of direct inter-residue and water-mediated contacts formed. Preferred contacts

are naturally passed down through the hierarchy, and branching results in finer segregation of structures and contacts, similar to ultrametric patterns [81]. We find evidence of salt bridges, hydrophobic, and hydrophilic transient contacts all contributing to basin definition. In addition, we observe water-mediated interactions forming unique indirect inter-residue contacts in addition to helping to facilitate direct contacts.

2.2 Methods

2.2.1 Simulation Details

An all-atom molecular dynamics simulation of wild type F10W eglin c was performed using the CHARMM27 protein-lipid force field [126] and the NAMD program suite [160]. The simulation was performed in 7758 explicit TIP3P water molecules under periodic boundary conditions. The charge of the protein was neutralized by counter ions, followed by the introduction of additional counter and co-ions to reproduce cell concentrations. The first of 25 NMR structures found in the eglin c PDB entry 1EGL was taken as the initial structure [161]. The tenth residue of the protein was mutated from phenylalanine to tryptophan in order to more closely mimic the conditions used in previous NMR studies(F10W) [141].

Prior to the commencement of the simulation, the system underwent a series of minimization steps. At constant volume, the entire protein was frozen in place and the water and ions were minimized for 10,000 steps. The protein side chains were then unfrozen and the system was minimized for an additional 10,000 steps. Finally, all constraints were removed from the system and it was minimized for an additional 20,000 steps.

The simulation proceeded with 2 fs time steps using the SHAKE algorithm and Ewald summation for long-range electrostatics. Short-range non-bonded interactions were calculated at each step and long-range interactions were calculated on even steps only. The pair list was updated every 10 steps. System coordinates were saved every 500 steps (1 ps) for later

analysis. At constant volume, the system was gradually heated via Langevin dynamics to 300 K in incremental steps of 5 K every 5 ps. Upon completion of the heating steps, the constant volume constraint was released, and the pressure was moderated by Langevin piston (set to 1 atm). The system was allowed to evolve under these conditions for 16 ns to allow some relaxation time. Data collection followed for an additional 336 ns for a total of 352 ns of simulation time.

2.2.2 Comparison with experimental results

Lipari-Szabo model free order parameters S^2 (backbone) and S^2_{axis} (side chain) values were calculated to show agreement of the simulation with experimentally derived dynamics [162]. Lipari-Szabo analysis is a common technique used to evaluate dynamics by examining the flexibility of bond vectors. The simulation was divided into 8 ns windows and the appropriate normalized bond vectors (backbone N-H and side chain terminal C-CH_3 respectively) were input into equation (2.1).

$$S^2 = 3/2\left[\langle x^2\rangle^2 + \langle y^2\rangle^2 + \langle z^2\rangle^2 + \right] + \left[2\langle xy\rangle^2 + 2\langle xz\rangle^2 + 2 + \langle yz\rangle^2\right] - 1/2 \tag{2.1}$$

In this equation, x,y, and z are the Cartesian components of the unit vectors that describe the direction of the selected bond. Experimental values for both backbone and side chain order parameters were kindly provided by Dr. Andrew Lee from the UNC Department of Pharmacy.

2.2.3 Principal component analysis

Principal component analysis (PCA) was used to simplify the analysis of the trajectory obtained from the MD simulation by redefining the phase space through an optimal linear transform such that the first few PCs contain most of the variance in the data. It has also been shown that

over 90% of dynamics can be captured by about 5% of the degrees of freedom [66, 163–167]. This promising prospect suggests that much of the high-dimensional phase space theoretically available to proteins may contain little to no interesting information because of constraints on those degrees of freedom. In this study, PCA of the internal coordinates defined by the backbone dihedral angles ϕ and ψ and the side chain dihedral angle χ_1 was performed. Dihedral angles are more appealing than raw Cartesian coordinates because they naturally lend themselves to physical interpretation and are independent of global translation or rotation of the molecule [79]. The susceptibility of Cartesian coordinate to produce artifacts has been shown by Mu *et al.* [168]. In order to eliminate discontinuity problems associated with angular coordinates 0 / 2π, each angle was instead defined by its sine and cosine components [79]. PCA is carried out by diagonalizing a covariance matrix **M** defined in equation (2.2)

$$\mathbf{M} = \left\langle xx^T \right\rangle \tag{2.2}$$

where x is the trajectory defined in terms of the appropriate trigonometric components of the ϕ, ψ, χ_1 dihedrals. Diagonalization of **M** produces a set of eigenvectors **u** which is a redefinition of the k degrees of freedom available to the system. Eigenvectors are sorted by descending eigenvalues representative of the variance of the data in that dimension so that the greatest variance lies along the vector $\mathbf{u}_1$ and the least along $\mathbf{u}_\mathrm{k}$. It is desirable to project the original trajectory data $x(t)$ into PC space to make use of this more efficient coordinate system in which all linear data correlations have been removed. Once the data is projected in to PC space, it is possible to determine which PCs should be considered essential degrees of freedom and which contain essentially Gaussian noise. PCs required for characterization of the essential dynamics possess multi-peaked probability distributions within that degree of freedom in which the peaks represent the multiple states in which the protein can explore and single peaked, Gaussian distributions conversely represent degrees of freedom that describe fluctuations within a single

state [66].

2.2.4 Identification of water-mediated contacts

In this study, water-mediated contacts are defined as water molecules, which simultaneously form hydrogen bonds with two or more residues. The water must meet distance and orientation requirements defined by Raiteri *et al.* [169]. Since the Raiteri *et al.* study considered only water-water hydrogen bonding, some additional requirements were necessary to correctly apply their method to proteins. In our work, any water whose oxygen atom came within 3.5 Å of an oxygen or nitrogen within the protein was added to a list of potential contacts. Selected waters were then tested for additional orientation and distance compliance. The equation used to evaluate this compliance was defined in Raiteri *et al.* work and can be shown in Equation (2.3) [169].

$$f(d) = \frac{1 - [(d - d_0)/\Delta]^n}{1 - [(d - d_0)/\Delta]^m} \tag{2.3}$$

This formula was used for both the distance and orientation requirements with a different set of parameters for each case which were also defined in Raiteri *et al.*. In the distance case, $d_0 = 2.75$, $\Delta = .45$, $n = 10$, $m = 16$, and d_0 is the $\overline{X_1X_2}$ (X = N or O) distance between the protein oxygen or nitrogen and the water oxygen with a cutoff of 3.5 Å or more. In the orientation case, $d_0 = 0$, $\Delta = .4$, $n = 4$, $m = 8$, and d_0 is the $\overline{X_1H} + \overline{X_2H} - \overline{X_1X_2}$ distance between the protein oxygen or nitrogen and the water oxygen with a cutoff of d = .5 Å or more. Additional care was taken to ensure that the formula was applied to the appropriate atoms that would be needed to confirm additional orientation requirements needed in the protein system. Waters simultaneously hydrogen bonded to 2 or more residues are counted as water mediated contacts.

2.3 Results

2.3.1 Comparison with experimental results

Lipari-Szabo model free S^2 analysis [162] was performed on the simulated trajectory to verify agreement of dynamics from the MD simulation with experiment. Quantitative agreement was found between simulated and experimentally derived backbone order parameters for most residues in the protein (Figure 2.1). Excellent agreement is shown for the residues of the N-terminus and the entire core of the protein. Some deviations from experiment are found in several residues of the flexible binding loop and a significant reduction in S^2 is shown in the last three residues of the C-terminus suggesting increased mobility in our simulation. Finding quantitative agreement of S^2_{axis} side chain order parameters obtained from simulation and those obtained from experiment is typically more difficult [154, 170, 171]. Our simulation results show a semi-quantitative agreement with those from experiment (data not shown).

2.3.2 Principal Component Analysis

PCA is a powerful linear technique used to aid in the comprehension of complex multidimensional systems by reducing the phase space while retaining essential degrees of freedom. On the other hand, using PCA on an MD trajectory has the potential to obscure some of the complexity of the system's dynamics. For example, blindly reducing the phase space to the first two PCs would average out important information about conformational substates contained in the higher order degrees of freedom. To circumvent this issue, we have classified each degree of freedom as belonging to either the essential or non-essential phase space. Characterization of the essential phase space begins by projecting the trajectory into PC space, then histogramming the data along each PC. In order to classify a PC as a member of the essential phase space, it is assumed that if the trajectory is projected onto a PC in the nonessential space, a histogram of this data will be Gaussian in nature, representing rapid fluctuation within a single

state. [66]. Additionally, it is assumed that a PC belonging to the essential subspace will possess multiple peaks representing the different states available in that degree of freedom [66]. A Gaussian was fitted to the histogrammed data by least squares analysis and using the R^2 coefficient of determination as a guide, we have created three levels of classification to describe the degree to which the histogrammed data may be considered Gaussian. Anything with a R^2 coefficient below .9 is considered to be Non-Gaussian, R^2 from .9-.98 the distribution retains some significant Non-Gaussian features and may represent a superposition of states which are poorly resolved in that specific dimension, distributions with R^2 from .98-1 are considered to be purely Gaussian in nature and are not considered part of the essential phase space. Representative histograms are shown in Figure 2.2. This analysis reveals that the first 4 PCs possess multi-peaked distributions with $R^2 < .9$ and next 7 PCs have distributions possessing R^2 values between .9 and .98. All consecutive PCs possess purely Gaussian distributions. These data suggest that the dynamics of eglin c over the course of this simulation should be described in at most an 11 dimensional manifold or roughly 3% of the total possible degrees of freedom.

2.3.3 Contact Analysis

The low dimensionality of the essential phase space permits the subsequent structural analysis. Two-dimensional histograms of the trajectory projected in the first two principal components may be examined to identify high density regions ($\geq$20% of maximum) within this space. These regions can be considered basins in the energy landscape and henceforth shall be referred to as such. It should be noted that the protein's dynamics are not tightly restricted to basin residence and much of the trajectory exists in more diffuse regions surrounding our tightly defined basins. There are a few additional highly diffuse regions which do not satisfy our definition of a basin and thus are not considered in this work. Data from the basins which meet our $\geq$20% of maximum criteria can be extracted from the full trajectory and then plotted in the space described by the second and third PCs in an effort to seek additional separation.

This process may be repeated until the incorporation of additional degrees of freedom no longer causes any basin splitting (Figure 2.3). Basin splitting should cease after the last essential dimension is reached, however, for purposes of maintaining sufficient sample size, our dimension expansion was cut off after 6 PCs. This cutoff ensured that all of the significantly multi-peaked dimensions were incorporated in addition to 2 possibly essential dimensions. The trajectory tends to enter a basin and linger for an extended period of time before moving on as opposed to rapid fluctuation between states. Originally, the full trajectory possessed an average RMSD of 3.01 Å; however, the average RMSD of each basin decreases as the levels of the hierarchy are traversed, implying that like structures are clustered and filtered into different branches as expected. The structural nature of these basins can be determined by extracting the associated frames from the rest of the trajectory. In principle, with sufficient sampling to ensure multiple basin visitations, it would be possible to construct a form of coarse grained DG similar to that found in Tarus *et al.* [156], and, subsequently, from the obtained DG postulate a master equation to describe transition dynamics. However, such unconstrained simulations would be computationally expensive at this time. Alternatively, a master equation may be formulated from computed free energy profiles between pairs of basins [172].

Average structures of each basin were computed based from the extracted frames and evaluated for defining structural features. The system as defined by basins from the first 2 PCs, adopts 4 globally distinct structures shown in Figure 2.5. Most of the divergent structural features which clearly characterize the differences in the 4 major basins unsurprisingly occur within the flexible binding loop and the solvent accessible N-terminal tail. A complete contact list was compiled for each basin in the hierarchy. Any non-nearest neighbor and non-next-nearest neighbor residues coming within 2.5 Å of one another were considered to be participating in contact. There are many divergent contacts between the four first level primary basins, so to simplify our consideration we focus on only the contacts formed by residues within the N-terminal tail and the flexible binding loop (For more detailed information on these contacts,

see the table included in the supplemental materials).

The majority of these transient competing contacts are hydrophobic in nature, however, there are many significant hydrogen bonds and several salt bridges observed. These data also show that the majority of the transient contacts are formed between the side chain atoms and to a lesser extent between backbone atoms or between the atoms of the side chains and backbone. Basin 1.1 is characterized by a collapse of the N-terminal tail on to the flexible binding loop and a coordinated migration of the C-terminal tail away from the core of the protein to form a hydrogen bond between L7 and H68 residues. None of the remaining basins possess significant interaction between the N-terminus and the binding loop, and the C-terminus remains closer to the protein core. Subsequent differentiation between the uppermost basins can be seen in the nature of the contacts possessed by the N-terminus. In basins 1.3 and 1.4, the N-terminus possesses a structure similar to a single turn of a coil linked by important contacts between the L7 residue and residues E2, F3, and G4 in addition to a contact between residues E2 and E6. This coil structure in the N-terminal tail is completely absent from basins 1.1 and 1.2. In this basin, the tail is extended and possesses no significant self-contacts. One significant contact, the side chain-backbone hydrogen bond between the backbone carbonyl group on L7 and the side chain amine hydrogen on W10, was found simultaneously in several basins. This is interesting because previous works [141] have stated that the W10 mutation from the wild-type F has negligible effects on structure. Even if this mutation does not produce any noticeable structural differences in the protein, there may be a significant unanticipated change in the dynamics.

Additional subtle differences between individual contacts formed may be found further along the hierarchy where it is possible to differentiate between structures that differ by as little as a single contact. With this type of analysis, individual local attractors become easily visible and are defined in a physically meaningful way. We hypothesize that the different direct contacts formed in each basin are representative of the inherent frustration in the energy

landscape of the protein, and serve to define the basin in which they reside. Interestingly, unexpected contacts were identified between residues of like charge. These contacts suggest that the solvent water may have played a role in mediating some interactions, thus, we evaluated this contribution to basin definition. In order to investigate these interactions, it was first necessary to establish the existence of persistent water-mediated contacts. Using a procedure developed by Raiteri et al. [169] for determining hydrogen bonds between liquid water molecules, a list of each residue and its respective hydrogen-bonded waters was created at each time step. This list was then scanned for waters which simultaneously participated in hydrogen bonds with multiple residues and these residues were counted as participating in a water-mediated contact. Results of this test indicated the presence of an abundance of significant water-mediated contacts in all basins. Some of these contacts are seen in over 90% of the frames associated with its particular basin. Specific waters typically had relatively low residence times of less than 500 ps but some persisted for over 10 nanoseconds. Across the 4 primary basins, there is a wide variation in the significant contacts detected, several of which are unique to one basin. Within each basin, many of the contacts remain conserved, however, there is a set of contacts that diverge upon further basin splitting. The increased structural refinement provided by mapping the basin in higher PC space reveals that the less resolved parent basin is in fact a collection of multiple states in which the individual water-mediated contacts exist in vastly different populations.

All of the contact populations were determined from samples containing over 100 snapshots; however, because of the ambiguity of determining appropriate error limits on the populations of water-mediated contacts, we used a somewhat arbitrary but reasonable definition of a significant population change of a contact. A population change was considered significant if there was a minimum of a 20% relative difference in its population between different substates. Water bridging was found to occur between side chain and backbone chemical groups alike. Water bridges are found to interact with residues independent of a direct contact in addition to mediating existing direct contacts. This mediation of direct contacts includes, but is not lim-

ited to, residues of like charge. Hydrogen bonding to water diminishes repulsion between the like charged residues, allowing them to maintain a direct contact. A representative example of the change in water-mediated contacts created by basin splitting are basins 1.4/3.1 and 1.4/3.2 which are shown in Figure 2.6. Details of each water bridge are shown above its population in Figure 2.6. This figure features the most prominent types of water bridges, between carbonyls (which reduces the repulsion of the like charges), over a direct hydrogen bond between an alcohol and a carbonyl, and at a salt bridge. Other existing water bridges are found between two alcohols or between an alcohol and a carbonyl in the absence of a direct contact. In some cases, a basin splitting may possess a significant reduction of the presence of a direct contact in one branch but maintenance of a water-mediated form in an opposing branch. Like direct contacts, some water-mediated contacts are binary in nature and can range from very strong to virtually non-existent in different branches of the free energy hierarchy, while others experience an attenuation associated with less favorable geometries.

We postulate that the basin splitting detected in this study is a manifestation of the inherent frustration possessed by the protein in the native state and that this frustration is exemplified by a number of distinct contacts which can be used to differentiate one basin from another. Figure 2.6 depicts a visual representation of the contact differences emerging from the splitting of basins 1.4/3.1 and 1.4/3.2. Several contacts which are shown to exist in specific regions of the protein in one basin are entirely absent in the other. Additionally, some of the significant identified contacts maintain their locality but swap the exact residues involved. In 1.4/3.1 and 1.4/3.2, R22, a residue in the α-helix forms a contact with the first β-sheet at either D33 or F36 respectively. Both direct and water-mediated contacts serve to increase structural stability of the protein and, barring an overpowering entropy loss, could explain the preference of the protein for structures within the basins identified in PC space.

2.4 Conclusion

The dynamics of wild-type eglin c has been simulated in order to obtain deeper insight into the energy landscape of the protein in its native state. The veracity of using molecular dynamics to model this system was bolstered both qualitatively and quantitatively by the results obtained by comparing simulated S^2 model free order parameters with those obtained from NMR experiment. We have, demonstrated the startlingly low dimensionality of the energy landscape occupied (roughly 3% of the total possible degrees of freedom) by the simulation over the course of a 336 ns simulation. Two-dimensional histogramming of the trajectory in the first two PCs reveals a set of basins in the energy landscape around which the system resides for an extended period of time. We have extended this observation by sequential two-dimensional examination of the trajectory in PC space with increasing PC index. By extracting only the data from a chosen basin and then viewing this data in the next indexed dimension, it is possible to separate large basins into finer scale basins which only become visible in higher dimensional space. This technique is only feasible because of the inherent low dimensionality of the active degrees of freedom when viewed in an efficient coordinate system. We have also observed, in agreement with NMR experiments, that the dynamical wealth of the protein resides primarily in the binding loop and the N-terminal tail. A contact list of the highly dynamic portions of this protein was created in an effort to understand the reason the system shows preference for these states. Snapshots of the frames residing within the basin of interest revealed a series of important contacts ranging in nature from salt bridges and hydrogen bonds to hydrophobic interactions. These contacts likely dictate the basin stability and hierarchy. There are a number of highly significant divergent contacts that are found to be important in the first of the basins. These contacts are also significant in the smaller subordinate basins, who themselves may be further differentiated by additional contact differences. In addition to direct contacts, we have also examined the role of water-mediated interactions in the energy landscape of a protein. We have found that there are a number of significant water mediated contacts formed by the pro-

tein throughout the simulation and that a subset of these contacts is unique to certain basins. Within a basin we have found that additional splitting observed upon expanding the data into higher dimensional PC vectors can be characterized by further segregation of both direct and water-mediated contacts. All subsequent basins found as we transcend the hierarchy retain the features of their ancestors. The results of this study exemplify the inherent residual frustration in the energy landscape of globular proteins and provides evidence of the importance of water mediated interactions.

2.5 Acknowledgments

This work was partially supported by the Arnold and Mabel Beckman Foundation Beckman Young Investigator Award. We thank Dr. Andrew Lee and Dr. Michael Rubinstein for stimulating and helpful discussions.

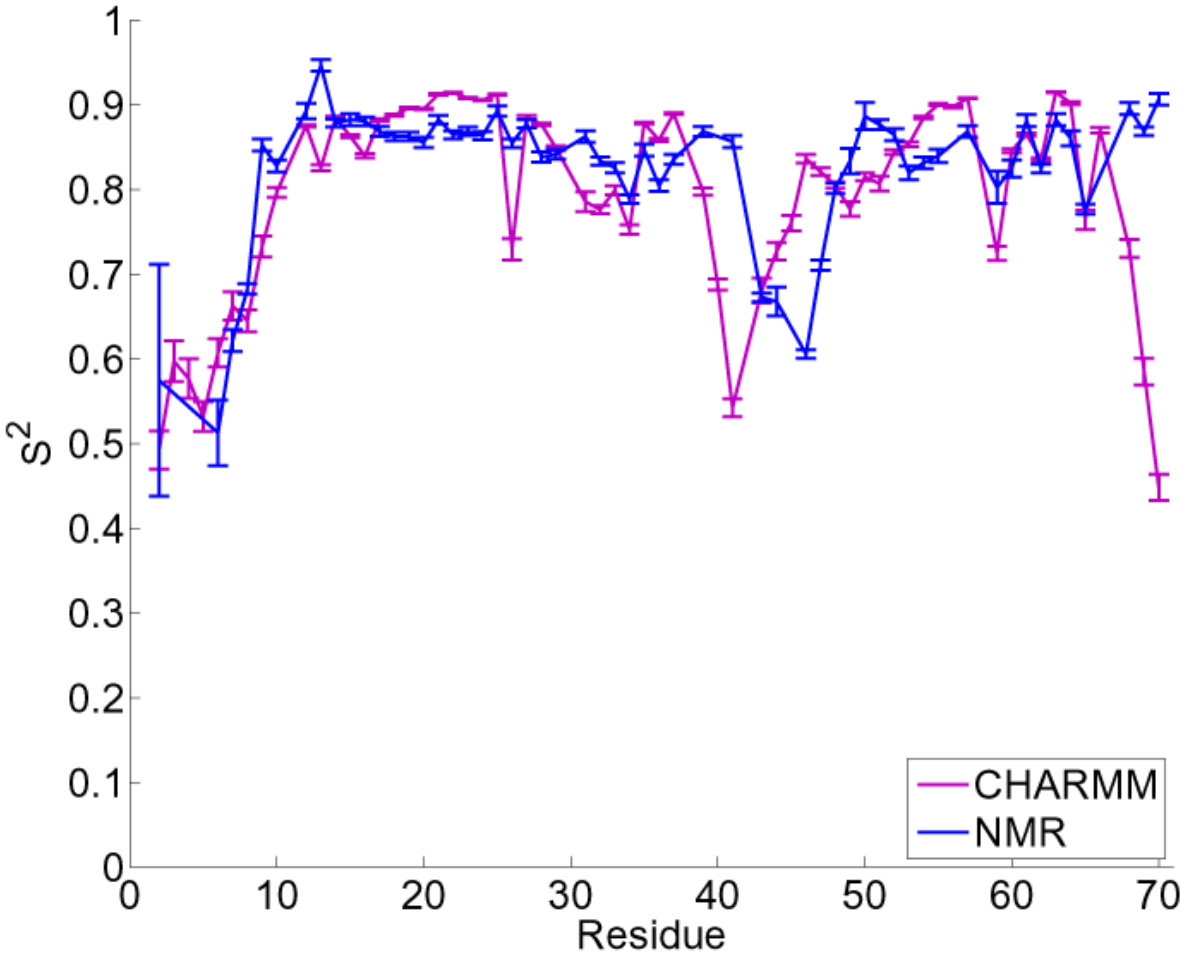


Figure 2.1: S^2 obtained from backbone N-H vectors. This figure shows a comparison of wild type eglin c from simulation and NMR experiments.

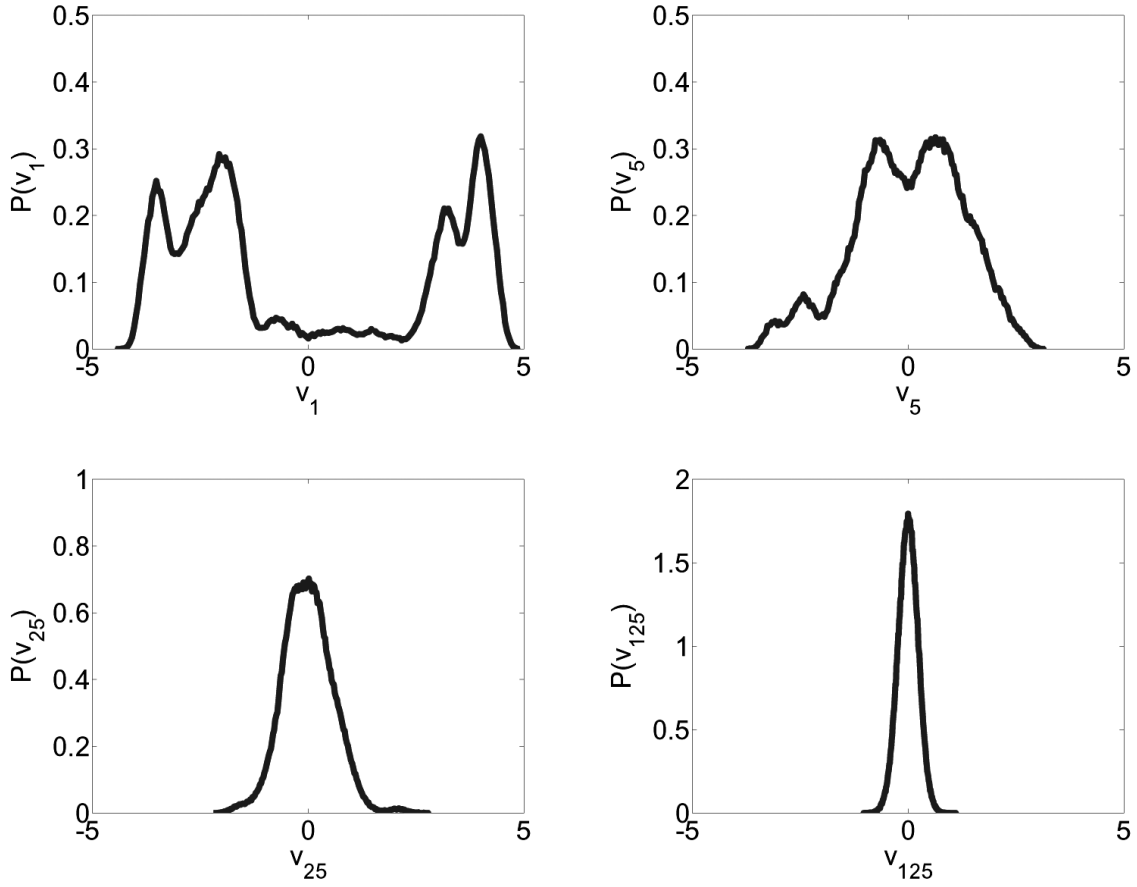


Figure 2.2: Probability distributions $P(v_i)$ for PCs 1, 5, 25 and 125 respectively. These histograms are representative examples of a non-Gaussian distribution (1), a primarily Gaussian distribution which maintains some nongaussian features (5), and purely Gaussian distributions (25, 125).

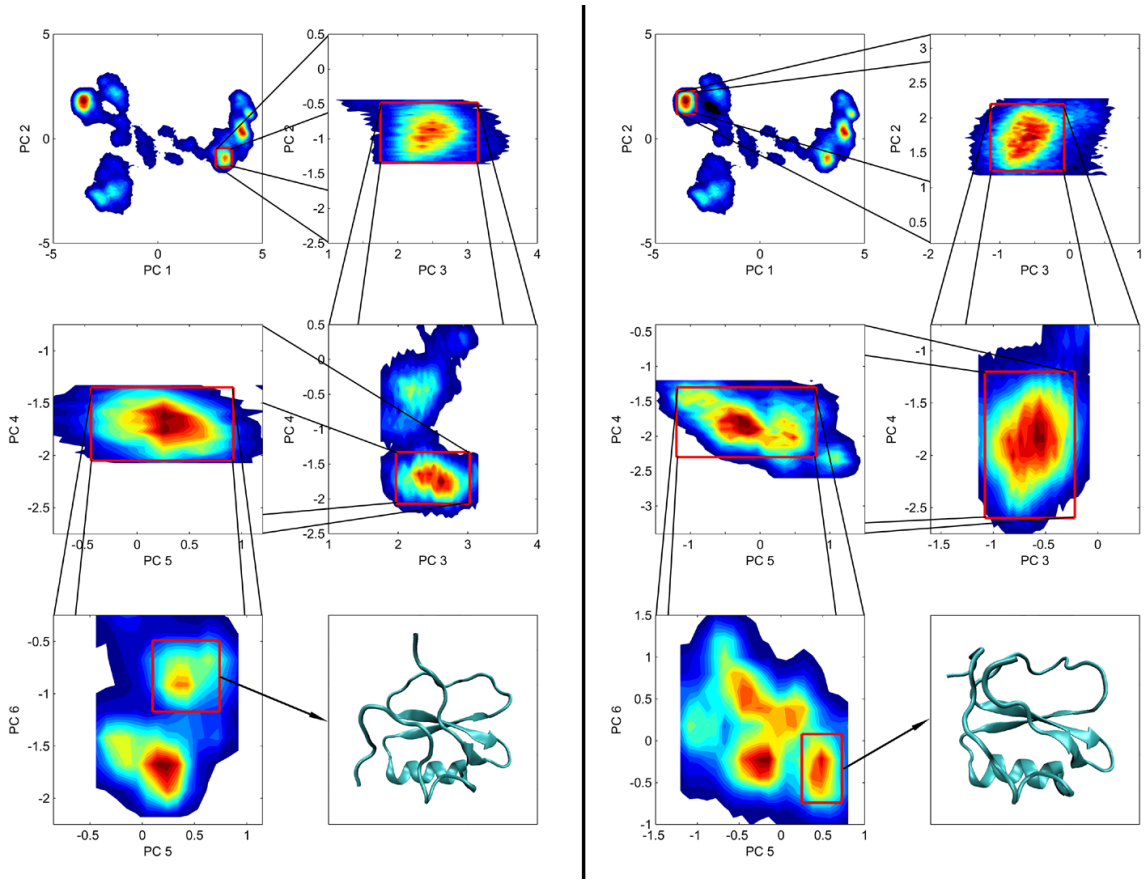


Figure 2.3: This figure shows the progression of the PC derived basin isolation through the 6th PC. The average structures are shown in the final boxes possessing RMSDs of 1.3 and 1.2 for the left and right respectively.

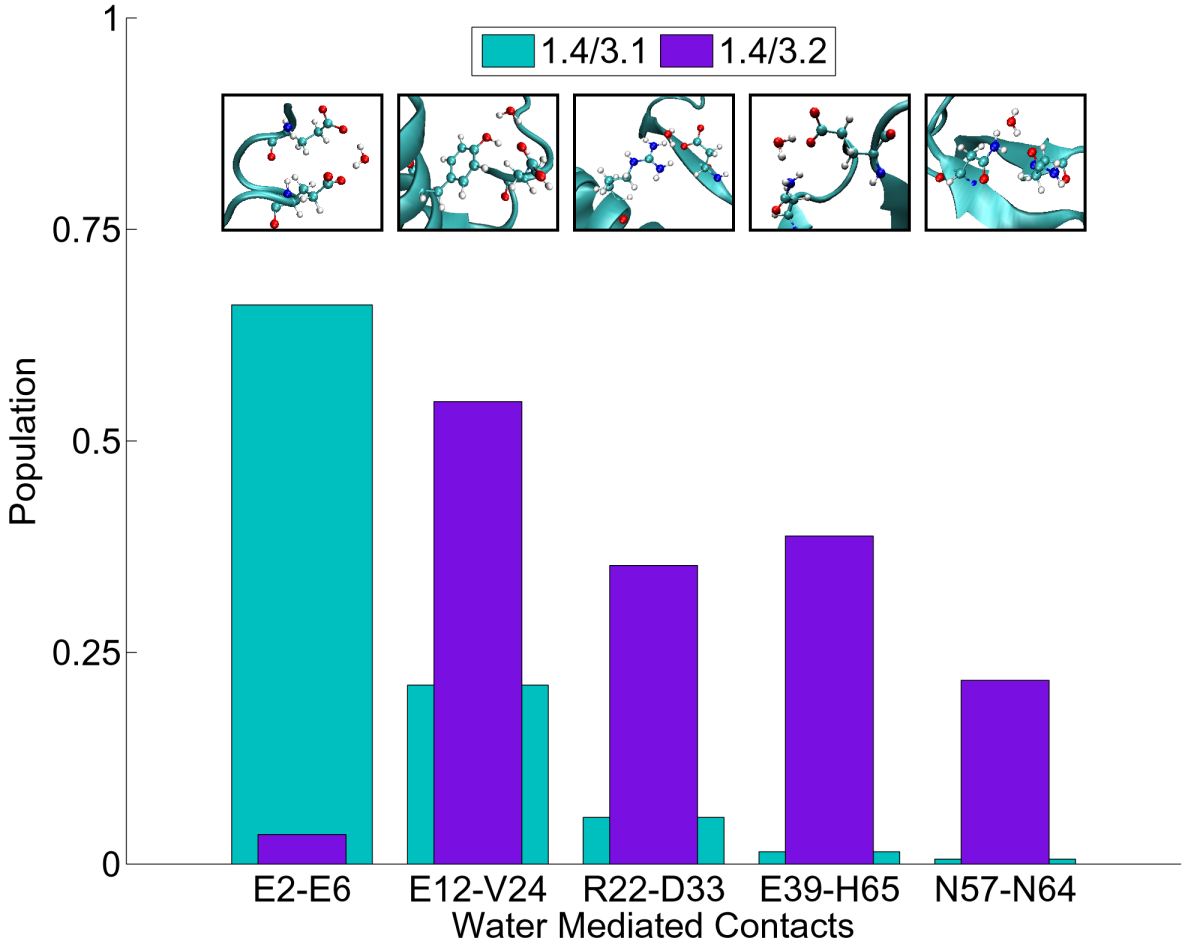


Figure 2.4: This figure depicts a representative example of how basin splitting segregates water mediated contacts.

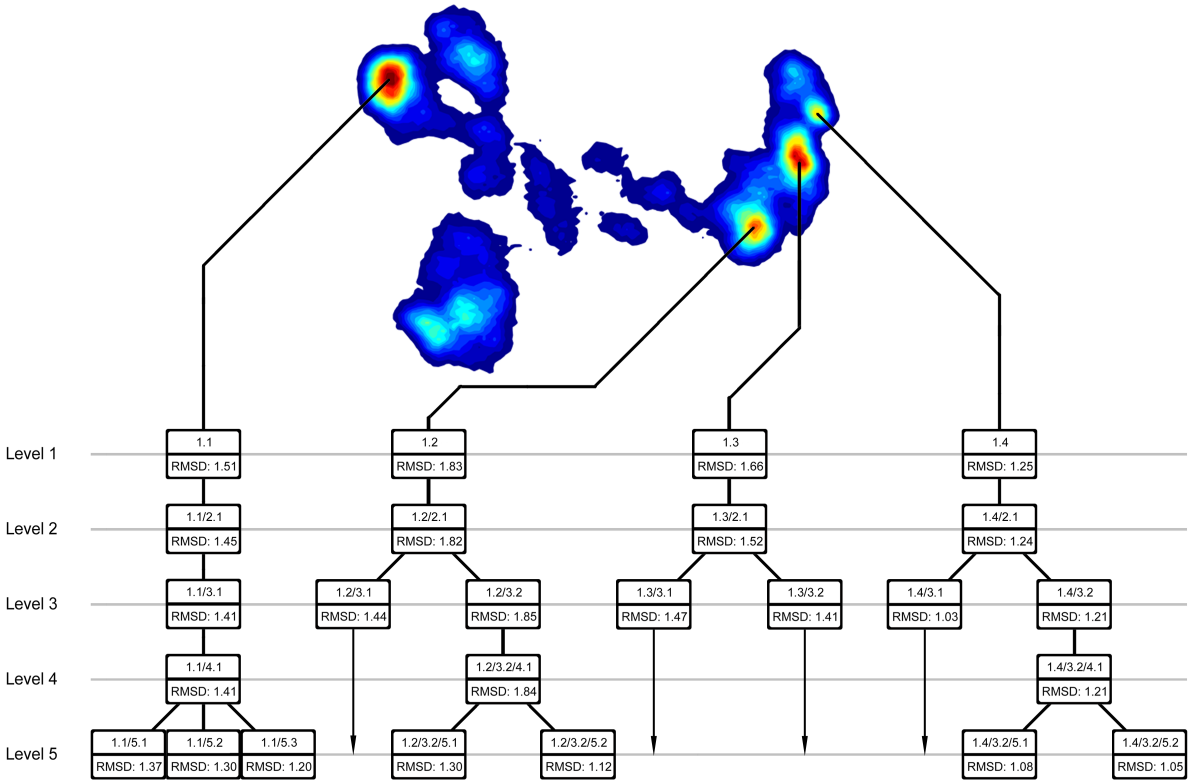


Figure 2.5: Tree of basin hierarchy. Branches terminated before the 5th level could be continued, however, no further separation of those branches was found within the range of PCs investigated. Location in the basin hierarchy is referenced by number in the format A/B/C/... where each letter represents a number X.Y where X indicates the level at which a basin splitting occurs and Y indicates the chosen branch.

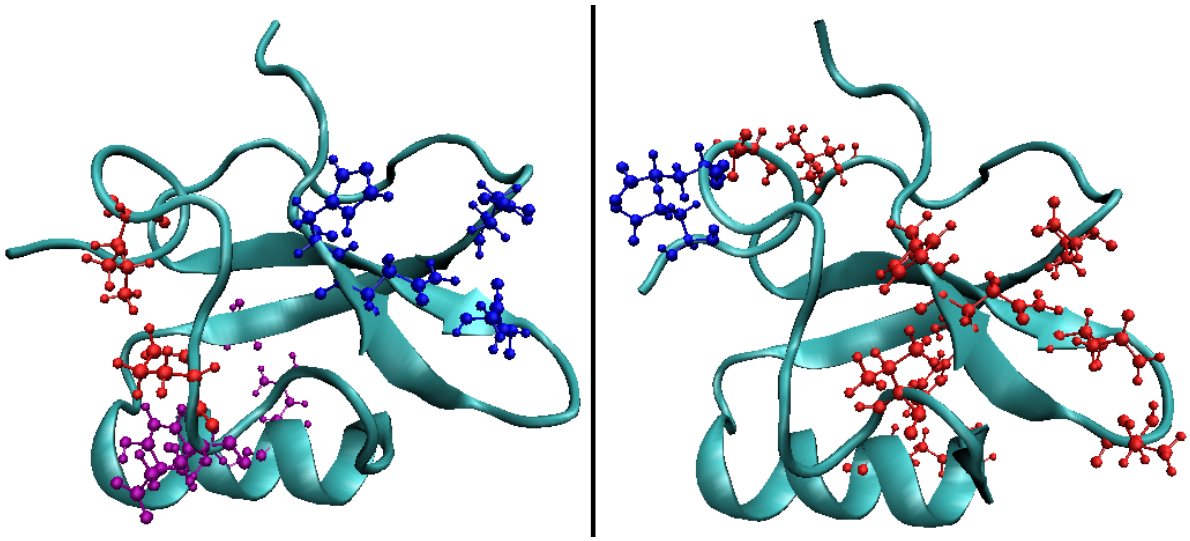


Figure 2.6: Highly significant contact differences emerging from basin splitting between basins 1.4/3.1 (left) and 1.4/3.2 (right).

Chapter 3

Counterion Atmosphere and Hydration Patterns near a Nucleosome Core Particle

Abstract

The chromatin folding problem is an exciting and rich field for modern research. On the most basic level, chromatin fiber consists of a collection of protein-nucleic acid complexes, known as nucleosomes, joined together by segments of linker DNA. Understanding how the cell successfully compacts meters of highly charged DNA into a micrometer size nucleus while still enabling rapid access to the genetic code for transcriptional processes is a challenging goal. In this work we shed light on the way mobile ions condense around the nucleosome core particle, as revealed by an extensive all-atom molecular dynamics simulation. On a hundred nanosecond timescale, the nucleosome exhibited only small conformational fluctuations. We found that nucleosomal DNA is better neutralized by the combination of histone charges and mobile ions compared with free DNA. We provide a detailed physical explanation of this effect using ideas from electrostatics in continuous media. We also discovered that sodium condensation around the histone core is dominated by the experimentally characterized acidic patch which is thought to play a significant role in chromatin compaction by binding with basic his-

tone tails. Finally, we found that the nucleosome is extensively permeated by over a thousand water molecules, which in turn, allows mobile ions to penetrate deeply into the complex. Overall, our work sheds light on the way ionic and hydration interactions within a nucleosome may affect inter-nucleosomal interactions in higher order chromatin fibers.

3.1 Introduction

DNA in eukaryotic cells folds into a highly compact and hierarchically organized structure called chromatin [101]. Genomic DNA contains on the order of 10^9 base pairs, corresponding to a contour length on the scale of a meter, which must be compacted by many orders of magnitude to fit inside of a eukaryotic nucleus which is several micrometers in diameter. DNA is a highly charged macromolecule, carrying two negative elementary charges per base pair, making charge neutralization essential to diminish strong self-repulsion and allow the chain's compaction [173]. Additionally, DNA is a semi-flexible polymer chain, with a persistence length of ~ 50 nm (at physiological conditions), thus, high level compaction would also incur elastic and entropic penalties [173]. These two problems are addressed to a significant extent through the formation of a DNA-protein complex, called a nucleosome (see Fig. 3.1), in which the DNA associates with a protein histone core, the latter carrying a large net positive charge [103]. Specifically, a nucleosome particle is composed of 146 DNA base pairs wrapped in ~ 1.7 turns around the core of the histone proteins. In addition to the histone core, neutralization of the remaining DNA charge is controlled by the surrounding aqueous salt environment. Aqueous salts play an important role in mediating various biological processes such as protein association [174, 175], RNA folding [176], nucleosome formation [100] and chromatin compaction. In this work we extensively study the environment of mobile counterions surrounding the nucleosome.

The nucleosome particle, together with the linker DNA segment connecting adjacent nucleosomes, represents a basic repeating unit of chromatin fiber [101]. The nucleosome complex may be viewed as the first, most basic level of the nuclear DNA compaction — subsequent compaction is achieved through the formation of secondary structure (30 nm fiber). The formation of the most basic chromatin structures are strongly dependent on the concentration of mobile ions in solution [104], whose role is to mitigate the electrostatic repulsion between linker DNA segments and also to mediate internucleosomal interactions. How mobile ions

interact with both the nucleosomal DNA and the protein histone core is an important issue which has been largely overlooked in prior computational studies of nucleosome particles and polynucleosomal arrays [177–179]. A precise description of site specific electrostatics of the nucleosome is likely needed for an accurate description of chromatin folding. For example, histone tails, which are flexible, highly charged, and unstructured portions of the main histone proteins, protrude from the nucleosome particle and mediate interactions between proximal nucleosomes [109]. Posttranslational modifications of these tails, which alter their electrostatics, are among the primary mechanisms regulating chromatin folding [109, 110]. It was recently suggested that a specific localized acidic patch on the histone core may play an important role in chromatin folding by providing an interaction site for histone tails [103, 111]. Though histone tails are not considered explicitly in the present study, we will show that sodium ions act as a probe and condense strongly around this acidic patch.

Both histone tail modifications and alterations to the surrounding ionic solution result in substantial changes to the nucleosome's electrostatic environment and crucially affect the formation of higher-order chromatin structures. Thus, to tackle the larger problem of chromatin folding, greater knowledge is needed at an atomic scale about the complex electrostatics of linker DNA, the nucleosome core particle, and histone tails. In prior works, we have analyzed short range electrostatics and the ionic atmosphere around linker DNA using all-atom Molecular Dynamics (MD) simulations [105–108]. In subsequent work, we developed coarse-grained models for linker DNA and electrolyte solutions, derived systematically by matching molecular correlation functions with all-atom MD simulation results [180, 181]. Analogously, it is expected that deeper understanding of the electrostatic environment of the nucleosome may be used to shed light on intermolecular forces driving chromatin folding.

In this work we endeavor to provide new insights into the nature of the nucleosome complex in its native state by performing a computationally extensive, 200 ns, explicit-solvent MD simulation of the yeast nucleosome (*Saccharomyces cerevisiae*) immersed into physiological

NaCl salt buffer. This simulation is an order of magnitude longer than a prior all-atom simulation of the nucleosome [182]. The longer simulation is needed to ensure that sufficient time has expired for mobile ion equilibration [183, 184]. We study the extent of the charge neutralization of the DNA surrounding the nucleosome. For added detail, we compute one- and two-dimensional radial distribution functions (RDFs), to identify spatially inhomogeneous association of mobile ions with DNA and histone proteins. Finally, we examine the high solvent accessibility of the nucleosome complex by identifying and visualizing water molecules residing inside the protein histone core and in between nucleosomal DNA strands.

The main results of the present study are as follows: It has been suggested that the complexation of the nucleosome core particle is driven in a part by the entropy gain from counterion release [100]. In this work we find that positive histone charges displace mobile counterions at a ratio that is larger than expected from simple counterion condensation arguments. We provide a detailed analysis of the causes of the excess positive charge and show that it is a result of multiple factors: First, the close wrapping of DNA leads to an additive effect of the electric fields produced by the DNA and causes an increase in sodium condensation relative to that normally observed in free DNA. Second, the difference in dielectric constants between the water and the histone core which the DNA encircles creates a mirroring effect that enhances sodium condensation. Third and finally, immobilizing some counter charges near the DNA leads to an overall enhancement of the charge neutralization. As a corollary to our findings, the histone core, in conjunction with counterions, is better able to neutralize the charge of the encircling DNA than what monovalent counterions alone accomplish on free, unbound segments of DNA, likely resulting in reduced inter-nucleosomal repulsion in compact chromatin fibers.

In addition to the expected high density of sodium condensation on the nucleosomal DNA, we also revealed highly localized sodium condensation on the largely positively charged histone core around sites consisting of predominately acidic residues. The most significant sodium condensation occurs at the acidic patch which is suggested to play a significant role as a poten-

tial internucleosomal interaction site for histone tail binding [103, 111]. We suggest that this finding adds new insight to the theory regarding histone tail binding to the acidic patch in that in addition to the assumed electrostatic favorability of association, there is an also an entropic gain due to the release of counterions from the site.

We also found over thousand water molecules extensively permeating the nucleosome particle interior, likely increasing the average interior dielectric constant from around 2-4 found in protein cores. Thus, as a physical object, nucleosome core particle resembles a sponge. This, in turn, allows mobile ions to easily penetrate the complex interior.

In summary, our extensive all-atom simulation provides new insight and understanding of the nucleosome core particle.

3.2 Methods

3.2.1 Simulation Details

An all-atom molecular dynamics simulation of a yeast nucleosome was performed using the CHARMM27 protein-nucleic acid force field [126] and the NAMD program suite [160]. A yeast nucleosome, (*Saccharomyces cerevisiae*), PDB entry 1ID3, was taken as the initial structure [185] (Figure 3.1). Aspartic and glutamic acid residues were assigned to be negatively charged, while glycine and arginine are assigned to be negatively charged. The 17 histidine residues in the histone core were assigned to their neutral state with a single proton at the $N^{\varepsilon 2}$ position, however, this residue may be protonated in the $N^{\delta 1}$ site or doubly protonated in a real nucleosome. The simulation was performed with over 54000 explicit TIP3P water molecules in a box with the dimensions $\approx 135 \times 109 \times 140$ Å. The simulation was performed under periodic boundary conditions with a minimum of 30 Å between nucleosomes in neighboring cells. The total system charge was neutralized by the addition of sodium counter ions, followed by the subsequent introduction of an additional 150mMol NaCl to approximate physiological con-

ditions. The system was initially held at constant volume, and the nucleosome was frozen in place while the water and ions were minimized for 100,000 steps. Subsequently, all constraints were removed from the system and it was minimized for an additional 400,000 steps. The system was gradually heated from 0 K via Langevin dynamics to 300 K in incremental steps of 5 K every 10 ps.

The production run proceeded with 2 fs time steps using the SHAKE algorithm and Ewald summation for long-range interactions. Weak harmonic positional restraints of 5×10^{-5} kcal / mol Å^2 were used to prevent large-scale translation or rotation of the nucleosome in an asymmetric box, however, small- and medium-scale nucleosomal motions were not affected. Short-range non-bonded interactions were calculated at each step, long-range interactions were calculated on even steps only, and the pair list was updated every 10 steps. System coordinates were saved every 1000 steps (2 ps) for later analysis. The production run was conducted under the constant pressure, moderated by Langevin piston (set to 1 atm), with a total simulation time of 200 ns ($\sim$ 205,000 CPU hours using 2.3 GHz Intel EM64T processors).

3.2.2 Verification of Convergence

It was shown previously that 50 ns was needed to equilibrate ions near DNA chains [183, 184]. To verify the validity of this assumption in our simulation, we divided the trajectory into four 50 ns sections and computed radial distribution functions for each section individually. The only significant change was observed between the 1st 50 ns and the rest of the trajectory. During the first 50 ns, the first two peaks of the nucleosome-sodium RDF showed a subtle shift of some sodium from the second peak to the first. It should be noted that we did not observe any significant changes in nucleosome-ion RDFs after the first 50 ns. The total number of ions within 1 nm of DNA was stable throughout the entire 200 ns of the simulation ($191 \pm .9$). As an additional check, we tracked ions in the histone core and found that the vast majority of ion condensation events (approach of within 7Å of nucleosome) were vanishingly short with only

1.7% of condensations lasting for more than 1 ns and only .1% of condensations lasting for more than 10 ns. We observed a single long lived condensation event which lasts for more than 100 ns which is elaborated upon below. These collective results suggest that though there may be a few highly specific sites which have not reached equilibration, the number of these sites is exceedingly small and has little effect on the primary conclusions of this work.

3.2.3 Analysis of Counterions with all-atom MD

A standard method for analyzing counterion distribution in solution is to compute a radial distribution function (RDF). The first step in the creation of a nucleosome-counterion RDF involves computing the distance between each ion in the system and the nearest atom in the nucleosome at each time step. These data are then histogrammed and normalized by a volume Jacobian to obtain the RDF. The Jacobian was obtained numerically and was defined for each distance as the total volume of a discrete shell defined in a 3-D cubic lattice at that distance from the nucleosome. The bin size for the RDF histogram and the lattice apportionment size for the Jacobian calculation were both 0.5 Å. Since the nucleosome is a highly rigid molecule the Jacobian was only recomputed every 1000 time steps. All in-house analysis programs were written with the assistance of the Biochemical Algorithms Library (BALL) [186].

We have also visualized ion density in 3-D through use of a technique that can be paralleled to long exposure photography. First, overall translation and rotation of the nucleosome are removed from the trajectory through least-squares fitting to the crystal structure. Then, simulation box is divided into $1.5 \times 1.5 \times 1.5$ Å cells and the location of each ion is recorded for the entire length of the simulation. Each cell is then assigned an average ion concentration.

3.2.4 Analysis of Counterions with the Poisson-Boltzmann Equation

In addition to our all atom studies, we have performed mean-field electrostatics calculations using the Poisson-Boltzmann formalism. For this work we used the Adaptive Poisson-Boltzmann

Solver (APBS) [187] program suite. A $257 \times 193 \times 257$ rectangular grid with a step size of 0.78 Å was used for the P-B calculation. Sodium and chloride bulk concentrations were both set to 150 mMol and the solvent probe radius was set to 1.4 Å. Solute and solvent dielectrics were set to 3.0 and 78 respectively. APBS was used to solve the non-linear P-B equation with the boundary condition potential specified by a Debye-Hückel model for multiple, non-interacting spheres. Identical parameters were used in calculations of one and two DNA chain systems, however, in these cases the grid step size was roughly halved (varying slightly between systems). Computation of RDFs arising from P-B results was carried out using similar methods to those previously used for all atom results.

3.3 Results and Discussion

3.3.1 Nucleosome is Highly Rigid on a 200 ns timescale

The first goal of our study was to characterize the conformational dynamics of the nucleosome on a 200 ns timescale. Visually, the structures appear remarkably similar throughout the simulation, except for a minor difference in the orientation of the loose entry and exit ends of the DNA, which are not tightly bound to histone core and are exposed to the solvent. To quantitatively examine this issue, we used dihedral angle principal component analysis (dPCA) to characterize the nucleosome's dynamics, similar to earlier work [188]. This technique revealed the presence of several dynamical basins in the space described by the first two principal components (data not shown). However, representative structures from the most dissimilar basins were found to be within $\sim$ 3 Å RMSD from each other. The robustness of the nucleosomal structure is subsequently useful later in this article when we generate a three-dimensional time averaged ion and water density plots around the nucleosome. In summary, on a 200 ns timescale, the nucleosome appears to undergo only small-scale structural fluctuations.

3.3.2 Nucleosomal DNA is More Neutralized than Free DNA

The primary goal of our study was to characterize the counter-ionic environment surrounding the nucleosome. For this discussion, we introduce the quantity γ, which is the percentage of the residual DNA charge after accounting for counterions and histone charges within 1 nm of the DNA surface. As a first step, we computed γ from our all atom simulation. The results of this calculation are shown in Table 3.1. Previous work has suggested that mobile ions are able to neutralize $\sim 75\%$ ($\gamma = 25\%$) of the charge of a strand of free DNA at a distance of .9 – 1 nm [105, 136, 189, 190]. The DNA surrounding the nucleosome has a charge of -292, therefore we would expect ~ 219 of these charges to be neutralized by counterions. The histone core provides a net charge of +76 within 1 nm of the DNA, therefore, a release of 76 mobile counterions is anticipated if we assume a 1 to 1 replacement. However, we observed a net +174 charge from counterions within 1 nm, indicating that only 45 counterions were released, instead of expected 76. This means that the histone and counterions ions collectively reduce residual DNA charge from $\gamma = 25\%$ in free DNA to $\gamma = 14\%$. Since this phenomenon results from multiple effects, we attempt to provide a clear explanation through a hypothetical, step by step, nucleosome assembly process (Figure 3.2). For each step, we performed an extensive series of non-linear Poisson-Boltzmann calculations to demonstrate how that particular step alters counterionic atmosphere around DNA.

Our assembly process starts (**State I**) with a a strand of free DNA in a mean field solvent and salt bath (150 mMol NaCl). In this state, we computed $\gamma = 34\%$ for the residual charge of DNA at 1 nm. It should be noted that this result is a higher baseline than predicted by Manning [190] and our previous all-atom simulations [105].

Next, as an approximation of the real system we measured the effect of bringing together two parallel strands of DNA from an initially large distance, to mimic the effect of **State II**. As shown in Fig. 3.3a, counterion condensation monotonically increases as the DNA center to center chain distance decreases as an effect of the combination of the electric fields created by

each strand. Bringing two DNA strands from an infinite separation to within a few Ångstroms of one another has a similar effect to increasing the charge density on a single strand, which is expected to result in enhanced condensation in order to lower the line charge density down to the Manning threshold. At the inter-DNA separation found in the nucleosome, this effect brings the DNA residual charge from $\gamma = 34\%$ to $\gamma = 23\%$. We observed similar, but less pronounced results from prior all-atom simulations [107], where we measured a $\gamma = 26\%$ to $\gamma = 18\%$ change in DNA residual charge. If instead of two parallel strands of DNA, we perform a P-B calculation with the coiled DNA from the nucleosome PDB with the histone cores removed we observe a small change from $\gamma = 23\%$ to $\gamma = 20\%$. This is likely a result of the curved geometry of the DNA chain in the latter as opposed to the two linear strands used in the former case.

The next step of inserting a histone core into spiral-shaped DNA chain introduces multiple effects due to: 1) excluded volume, 2) low dielectric and 3) immobilized histone charges. These effects change the extent of counterionic condensation in profoundly different ways. Thus, to individually understand each of these effects, we subdivided the histone insertion into several sub-steps. In **State III**, to reveal the effect of the excluded volume interactions, we introduce a completely neutral histone core, impermeable to the ionic bath, and with the same dielectric constant as the outside solvent. The effect of this excluded volume is a small change in residual DNA charge from $\gamma = 20\%$ to $\gamma = 22\%$.

In **State IV**, we continue to keep the histone core devoid of any charge, however, we reduce the dielectric constant of the material from that of the solvent $\varepsilon \sim 78.54$ to $\varepsilon \sim 3.0$ (however, this might significantly underestimate the dielectric constant of nucleosome complex if treated as a continuum material, as discussed below). It has been argued that the assignment of a dielectric constant for a protein in a P-B system is highly non trivial, and a higher value may be more appropriate [191]. Thus, the results from this section may be scaled but the trends should be the same. The effect of the reduction of the core dielectric (Fig. 3.3 b) can be explained by

the dielectric mirroring effect in which the higher permittivity of the solvent keeps the force lines resulting from DNA electric field in the solvent, effectively creating an image charge of the same sign within the lower dielectric material [192]. As the dielectric constant of the histone core is decreased, counterion condensation increases, since the image charge effect grows stronger. This leads to a change in residual DNA charge from from $\gamma = 22\%$ to $\gamma = 18\%$.

Finally, in **State V**, we gradually introduce the histone charges (Fig. 3.3 c) by uniformly changing the normal atom charges with a constant scaling factor and observe a final small change in residual DNA charge from $\gamma = 18\%$ to $\gamma = 17\%$. Using the analytical solution of the Poisson-Boltzmann equation for a solution of uniformly charged cylinders [193], we found qualitatively similar results. Although in the case of the nucleosome this effect is small, it can become quite significant, when around 50% of mobile charges become stationary. To explain this effect we consider that the introduction of immobile positive charges from the histone core effectively reduces the DNA line charge density. Despite the fact that such a system results in diminished counterion condensation, we still find a net increase in DNA neutralization when we retroactively subtract the immobilized charges from the DNA line charge and add them to the total number of counterions. Since the counterion condensation profile results from competition between electrostatics and mobile ion translational entropy, making some mobile ions stationary reduces the entropic penalty, allowing larger overall charge neutralization.

Poisson-Boltzmann results indicate that the most significant contribution to DNA charge neutralization is simply the result of the enhancement of the electric field caused by the tight wrapping of the DNA chain around the histone core. However since Poisson-Boltzmann results are not entirely numerically consistent with all-atom observations it is possible that the relative importance may shift between these effects. Further numerical experiments based on all-atom simulations could be used to examine the magnitude of each effect.

To verify that the electrostatic interactions between DNA and positively charged histone residues were not significantly diminished because of the latter being deeply buried, we defined

an effective radial center of the charge distribution within 1 nm of the DNA for both the charged histone residues and the ions in solution using the same method as would be used to compute a center of mass. We find that both the histone core and the mobile ions possess nearly the same center of charge (Table 3.1). This means that within 1 nm, the average distance of a charged residue to the DNA chain is the same as the average distance of an ion to the DNA chain. There are 13 histidine residues within 1 nm of DNA which could become deprotonated or could change protonation sites in a real system. Deprotonation of these sites could cause some additional counterion release, but we do not expect this change to significantly change our conclusions. Collectively, these results likely carry significant consequences for higher order chromatin compaction, because they demonstrate that the combination of a histone core and monovalent counterions is more effective at neutralizing the charge of a segment of DNA up to a distance of 1 nm than if the same segment were in the presence of only the counterions. It has been noted that some histone tails preferentially bind to linker DNA over nucleosomal DNA [194, 195]. We postulate that this effect may in part be caused by the increased charge neutralization of nucleosomal DNA relative to linker DNA.

3.3.3 High-Resolution Description of Counterionic Atmosphere

As a next step, we computed nucleosome-sodium RDFs for both all-atom and mean-field Poisson-Boltzmann results (Figure 3.4). Both results showed a significant increase in sodium concentration around the nucleosome relative to that of the bulk, which is unsurprising given the large net negative charge of the DNA histone complex. Significant qualitative differences, however, exist between the all-atom and the mean-field results. Three distinct ionic shells surrounding the nucleosome were observed in atomistic simulations, while the P-B calculations lack this fine detail and show one single peak. This is not unanticipated because P-B calculations neglect the discrete nature of water and ions. A similar result was found in our previous work where ion condensation around a segment of linker DNA was analyzed using atomistic

and mean field calculations [105]. As discussed in our previous work, the single P-B peak near the nucleosome is equivalent to the second peak from the all-atom result while the first peak is absent. This first peak is representative of direct association of ions with the nucleosomal atoms. As with linker DNA, the strong close condensation of ions around the nucleosome significantly reduces the short range (< 1 nm) nucleosomal charge and may promote folding of a polynucleosomal array into chromatin fiber.

While a nucleosome-sodium RDF can provide useful information about the surrounding ionic environment, its resolution is very low. The nucleosome is far from a homogeneous object and is large relative to the length scale of interesting ionic condensation. As previously stated, the histone core has a large net positive charge, while the surrounding DNA is highly negatively charged. The dissimilarity between the two components of this system suggests that there is likely a significant difference between the propensity for a sodium ion to associate with DNA compared with histones. To address this issue, we have computed two-dimensional RDFs. The first dimension was chosen to be the distance between a sodium ion and the closest atom belonging to the DNA chain, while the second dimension corresponds to the distance between that same sodium atom and the closest atom belonging to the protein core. Two dimensional RDFs were created for the all-atom systems (P-B results are qualitatively similar with the same deviations observed in the 1-D case) with interesting results (Figure 3.5). Unsurprisingly, sodium ions preferentially associate with DNA however, we also observed a significant condensation of sodium around the histone protein at locations distant from the DNA.

To understand the source of the sodium concentration around the histone core, we scanned the all-atom trajectory, selectively choosing only the ions within the two-dimensional RDF region identified by our plots. These sodium atoms were then examined against all residues of the nucleosome to identify the closest. It was found that this peak did not correspond to a large condensation at a single location, but contributions arose from several locations around specific residue types. Dense sodium condensation was observed primarily in regions rich in

solvent accessible acidic residues. The plurality of the sodium condensation, 46%, was found closest to glutamic acid residues (negatively charged), followed by aspartic acid (negatively charged), 19%, and serine (neutral; polar), 19%. To further explore this issue, we also extended the search to include neighboring residues on the nucleosome. The results showed that ions condense most significantly when there are two or more acidic residues in close proximity.

Curiously, we identified a single condensation site around which the surrounding residues within $\sim$7Å (the Bjerrum length in water) possessed no excess of negative charge. We found that this condensation could be attributed to a single sodium atom in the interior of the histone core which occupied the site for over 100 ns and had not left by the end of the simulation. Visual inspection of the sight, revealed that the sodium was forming a chelate complex with carboxylate group on the aspartic acid, a carbonyl oxygen on the serine backbone and the hydroxyl group on the serine side chain (Figure 3.6). Thus we hypothesize that the highly localized favorable interaction with these chelating oxygens and the associated entropy gain by sodium's ionic shell waters released into the bulk provided the remarkable longevity of the occupancy. While it is possible that there exist other neutral sites capable of chelating a sodium ion, we do not believe that this effect would be of much significance in terms of internucleosomal interactions because of steric considerations.

We compared our detailed sodium condensation data with a crystallographically identified acidic patch [103] which may play a significant role in internucleosomal interactions, and discovered that this patch was responsible for much of the observed sodium condensation. This observation may provide an additional suggestion regarding the method of histone tail association. In addition to the expected electrostatic favorability of bringing the basic residues of the histone tail in contact with the acidic patch, we now also expect there to be an entropic gain associated with counterion release from the site upon binding. Additionally, this is exciting because it may be possible to use counterion condensation data to search for other potential internucleosomal interaction sites *in silico*. In addition to radial distribution functions, we also

produced a 3-D mapping of sodium density around the nucleosome (Figure 3.7). This plot clearly identifies the high level of sodium condensation around both the DNA and around the crystallographically identified acidic patch.

3.3.4 Nucleosome as a Sponge Filled with Water

Examination of the 3-D sodium density revealed ion penetration into the interior of the histone core, such as the previously described case of the chelated sodium. We verified that these ions were not present at the start of the simulation, naturally leading to the conclusion that they had arrived at some later time. This unanticipated result led us to examine the permeability of the histone core to the solvent. Crystallographic experiments have shown the presence of solvent within the core and the existence of significant water mediated interactions [196]. In a similar fashion to our sodium density calculations, we computed a 3-D solvent density of a space that can be described as an imaginary capsule capable of encompassing the interior of the histone core (Figure 3.8). Our analysis revealed an enormous solvent accessibility of the interior of the histone core. Although individual histone monomers effectively exclude water from their positions within the core, the space between histones is significantly solvated with a very large solvent channel in the center perpendicular to the DNA wrapping. This yields two important observations: First, the solvent allows ions to penetrate the core and may provide a local stabilization in regions of numerous like charges. Second, the solvent permeability mitigates the effect of the large net charge of the histone core by effectively providing a high dielectric bath. This emphasizes the need for a more detailed description of the core in the context of chromatin folding, which is almost always overlooked in modern coarse grained models of chromatin.

We also investigated the solvent mobility in the core to understand whether or not the water was more significantly structured in the core than in the surrounding solvent. We first computed a radial distribution function between the nucleosome and the oxygen atom of surrounding wa-

ter molecules and determined the depth of the first solvation layer. From there, we computed autocorrelation functions of the water dipole moments within the first solvation layer and compared them with bulk water outside of this layer in the cases of both the interior and exterior of the nucleosome. Unsurprisingly, external bulk water was highly mobile with the average autocorrelation decaying to less than 10% after 10 ps. This result stands in stark contrast to water within the first hydration layer on the exterior of the nucleosome. These waters were highly structured and retained an average autocorrelation of $\sim 65\%$ after 10 ps, only decaying to $\sim 40\%$ after 100 ps (the size of our autocorrelation windows). Though autocorrelation of the dipole bond vectors can not produce meaningful rotational diffusion constants [197], they still provide a useful gauge of the relative mobility of the selected water. We found no significant changes in these autocorrelation functions between interior surface waters and exterior surface water. Additionally, there was also no significant change between the small amount of interior bulk water and exterior bulk water. Our results suggest that the majority of the water ($\sim$950 molecules) in the interior of the nucleosome are significantly constrained in dynamics, to the same extent as exterior water that resides within the first hydration layer. However, there is a significant amount of water ($\sim$200 molecules) in the core that behaves in a bulk-like fashion.

3.4 Conclusions

In this work, we described the ionic atmosphere around the nucleosome complex with atomic detail from an extensive, 200 ns all-atom MD simulation of the yeast nucleosome. We analyzed conformational rearrangements over the course of the MD simulation using dPCA and concluded that although the nucleosome is dynamic, large scale conformational changes are minimal over the accessible timescale.

Because subtle electrostatic effects play a crucial role in the process of condensation and decondensation of chromatin fiber, the need to understand electrostatics on the level of a single

nucleosome particle arises naturally. We found in this work that the combination of the histone core and mobile counterions is more effective at neutralizing the DNA charge on a short, 1 nm, length scale than mobile counterions alone. We demonstrated that this effect is a combination of 4 factors: First, winding the DNA around the histone core results in an additive combination of the electric field between closely wrapped bases and an increase in sodium condensation and thus greater charge neutralization. Second, the excluded volume consumed by the histone core causes a slight reduction in the ability of mobile counterions to neutralize the DNA chain. Third, the differences between the solvent and histone core dielectric constants creates image charges of the DNA charge which increases sodium condensation and charge neutralization. Finally, the stationary positive histone charges also cause an increase in charge neutralization. Our results emphasize the role of proximal DNA chain interactions within a single nucleosome: wrapping DNA around the histone core results not only in significant elastic penalty, but also unfavorable inter-DNA electrostatic repulsion, where the latter effect is mitigated by additional counterion condensation. Thus, these unfavorable interactions must be overcome by both the free energy gain due to released counterions upon histone core insertion, and and a combination of hydrophobic and hydrogen bond interactions between the histone core and the DNA. Larger neutralization of DNA in a nucleosome is expected to diminish inter-nucleosomal repulsion in higher order chromatin arrays and may play a role in determining preferential binding of histone tails to linker DNA.

To obtain a more detailed description of the ionic atmosphere around the nucleosome, we computed one and two-dimensional distribution functions of the surrounding ions. The results indicate that sodium ions condense primarily on the DNA molecule, as expected from simple electrostatic arguments, exhibiting characteristic patterns of pronounced peaks and minima in the 1D ion-DNA RDF, due to hydration shells. Far less trivial behavior was observed for the ions interacting with the protein part of the nucleosome, far from the DNA. Specifically, sodium seemed to heterogeneously condense around a localized region of acidic residues, which are

required for nucleosome array folding [111]. Hence, we suggest that in addition to the electrostatic favorability of binding between the acidic patch and the basic residues of the histone tail, there may also be a favorable entropic effect created by the release of counterions from the patch site upon binding.

In this work we found that the interior of the nucleosome core particle is highly solvent accessible and contains both surface and bulk-like water. This conclusion is consistent with the recently resolved crystal structure of the nucleosome core particle demonstrating the presence of many water molecules mediating histone-DNA interactions and the interaction among various core histone proteins [196]. However, this result runs contrary to many simplified models that treat the nucleosome as a dense, water impermeable, homogeneous, solid-like object with low a dielectric constant. The results of the present study indicate that in simplified, coarse-grained representation of the nucleosome, the effect of solvent will have to be taken into account by setting a significantly higher dielectric constant, than is customarily used ($\varepsilon \sim 2-4$) [198]. Finally, solvent accessibility appears to be responsible for the penetration of ions deep into the core of nucleosome. Particularly, we observed a remarkably long lived sodium condensation in the protein histone core, where the ions may form a chelate complex with the residues of histone protein.

In summary, our work provides a detailed atomistic picture of the nucleosome complex in its native state. New insights into the nucleosome core particle ionic atmosphere, solvent accessibility and dynamics point to important details that may help to better understand and model structure and dynamics of chromatin fiber.

3.5 Acknowledgments

We thank Dr. Max Berkowitz and Dr. Michael Rubinstein for insightful comments. This work was supported by the Beckman Young Investigator Award.

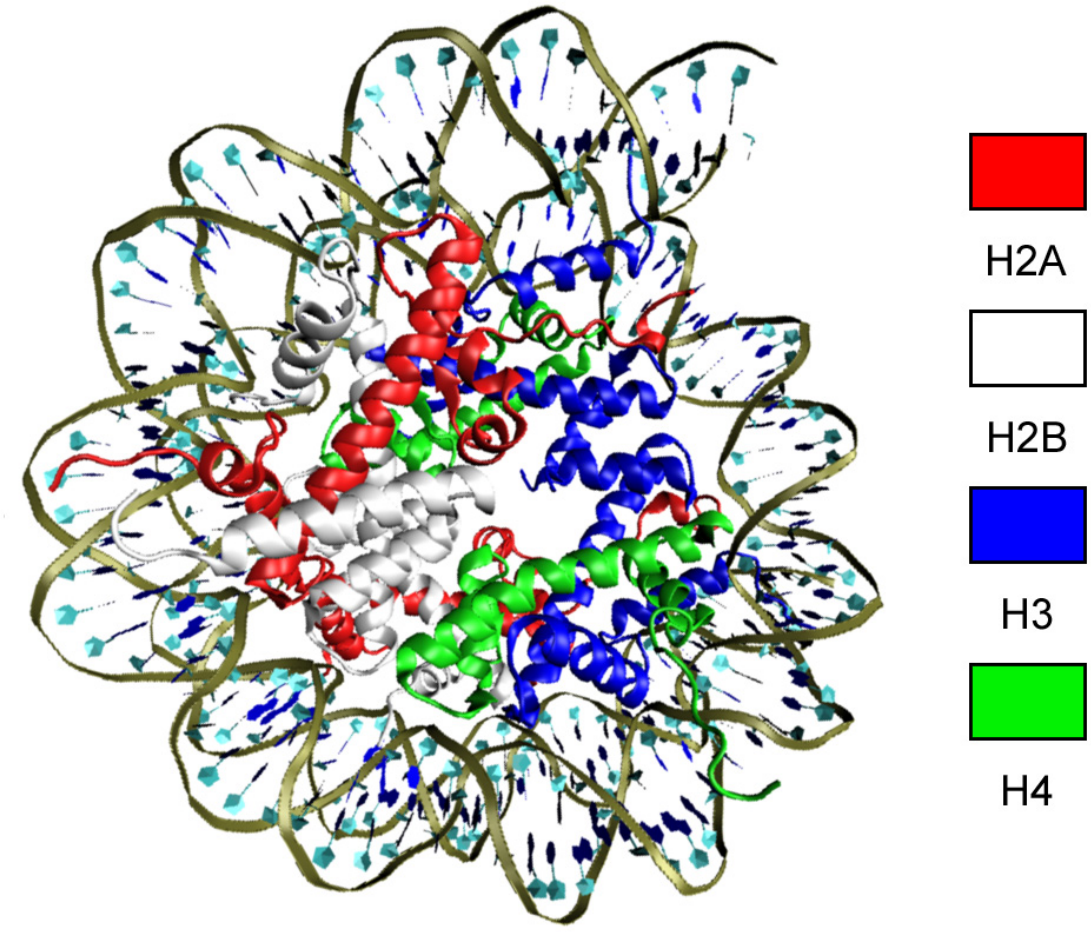


Figure 3.1: The yeast nucleosome used in this study (PDB entry 1id3) is shown here with its histone proteins color coded.

	Charge	% of DNA charge
Net charge of DNA	-292	100%
Net charge of counterions	174	60%
Net charge of histones	76	26%
Net charge of system within 10 Å	-42	$\gamma = 14\%$

Radial center of mobile charges	4.40 Å
Radial center of protein charges	4.34 Å

Table 3.1: Summary of results from our all-atom nucleosome simulation describing the charge distribution within 1 nm of the DNA surface

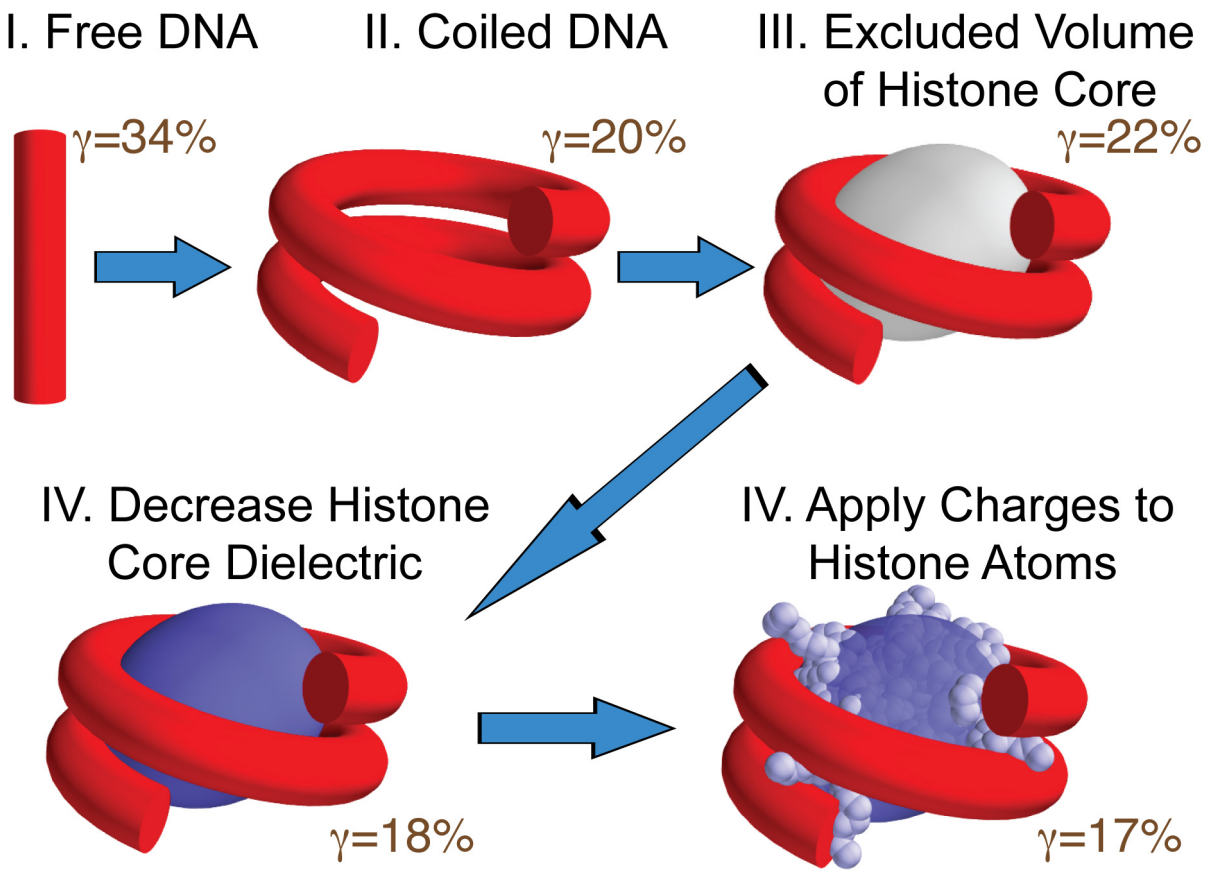
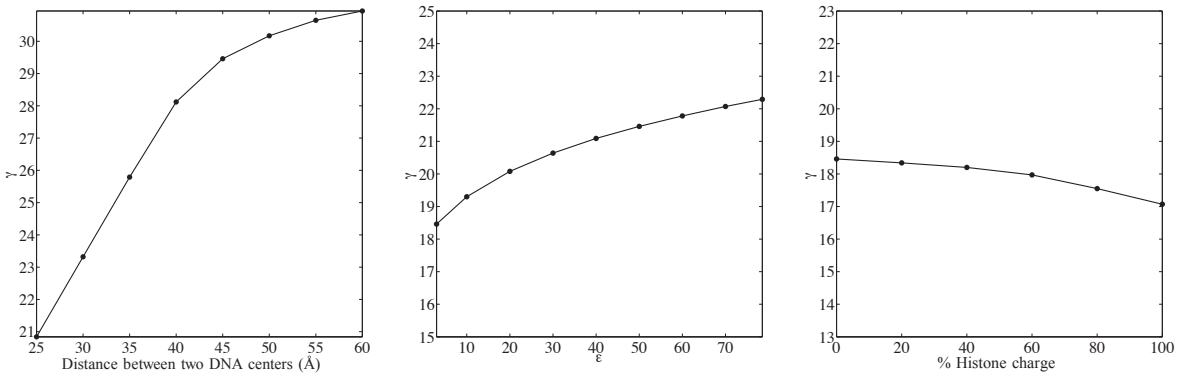


Figure 3.2: This figure shows a conceptual step by step nucleosomal assembly pathway for understanding the effects of nucleosome formation on counterion condensation.



(a) State II (b) State IV (c) State V

Figure 3.3: These figure shows the % of DNA charge neutralization a) of two strands of parallel DNA at multiple distances between the centers of the DNA strands, b) of the nucleosomal DNA wrapped around a histone core with different dielectric constants, c) of the nucleosomal DNA with the charges of the histone core uniformly scaled.

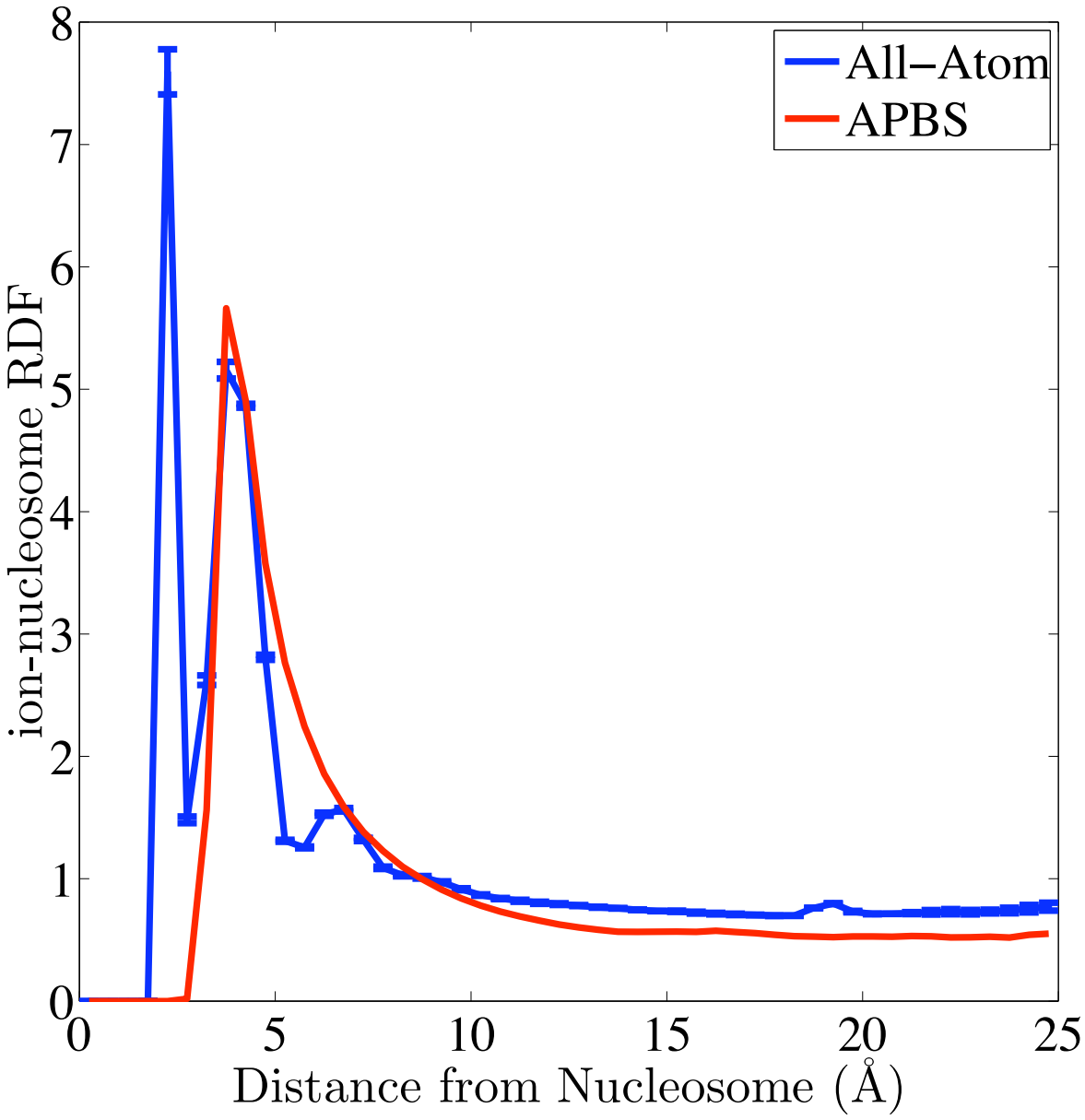


Figure 3.4: A comparison of the 1-D sodium-nucleosome radial distribution functions obtained from APBS and all-atom simulations reveals important qualitative difference. Specifically, the first peak of the all atom result, indicative of sodium condensation is absent from the mean field result. The all-atom plot is averaged from the final 3 50 ns segments with error bars shown for each point.

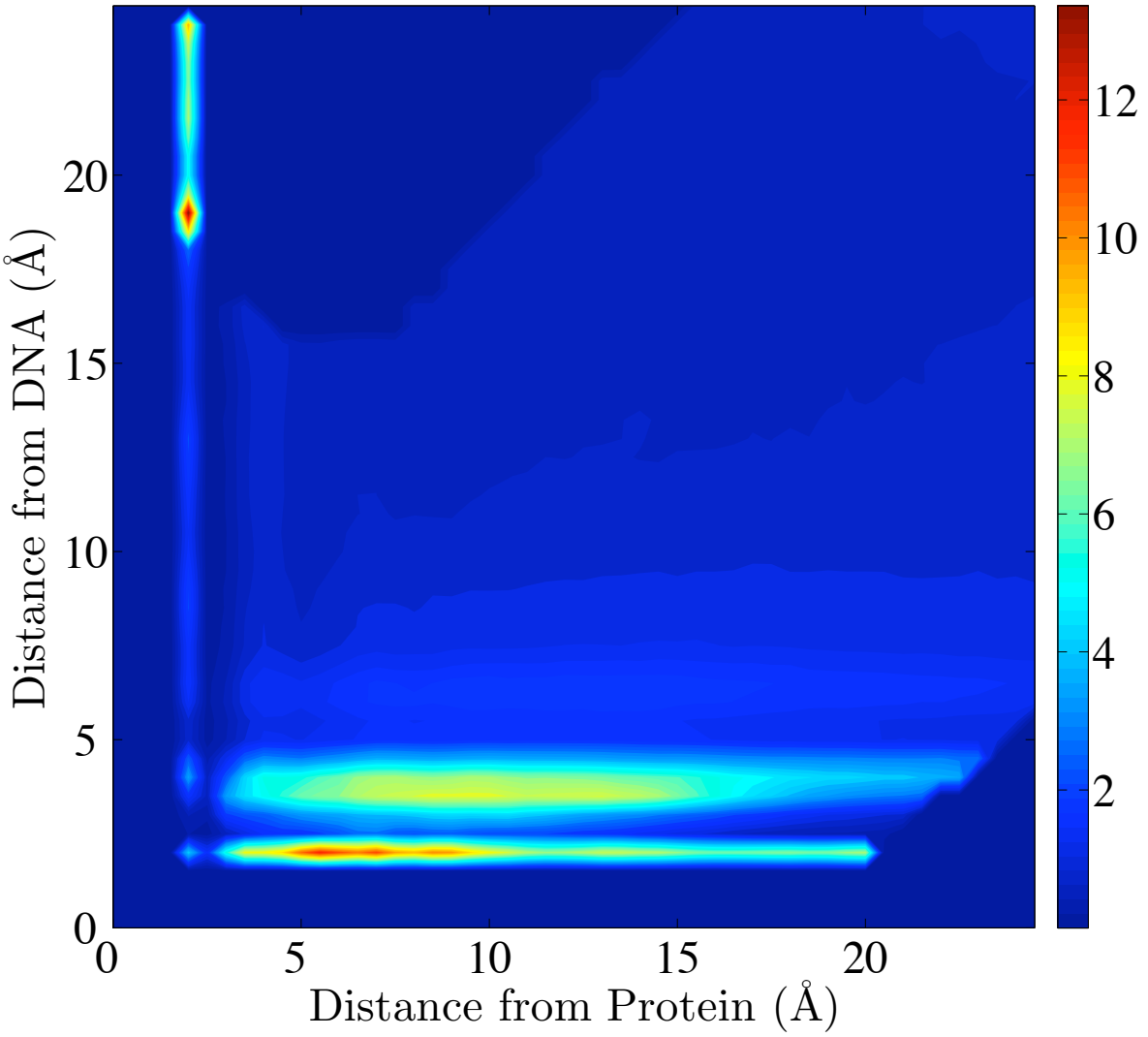


Figure 3.5: The 2-D sodium-nucleosome radial distribution function obtained from our all-atom simulation shows the layering of the sodium cloud around the DNA as shown in our 1-D result. It also reveals strong, highly localized sodium condensation on the histone core.

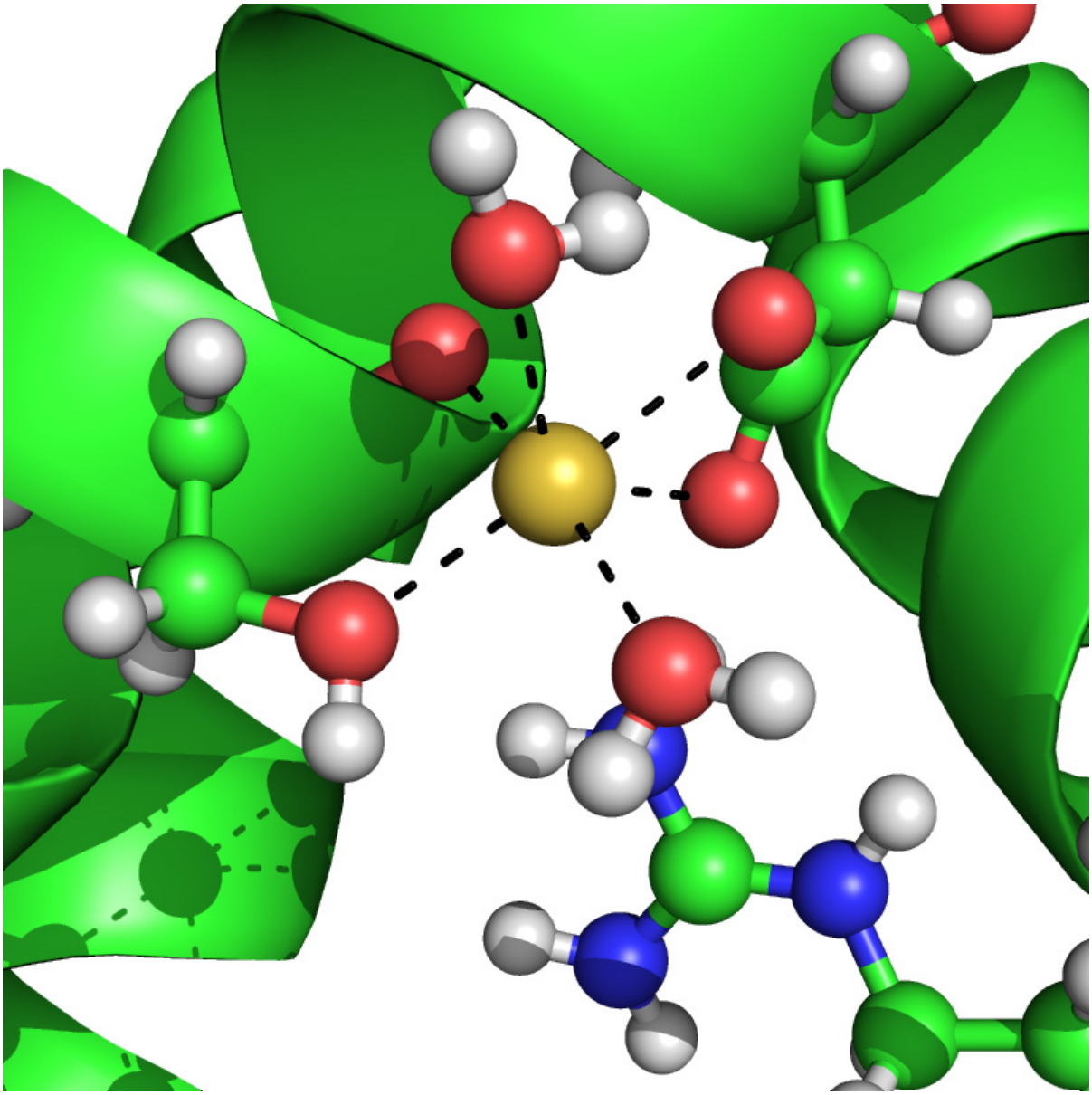


Figure 3.6: The chelate complex formed between a sodium ion (yellow) and oxygens (red) from the carboxylate group of a nearby aspartic acid residue in addition to both the backbone carbonyl and side chain hydroxyl group on a nearby serine residue (the atoms of the amino acid triplet and water in the first solvation shell around the sodium is shown as spheres).

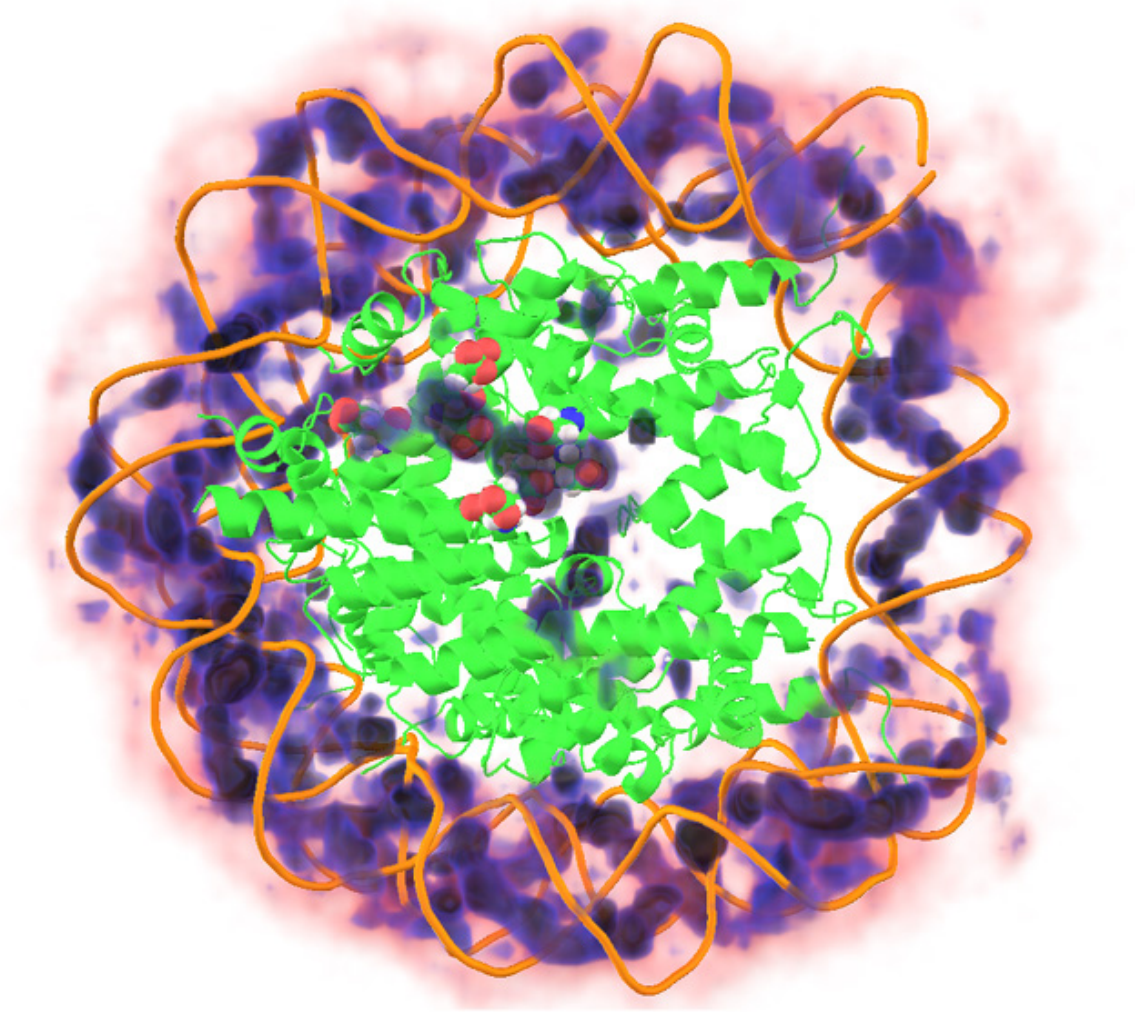


Figure 3.7: A 3-dimensional distribution of sodium around the nucleosome. The crystallographically identified acidic patch has been highlighted as spheres on the surface of the histone core and a high level of sodium condensation is observed around these residues.

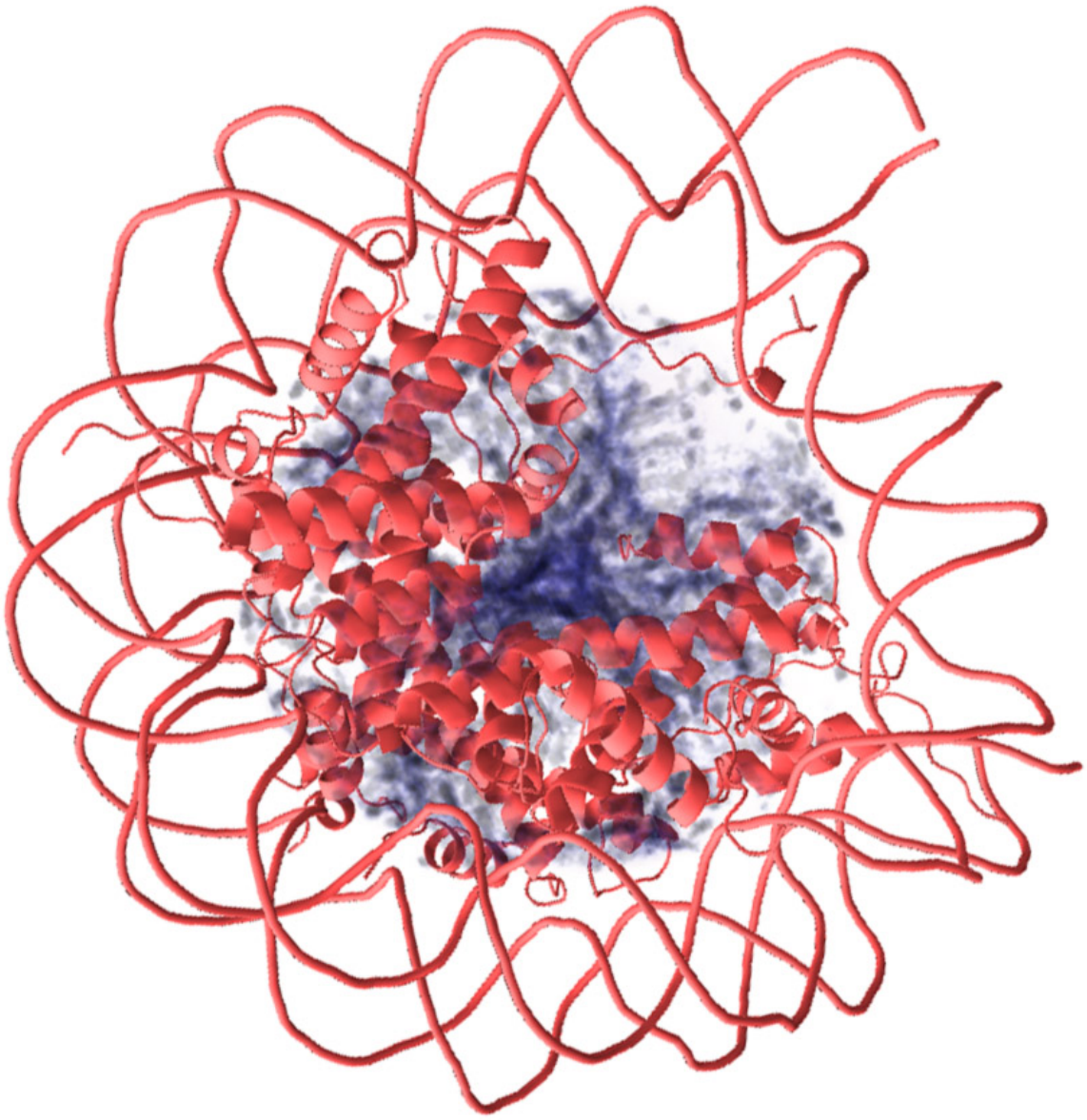


Figure 3.8: The core of the nucleosome is highly solvated as shown in this figure. One H2 histone has been removed for clarity.

Chapter 4

Dynamics of Light Harvesting Peptides

Abstract

Light harvesting antennae may play an essential role in large-scale molecular assemblies designed for the production of solar fuels. One of the most important features of these antennae is the ability to perform efficient energy transfer along its length. A regular pattern and controlled positioning of chromophores makes their behavior more predictable. Marcey Waters' group is developing such scaffolds using small coiled-coil peptides which may be useful because of their predictable structure. Chromophores may be tethered to a range of sites on these peptides, offering some degree of positional control. Despite this control, the tethers are long, highly flexible chains which provide the chromophores with a potentially wide range of conformational freedom. This conformational freedom may play an important role in regulating energy transfer dynamics. To that end, we have performed extensive all-atom molecular dynamics simulations of these peptide systems with the tethers in three distinct positions. Over the course of our simulations, we find that the tethers are highly dynamic and explore multiple conformational states. Structurally, we predict that the chromophores may have an impact on the stability of the coiled-coil structure as they seem to induce fraying in the termini. Our simulations are in agreement with experimental energy relaxation experiments. When the tethers

are placed at their greatest separation, we predict no energy transfer to occur, in agreement with experiment. We also predict that energy transfer could occur in the other two systems. In one case, this has already been proven by experiment; the other is yet to be tested.

4.1 Introduction

Photosynthesis is an example of a biological process which captures solar energy and stores it in a useful chemical form. Currently, chemists are trying to mimic these capabilities efficiently through the production of artificial photosynthetic systems. The difficulties in accomplishing this arise because such a system requires "multiple chemical functions to exist in a stable architecture [199]." Thus, scientists are trying to use the principles of physics and chemistry to bypass the billions of years of evolution in natural photosynthetic systems. Pathways for artificial photosynthesis involve the splitting of H_2O into hydrogen and oxygen and the reduction of CO_2 to methanol or some form of hydrocarbon. In natural photosynthesis, CO_2 reduction and H_2O oxidation occur in photosystems I and II respectively. These photosystems are surrounded by what are known as thylakoid membranes which support light harvesting chlorophyll. Light harvesting events in the chlorophyll lead to excited state energy transfer to the photosystems, driving the chemical reactions. Analogously, one approach, which is relevant for the research performed in this study, is based on spatially integrated molecular assemblies. Put simply, the system is divided into modules each with its own function, that, when combined, form a complete artificial photosynthetic assembly. Hypothetical minimal requirements for such a system were described in Alstrum-Acevedo *et. al.* [199]: First, light energy must be absorbed by a single chromophore or an antenna array of such chromophores. Next, electron transfer quenching of the excited state must occur followed by charge separation. These charges must then migrate to their respective oxidative and reductive catalysts, upon the arrival of sufficient charges to perform the needed half-reactions, they can react with CO_2 or H_2O and return the catalysts to their original states. In this work, we focus on the light harvesting module of this assembly.

The research group of Marcey Waters, working collaboratively within the UNC Energy Frontier Research Center, has developed a new form of light harvesting antenna using Ruthenium (II) based chromophores tethered to a coiled-coil peptide scaffold. Several of these pep-

tide systems have been produced with the chromophores tethered at different positions and energy transfer has been measured by introducing Osmium (II) chromophores which quench excited state Ru* complexes by acting act as low-lying energy traps. To understand these systems more completely, it is necessary to first consider their constituent components.

The activity centers of light harvesting antennae are the Ru(II) based chromophores. Ruthenium complexes have been studied for their photovoltaic properties, namely the ability to undergo electron transfer since the 1970s [200–202]. Photoexcitation of the chromophores used in this study, which are based on $[Ru(bpy)_3]^{2+}$ (Figure 4.1), undergo singlet metal-to-ligand charge transfer (^{1}MLCT), and a subsequent intersystem crossing to a (^{3}MLCT) state within tenths of a picosecond [203]. This triplet state has a relatively long lifetime of $\sim$1 μs because relaxation to the ground state is spin forbidden, providing sufficient time for energy transfer to an adjacent chromophore to occur [204]. Monte Carlo simulations of Ru(II) chromophore derivatized polystyrene suggest that the Ru*$\rightarrow$Ru energy transfer can occur on a 1–4 ns timescale with a single hop efficiency of 99.6 to 99.9% [204]. This implies that with a well designed polymer, it should be possible to conduct excited state energy over long distances with good efficiency. In order to quantify energy transfer dynamics in experiments, energy traps may be introduced in known quantities to terminate the migration process. For these systems, $[Os(bpy)_3]^{2+}$ based chromophores are introduced which possess this trapping property because Os excited states lie $\sim$0.34 eV below Ru excited states. Despite the rather large fraction of singlet character introduced to the excited state through spin-orbit coupling, it is believed that energy transfer in these systems occurs primarily through the Dexter mechanism since Förster energy transfer requires transitions in both donor and acceptor groups to be spin allowed. The Dexter mechanism requires molecular orbital overlap for electron exchange causing the energy transfer rate to decay exponentially with distance. This implies that for energy transfer to occur, the chromophores must be very close, with minimal external interference from solvent or other sources. Thus it is important to design light harvesting antennae with

chromophore-chromophore distances in mind.

The next component of this system is the polymer scaffold. As previously stated, careful regulation of chromophore positions is paramount to an effective light-harvesting antenna. To that end, the Waters group has looked to biology for an answer. Coiled-coils are a well characterized, clearly defined, and relatively stable structural motif in proteins. Specifically, they are composed of coiled α-helical peptide segments held together by hydrophobic interactions. Coiled-coils have a regular heptad repeat pattern which leads to highly predictable positions for each residue of the amino-acid sequence (see Figure 4.2). Self-assembling coiled coil peptides, have been been developed with "sticky" ends which promote long fibril formation [205]. A modified pair of complimentary peptide sequences [206] have been adopted by the Waters group for testing as a chromophore scaffold in light-harvesting antennae. The chromophores can be tethered to the peptide backbone at any of the 5 positions in the heptad repeat that do no disrupt the hydrophobic strip which holds the coil dimer together. This may allow researchers to have a great deal control over the positioning of chromophores over long distances; optimal features for designing a functional light-harvesting antenna module.

Despite the suggested simplicity of this system, and the control over tether positions, the relative positions of the chromophores themselves are more complicated. The tethers are fairly long ($>$1 nm when fully extended), flexible, mostly hydrophobic chains which allow a great deal of conformational freedom. Because of this, it is important to understand both the dynamics of the chromophores and the dynamics of the base peptide itself. To that end, we have conducted extensive molecular dynamics simulations of these modified peptide systems with differing chromophore tether positions. Our initial results suggest that chromophore dynamics do play an important role in regulating energy transfer and that conformational changes occur on the timescale of energy relaxation studies performed in the Papanikolas group.

4.2 Methods

4.2.1 Building the System

Since no crystal structure was available for the system, the initial structure was generated using PyMOL [123], which is a molecular visualization tool capable of constructing simple peptides. Peptides 1 and 2 were initially generated independently using PyMOL's helical parameters and were then manually aligned with care taken to avoid steric clashes and satisfy the hydrophobic interface. The segment which links the peptide coil to the Ruthenium chromophore is a modified lysine with four methylene units connected to a 1,4-triazole ring, with a methyl amide unit attaching to the bipyridine ligand. This segment was constructed using Gaussview, part of the Gaussian 03 suite. In order to examine the effect of linker positioning on the chromophores, the following three systems were created: In System 1, the linkers were positioned at F-F coiled coil positions replacing residues 14 and 14 on chains 1 and 2 respectively. In System 2, the linkers were positioned at C-B coiled coil positions replacing residues 11 and 10 on chains 1 and 2 respectively. In System 3, the linkers were positioned at G-E coiled coil positions replacing residues 8 and 13 on chains 1 and 2 respectively. The initial structures of these systems are shown in Figure 4.2.

4.2.2 Parameterization of the Chromophores

The simulation was prepared using we use the AMBER force field [116] with the ff99SB parameter set [117]. Since the AMBER libraries do no possess parameters for the artificial amino acids used as tethers, or for the chromophores themselves, they needed to be collected from literature or obtained through quantum calculations. Partial charges of the linker and chromophores were obtained from Gaussian calculations using restricted B3LYP [118] with the LANL2DZ basis set [119–121]. Charges derived using the restricted electrostatic potential (RESP) technique [122] gave spurious results for Ruthenium and the chelating nitrogen atoms

in the bipyridine ligands. RESP has difficulty predicting the correct charge for buried atoms since the charges are assigned in an effort to reproduce the external electrostatic potential [122]. Because of this, Mulliken charges were used in lieu of RESP charges. In general, Mulliken charges tend to be slightly more exaggerated than RESP charges with an average difference in predicted charge of $0.1(\pm 0.1)e$ for any atom aside from the Ruthenium and those atoms immediately surrounding it. There was insufficient memory to compute partial charges for the entire linker and chromophore. In order to deal with this issue, the partial charges for the base of the peptide up to the Cγ carbon of the side chain were extracted from the standard lysine amino acid residue. Force constants for Ru-N stretches, N-Ru-N (cis/trans) bends, C-C-N-Ru dihedrals, H-C-N-Ru dihedrals and van der Waals parameters were obtained from Brandt et al. [207]. Since AMBER does not explicitly support Octahedral geometry, chelating nitrogen atoms were divided into three distinctly named but chemically identical types in order to establish different bending force constants for cis and trans positions.

4.2.3 Simulation Details

Each of the three simulations were performed with $\sim$ 13000 explicit TIP3P water molecules in a box with the dimensions $\sim 75 \times 75 \times 75$ Å under periodic boundary conditions. The charge of each system was neutralized by the addition of sodium counter ions, followed by the subsequent introduction of an additional 10mM NaCl. Each system was held at constant volume, and the peptides were frozen in place while the water and ions were minimized for 200,000 steps. Subsequently, all constraints were removed from the systems and they were minimized for an additional 200,000 steps. The systems were gradually heated via Langevin temperature control to 300 K in incremental steps of 5 K every 50 ps. The production runs proceeded under the constant pressure, moderated by Langevin piston (set to 1 atm), with 2 fs time steps using the SHAKE algorithm and Ewald summation for long-range interactions. Short-range non-bonded interactions were calculated at each step, long-range interactions were

only calculated on even steps and the pair list was updated every 10 steps. System coordinates were saved every 1000 steps (2 ps) for analysis.

4.3 Results and Discussion

4.3.1 System 1 (F–F)

Both chromophores start from an extended initial conformation and quickly collapse to the peptide backbone in the direction of the N-termini in the dorsal position (see Figure 4.2 for a positional reference). The chromophore on peptide 1 (henceforth referred to as C1; the chromophore on chain 2 will be referred to as C2) quickly shifts to the ventral position and then both chromophores to explore their respective local phase spaces for $\sim$ 20 ns. At approximately 42 ns, C2 and its tether dissociate from the peptide backbone and reassociate in the direction of the C-termini, making contacts with its somewhat frayed C-terminal tail and significantly increasing the chromophore-chromophore distance. Around 90 ns, C1 dissociates and from the peptide backbone and then reassociates in the dorsal position toward the middle of the peptide, bringing the chromophores closer together on average. Finally, C1 unbinds once more and subsequently rebinds pointed toward the C-termini in the ventral position, associating with the frayed C-terminus of its parent peptide. The chromophores and tethers spend the remainder of the recorded time exploring their local phase space. The chromophore–chromophore distance distributions and the trajectory which produced it can be found in Figure 4.3. This distribution is broad (over a 3 nm range) and highly non-Gaussian. Throughout the course of this simulation, the large chromophore–chromophore separation distances observed would effectively preclude energy transfer via the Dexter mechanism. This result is consistent with experiments performed in the Papanikolas lab.

4.3.2 System 2 (C–B)

Both chromophores start from an extended initial state and C1 quickly collapses onto middle of peptide 2 in the dorsal position. After approximately 9 ns, C2 condenses onto the ventral surface of peptide 2 pointed toward the C-termini. Briefly, C2 dissociates from the backbone and comes into close proximity of C1 before collapsing back to its original position. While remaining closely bound to the peptide backbone, C2 drifts closer to C1 then both chromophores explore their local phase space for approximately 20 ns. At $\sim$ 53 ns into the simulation, C2 once again becomes unbound from the peptide backbone and then collapses onto the opposing chromophore C1. The chromophores remain tightly bound in this new state for $\sim$ 50 ns, however, during this time the peptide backbone is quite dynamic and seems to fray in response to their presence. Eventually, the frayed C-terminal tail of P2 condenses onto the chromophores. The ligands of the chromophores remain tightly stacked until $\sim$ 109 ns where fluctuations seem to disrupt the stacking for $\sim$ 10 ns. The chromophores are then briefly driven apart as the backbone of peptide 2 pulls away, initially pulling C2 with it. At the point of closest approach, the ligands of the chromophore are in Van der Waals contact and are close enough for Dexter energy transfer and the lifetime of the contact is longer than the observed timescale for energy transfer (4 ns). This result is also consistent with experiments performed in the Papanikolas lab.

4.3.3 System 3 (G–E)

In the final system, C1 quickly collapses to the dorsal portion of the N-terminal tail of its own peptide and begins to fray the coil. Meanwhile, C2 begins to drift towards C1, until finally both C2 and its tether are brought into Van der Waals contact with the opposing chromophore. After C1 and C2 have been brought together, the frayed N-terminal tail of peptide 2 migrates to a position where it lies on top of the C1 tether and inserts itself into a groove formed by

the chromophores. The orientation of the chromophores and the presence of the tail seems to disrupt the very close ligand stacking observed in System 2, however, the chromophores are still in Van der Waals contact. At $\sim$ 35 ns, the C1 tether undergoes a slight conformational rearrangement and the N-terminal tail of peptide 1 dislodges itself the grove between the chromophores. These changes allow the chromophores to make a brief transition to the tightly stacked state. This state is then disrupted by the return of the N-terminus, reformed after $\sim$ 5 ns, then disrupted again. At $\sim$ 64 ns, C1 dissociates from the peptide backbone and rebinds in a ventral position facing the C-termini while C2 remains bound to the N-terminus of peptide 2. Over the next 10 ns, C1 drifts back toward the ventral side of the N-terminus of its parent peptide. For the next 36 ns, the chromophores explore their local phase space until C1 eventually dissociates from its backbone and rebinds to C2. For the remainder of the simulation, the chromophores fluctuate from a tight ligand stacking to a more loose state. The close proximity of the ligands, even when not stacked, implies that energy transfer could occur in this system, though possibly at a diminished rate. There is currently no experimental energy transfer data to compare with the simulation results. It should be noted that in the primary conformation in which the chromophores are in Van der Waals contact, the tether of C2 is highly extended in order to reach the other chromophore.

4.3.4 General Discussion

Overall, there are several main observations from these simulations, all of which point to the importance of understanding the dynamics of these systems. First, although the coiled-coil design does offer some level of positional control over the location of chromophores, there is still a large region in which they are free to move. The chromophores and their tethers undergo multiple conformational changes over the course of the simulation which would have an impact on energy transfer. Hydrophobic collapse tends to drive the chromophores and their tethers tend to conformations in which they are bound to either the peptide backbone, or each other.

This means that very little time is spent with the chromophores in an unbound state exploring the solvent. Although the termini of the peptides spontaneously fray on their own, the chromophores seem to exacerbate this problem since they provide competition for the formation of hydrophobic contacts. This may lead to destabilization of hydrophobic strip of the coiled-coil in the real peptide systems and may drive the equilibrium toward the dissociated state. Tether position may also play a role in the level of fraying as System 1 (with the tethers placed that the center of the chain) seems to fray the least, although more simulations should be conducted to confirm this. Most importantly, these simulations shed light on the energy transfer experiments conducted by the Papanikolas lab. First, our simulations suggest that System 1 can not undergo energy transfer because the chromophores are never in close proximity. Although it is possible that we have simply not sampled enough of the phase space to observe conformations in which the chromophores come into contact, our current results for System 1 are consistent with those found in experiment. Next, our simulations of Systems 2 and 3 suggest that the chromophores prefer conformations which bring them into contact with each other. These conformations persist on a timescale that is roughly an order of magnitude larger than that of energy transfer observed from experiments suggesting it may be possible to predict energy transfer rates from distance distributions.

4.4 Summary and Future Direction:

We have conducted three long timescale simulations of coiled-coil peptide systems that have been designed for eventual use in light-harvesting antennae. The introduction of the chromophore tethers to different positions in the coiled-coil structure does play a role in regulating their position; however, the long, highly flexible tethers provide the chromophores with access to a wide range of conformations. Our simulations suggest dynamics likely play an important role in regulating energy transfer as we observe many conformational changes within the

timescale of an energy relaxation experiment. Our simulations seem to indicate that the chromophores may have a disruptive effect on the stability of the coiled-coils by promoting fraying of the termini. Finally, these simulations are consistent with energy transfer experiments conducted by the Papanikolas lab. We show that System 1 cannot undergo energy transfer because of the large chromophore-chromophore separation even at its closest approach. Likewise, we also show that Systems 2 and 3 should be able to undergo energy transfer.

This work will be extended by running additional simulations with the tethers at different positions. Additionally, the base peptide with no chromophores will be simulated for stability comparisons. In order to improve the confidence in our simulation results and sample as much of the chromophores' available phase space as possible, we will extend our current simulations to several hundred ns. In addition to atomistic simulations, it is also possible to expand the scope of our work to model energy transfer. It may be possible to use the chromophore-chromophore distance distributions to perform stochastic energy transfer simulations. It is also possible to extend our work to other light-harvesting polymer models should the need arise. Finally, to enable the investigation on a larger scale, it may be possible to create coarse grained models of polymer systems using the molecular renormalization group technique developed in our lab [180].

4.5 Acknowledgments

This work was done in collaboration with Dr. John Papanikolas, Dr. Marcey Waters, Dale Wilger and Stephanie Bettis. This work was supported by the UNC EFRC.

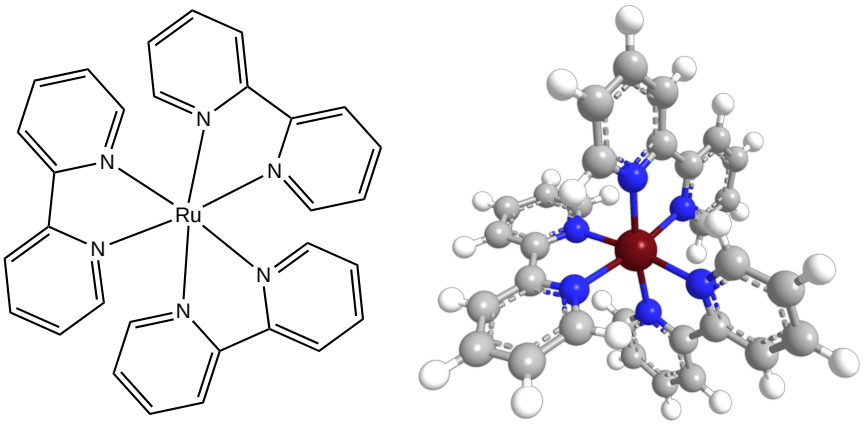


Figure 4.1: Schematic and 3-D view of $[Ru(bpy)_3]^{2+}$.

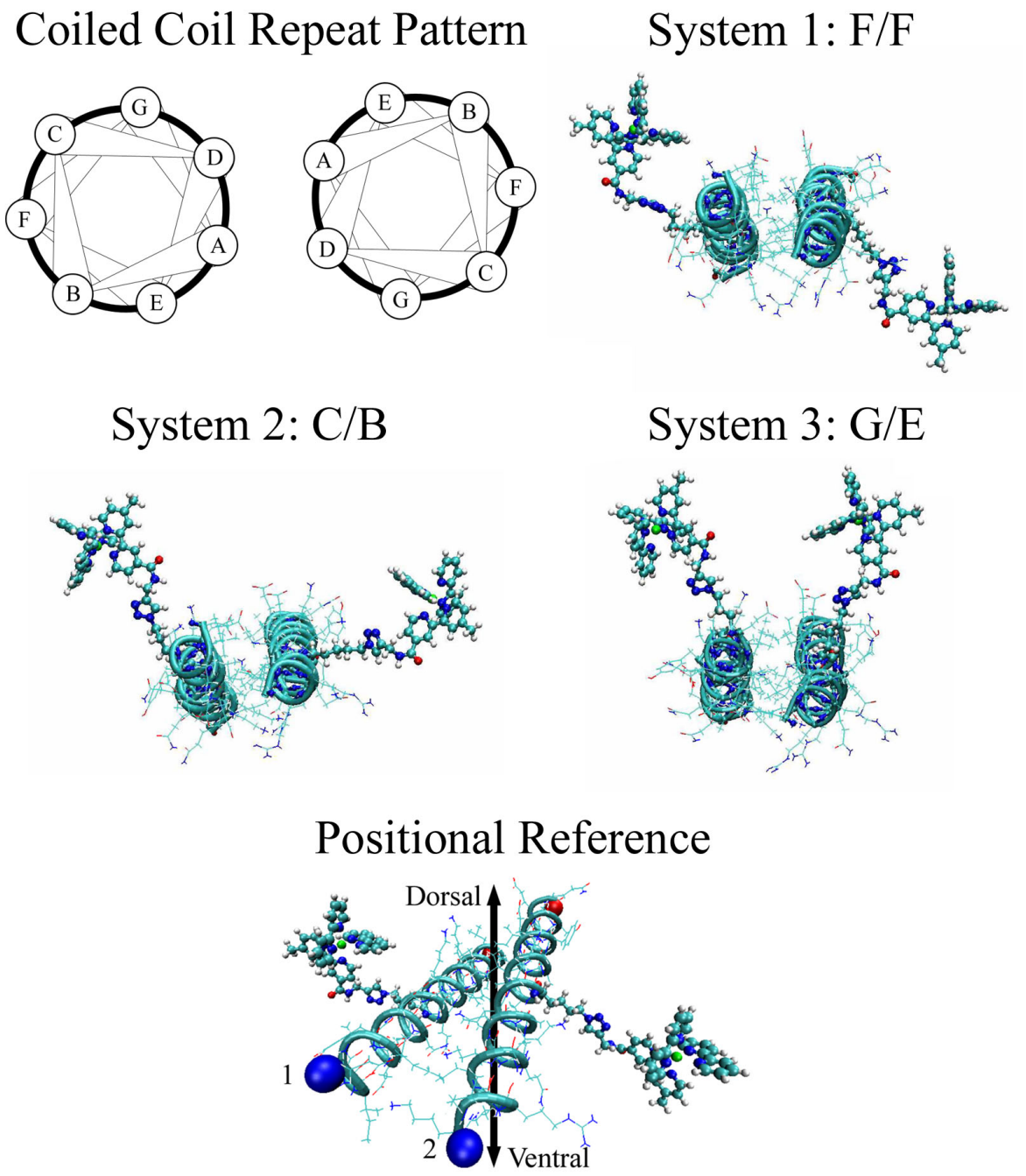


Figure 4.2: Top left: The coiled coil repeat pattern. Top right and middle: Tether positions for Systems 1, 2, and 3. Bottom: Positional reference with N-termini (blue) and C-termini (red) labeled

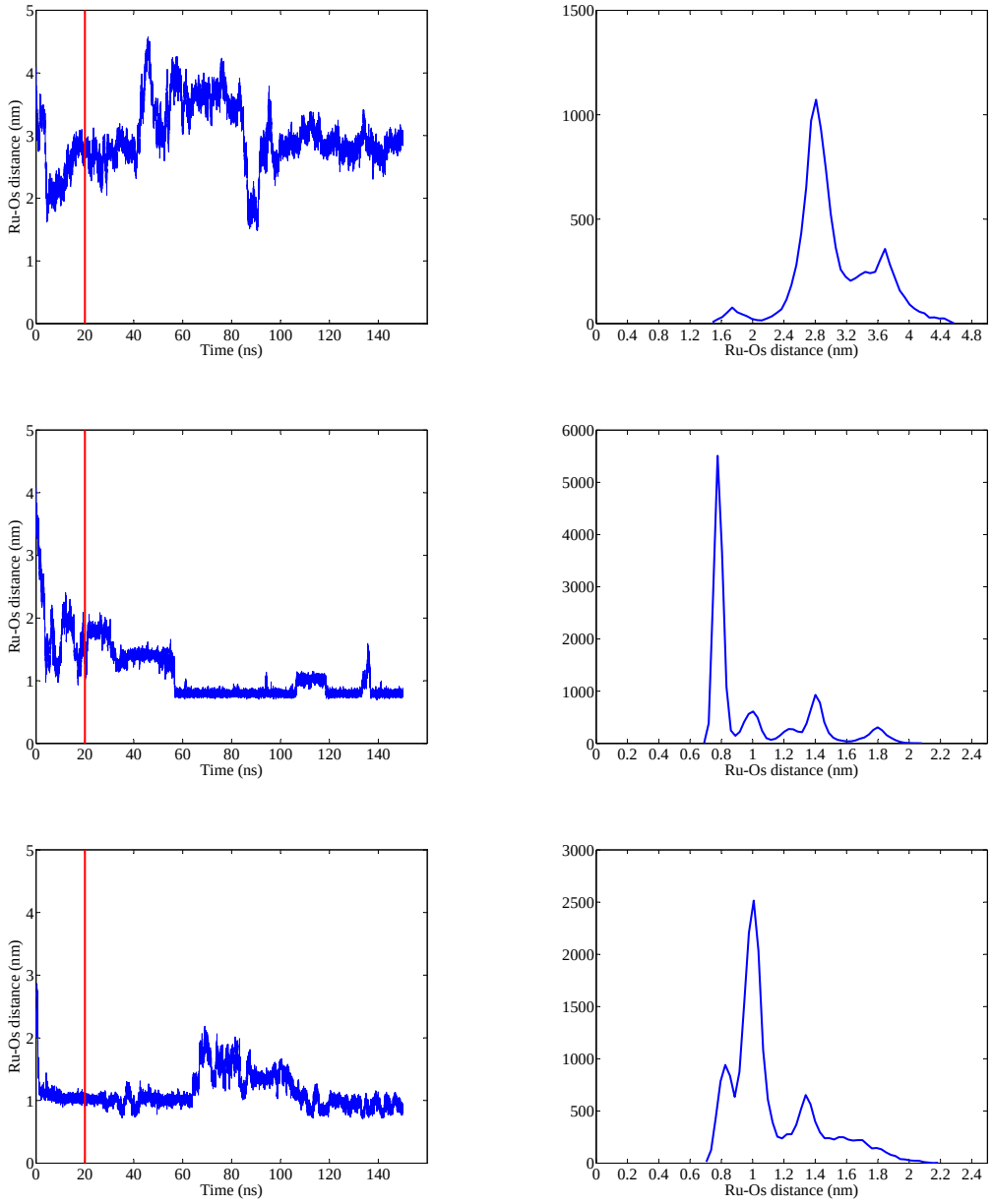


Figure 4.3: From top to bottom: Systems F/F, C/B, and G/E. Left: Ru-Os center to center distance distributions. The red line indicates the 20 ns mark before which data was discarded. Right: Ru-Os center to center distance histograms. Please note the axis scale differences.

Bibliography

[1] Levinthal, C. *J Chim Phys* **1968,** *65,* 44-45.

[2] Scheraga, H. A. *Biopolymers* **1983,** *22,* 1–14.

[3] Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991,** *254,* 1598–1603.

[4] Frauenfelder, H.; Alberding, N. A.; Ansari, A.; Braunstein, D.; Cowen, B. R.; Hong, M. K.; Iben, I. E. T.; Johnson, J. B.; Luck, S.; et al., *Journal of Physical Chemistry* **1990,** *94,* 1024-1037.

[5] Onuchic, J. N.; Wolynes, P. G.; Luthey-Schulten, Z.; Socci, N. D. *Proc Natl Acad Sci U S A* **1995,** *92,* 3626–3630.

[6] Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995,** *21,* 167–195.

[7] Pande, V. S.; Grosberg, A. Y.; Tanaka, T. *Biophys J* **1997,** *73,* 3192–3210.

[8] Pande, V. S.; Grosberg, A. Y.; Tanaka, T. *Rev Mod Phys* **2000,** *72,* 259-314.

[9] Plotkin, S. S.; Onuchic, J. N. *Q Rev Biophys* **2002,** *35,* 205–286.

[10] Scheraga, H. A.; Khalili, M.; Liwo, A. *Annu Rev Phys Chem* **2007,** *58,* 57–83.

[11] Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. *Annu Rev Biophys* **2008,** *37,* 289–316.

[12] Chen, Y.; Ding, F.; Nie, H.; Serohijos, A. W.; Sharma, S.; Wilcox, K. C.; Yin, S.; Dokholyan, N. V. *Arch Biochem Biophys* **2008,** *469,* 4–19.

[13] Grosberg, A.; Khokhlov, A. R. *Statistical Physics of Macromolecules;* Am. Inst. of Phys., Melville, NY: 1994.

[14] Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. *Fold Des* **1998,** *3,* 577–587.

[15] Sherman, E.; Haran, G. *Proc Natl Acad Sci U S A* **2006,** *103,* 11539–11543.

[16] Pande, V. S.; Grosberg, A. Y.; Tanaka, T. *Proc Natl Acad Sci U S A* **1994,** *91,* 12972–12975.

[17] Shakhnovich, E. I.; Gutin, A. M. *Proc Natl Acad Sci U S A* **1993,** *90,* 7195–7199.

[18] Henzler-Wildman, K.; Kern, D. *Nature* **2007,** *450,* 964–972.

[19] Fenimore, P. W.; Frauenfelder, H.; McMahon, B. H.; Young, R. D. *Proc Natl Acad Sci U S A* **2004,** *101,* 14408–14413.

[20] Plaxco, K. W.; Gross, M. *Nat Struct Biol* **2001,** *8,* 659–660.

[21] Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. *Proteins* **2001,** *42,* 38–48.

[22] Papoian, G. A.; Wolynes, P. G. *Biopolymers* **2003,** *68,* 333–349.

[23] McCammon, J.; Harvey, S. *Dynamics of Proteins and Nucleic Acids;* Cambridge Univ. Press, New York: 1987.

[24] McCarney, E. R.; Kohn, J. E.; Plaxco, K. W. *Crit Rev Biochem Mol Biol* **2005,** *40,* 181–189.

[25] Doi, M.; Edwards, S. *Theory of Polymer Dynamics;* Clarendon Press, Oxford, UK: 1986.

[26] Guo, Z.; Thirumalai, D. *Biopolymers* **1995,** *36,* 83 - 102.

[27] Gast, K.; Damaschun, H.; Misselwitz, R.; Mller-Frohne, M.; Zirwer, D.; Damaschun, G. *Eur Biophys J* **1994,** *23,* 297–305.

[28] Ptitsyn, O. B.; Pain, R. H.; Semisotnov, G. V.; Zerovnik, E.; Razgulyaev, O. I. *FEBS Lett* **1990,** *262,* 20–24.

[29] Ptitsyn, O. B. *Nat Struct Biol* **1996,** *3,* 488–490.

[30] Ptitsyn, O. B. *Trends Biochem Sci* **1995,** *20,* 376–379.

[31] Ptitsyn, O. B.; Uversky, V. N. *FEBS Lett* **1994,** *341,* 15–18.

[32] Ptitsyn, O. B. Protein Folding. In ; Creighton, T. E., Ed.; Freeman, New York: 1992.

[33] Ptitsyn, O. B. *Curr Opin Struct Biol* **1995,** *5,* 74–78.

[34] Ptitsyn, O. B.; Bychkova, V. E.; Uversky, V. N. *Philos Trans R Soc Lond B Biol Sci* **1995,** *348,* 35–41.

[35] Ding, F.; Jha, R. K.; Dokholyan, N. V. *Structure* **2005,** *13,* 1047–1054.

[36] Tran, H. T.; Wang, X.; Pappu, R. V. *Biochemistry* **2005,** *44,* 11369–11380.

[37] Lin, L.; Pinker, R. J.; Forde, K.; Rose, G. D.; Kallenbach, N. R. *Nat Struct Biol* **1994,** *1,* 447–452.

[38] Flanagan, J. M.; Kataoka, M.; Fujisawa, T.; Engelman, D. M. *Biochemistry* **1993,** *32,* 10359–10370.

[39] Plaxco, K. W.; Dobson, C. M. *Curr Opin Struct Biol* **1996,** *6,* 630–636.

[40] Klimov, D. K.; Thirumalai, D. *Fold Des* **1998,** *3,* 127–139.

[41] Bu, Z.; Neumann, D. A.; Lee, S. H.; Brown, C. M.; Engelman, D. M.; Han, C. C. *J Mol Biol* **2000,** *301,* 525–536.

[42] Lubchenko, V. *Proc Natl Acad Sci U S A* **2008,** *105,* 10635–10636.

[43] Lubchenko, V. *Journal of Non-Crystalline Solids* **2006,** *352,* 4400-4409.

[44] Ringe, D.; Petsko, G. A. *Biophys Chem* **2003,** *105,* 667–680.

[45] Ding, X.; Rasmussen, B. F.; Petsko, G. A.; Ringe, D. *Bioorg Chem* **2006,** *34,* 410–423.

[46] Vitkup, D.; Ringe, D.; Petsko, G. A.; Karplus, M. *Nat Struct Biol* **2000,** *7,* 34–38.

[47] Camacho, C. J.; Thirumalai, D. *Proc Natl Acad Sci U S A* **1993,** *90,* 6369–6372.

[48] Klimov, D. K.; Thirumalai, D. *Proteins* **1996,** *26,* 411–441.

[49] Schrodinger, E. *What Is Life?: The Physical Aspect of the Living Cells;* Cambridge Univ. Press, Cambridge, New York: 1944.

[50] Shakhnovich, E. I.; Finkelstein, A. V. *Biopolymers* **1989,** *28,* 1667–1680.

[51] Bryngelson, J. D.; Wolynes, P. G. *Proc Natl Acad Sci U S A* **1987,** *84,* 7524–7528.

[52] Dill, K. A.; Chan, H. S. *Nat Struct Biol* **1997,** *4,* 10–19.

[53] Yue, K.; Dill, K. A. *Proc Natl Acad Sci U S A* **1995,** *92,* 146–150.

[54] Dobson, C. M.; Sali, A.; Karplus, M. *Angew. Chem. Int. Ed.* **1998,** *37,* 868 - 893.

[55] Wolynes, P. *Proceedings of the American Philosophical Society* **2001,** *145,* 555-563.

[56] Ferreiro, D. U.; Hegler, J. A.; Komives, E. A.; Wolynes, P. G. *Proc Natl Acad Sci U S A* **2007,** *104,* 19819–19824.

[57] Sutto, L.; Latzer, J.; Hegler, J. A.; Ferreiro, D. U.; Wolynes, P. G. *Proc Natl Acad Sci U S A* **2007,** *104,* 19825–19830.

[58] Austin, R. H.; Beeson, K. W.; Eisenstein, L.; Frauenfelder, H.; Gunsalus, I. C. *Biochemistry* **1975,** *14,* 5355–5373.

[59] Ansari, A.; Hattikudur, N. S.; Joshi, S. R.; Medeira, M. A. *J Immunol Methods* **1985,** *84,* 117–124.

[60] Schwille, P.; Kummer, S.; Heikal, A. A.; Moerner, W. E.; Webb, W. W. *Proc Natl Acad Sci U S A* **2000,** *97,* 151–156.

[61] Kern, D.; Zuiderweg, E. R. P. *Curr Opin Struct Biol* **2003,** *13,* 748–757.

[62] McCammon, J. A.; Karplus, M. *Nature* **1977,** *268,* 765–766.

[63] Bahar, I.; Rader, A. J. *Curr Opin Struct Biol* **2005,** *15,* 586–592.

[64] Hayward, S.; Go, N. *Annual Review of Physical Chemistry* **1995,** *46,* 223-250.

[65] Ma, J. *Structure* **2005,** *13,* 373–380.

[66] Amadei, A.; Linssen, A. B.; Berendsen, H. J. *Proteins* **1993,** *17,* 412–425.

[67] Schmitz, M.; Tavan, P. *J Chem Phys* **2004,** *121,* 12233–12246.

[68] Ham, S.; Hahn, S.; Lee, C.; Kim, T.-K.; Kwak, K.; Cho, M. *The Journal of Physical Chemistry B* **2004,** *108,* 9333-9345.

[69] Hahn, S.; Ham, S.; Cho, M. *Journal of Physical Chemistry B* **2005,** *109,* 11789-11801.

[70] Schrader, T. E.; Schreier, W. J.; Cordes, T.; Koller, F. O.; Babitzki, G.; Denschlag, R.; Renner, C.; Loweneck, M.; Dong, S.-L.; Moroder, L.; Tavan, P.; Zinth, W. *Proc Natl Acad Sci U S A* **2007,** *104,* 15729–15734.

[71] Caflisch, A. *Curr Opin Struct Biol* **2006,** *16,* 71–78.

[72] Becker, O. M.; Karplus, M. *The Journal of Chemical Physics* **1997,** *106,* 1495-1517.

[73] Frauenfelder, H.; McMahon, B. *Proc Natl Acad Sci U S A* **1998,** *95,* 4795–4797.

[74] Wales, D. J.; Bogdan, T. V. *J Phys Chem B* **2006,** *110,* 20765–20776.

[75] Rao, F.; Caflisch, A. *J Mol Biol* **2004,** *342,* 299–306.

[76] Garcia, A. E.; Sanbonmatsu, K. Y. *Proteins* **2001,** *42,* 345–354.

[77] Becker, O. M. *Journal of Computational Chemistry* **1998,** *19,* 1255–1267.

[78] Levy, Y.; Becker, O. M. *Journal of Chemical Physics* **2000,** *114,* 993–1009.

[79] Altis, A.; Nguyen, P. H.; Hegger, R.; Stock, G. *J Chem Phys* **2007,** *126,* 244111.

[80] Hegger, R.; Altis, A.; Nguyen, P. H.; Stock, G. *Phys. Rev. Lett.* **2007,** *98,* 28102.

[81] Rammal, R.; Toulouse, G.; Virasoro, M. *Rev Mod Phys* **1986,** *58,* 765-788.

[82] Franz, S.; Mezard, M.; Parisi, G.; Peliti, L. *J Stat Phys* **1999,** *97,* 459-488.

[83] Cheung, M. S.; Garcia, A. E.; Onuchic, J. N. *Proc Natl Acad Sci U S A* **2002,** *99,* 685–690.

[84] Papoian, G. A.; Ulander, J.; Wolynes, P. G. *J Am Chem Soc* **2003,** *125,* 9170–9178.

[85] Papoian, G. A.; Ulander, J.; Eastwood, M. P.; Luthey-Schulten, Z.; Wolynes, P. G. *Proc Natl Acad Sci U S A* **2004,** *101,* 3352–3357.

[86] Szep, S.; Park, S.; Boder, E. T.; Duyne, G. D. V.; Saven, J. G. *Proteins* **2009,** *74,* 603–611.

[87] Head-Gordon, T.; Brown, S. *Curr Opin Struct Biol* **2003,** *13,* 160–167.

[88] van der Vaart, A.; Bursulaya, B. D.; Brooks, C. L.; Merz, K. M. *The Journal of Physical Chemistry B* **2000,** *104,* 9554-9563.

[89] Kaya, H.; Chan, H. S. *J Mol Biol* **2003,** *326,* 911–931.

[90] Kauzmann, W. *Adv Protein Chem* **1959,** *14,* 1–63.

[91] Dyson, H. J.; Wright, P. E.; Scheraga, H. A. *Proc Natl Acad Sci U S A* **2006,** *103,* 13057–13061.

[92] Lau, K. F.; Dill, K. A. *Macromolecules* **1989,** *22,* 3986-3997.

[93] Yue, K.; Fiebig, K. M.; Thomas, P. D.; Chan, H. S.; Shakhnovich, E. I.; Dill, K. A. *Proc Natl Acad Sci U S A* **1995,** *92,* 325–329.

[94] Honeycutt, J. D.; Thirumalai, D. *Biopolymers* **1992,** *32,* 695–709.

[95] Vasilevskaya, V.; Khalatur, P.; Khokhlov, A. *Macromolecules* **2003,** *36,* 10103-10111.

[96] Zong, C.; Papoian, G. A.; Ulander, J.; Wolynes, P. G. *J Am Chem Soc* **2006,** *128,* 5168–5176.

[97] Frauenfelder, H.; Fenimore, P. W.; Young, R. D. *IUBMB Life* **2007,** *59,* 506–512.

[98] He, Y.; Ku, P. I.; Knab, J. R.; Chen, J. Y.; Markelz, A. G. *Physical Review Letters* **2008,** *101,* 178103.

[99] Ebbinghaus, S.; Kim, S. J.; Heyden, M.; Yu, X.; Gruebele, M.; Leitner, D. M.; Havenith, M. *J Am Chem Soc* **2008,** *130,* 2374–2375.

[100] Schiessel, H. *J Phys: Condens Matter* **2003,** *15,* R699–R774.

[101] van Holde, K. E. *Chromatin;* Springer, New York: 1989.

[102] Kornberg, R.; Thomas, J. *Science* **1974,** *184,* 865–868.

[103] Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. *Nature* **1997,** *389,* 251–260.

[104] Bertin, A.; Mangenot, S.; Renouard, M.; Durand, D.; Livolant, F. *Biophys J* **2007,** *93,* 3652–3663.

[105] Savelyev, A.; Papoian, G. A. *J Am Chem Soc* **2006,** *128,* 14506–14518.

[106] Savelyev, A.; Papoian, G. A. *Mendeleev Commun* **2007,** *17,* 97–99.

[107] Savelyev, A.; Papoian, G. A. *J Am Chem Soc* **2007,** *129,* 6060–6061.

[108] Savelyev, A.; Papoian, G. A. *J Phys Chem B* **2008,** *112,* 9135–9145.

[109] Fischle, W.; Wang, Y.; Allis, C. D. *Curr Opin Cell Biol* **2003,** *15,* 172–183.

[110] Strahl, B. D.; Allis, C. D. *Nature* **2000,** *403,* 41–45.

[111] Zhou, J.; Fan, J. Y.; Rangasamy, D.; Tremethick, D. J. *Nat Struct Mol Biol* **2007,** *14,* 1070–1076.

[112] DoE, **2009,** .

[113] Pimentel, D.; Patzek, T. W. *Natural Resources Research* **2005,** *14,* 65–76.

[114] Service, R. F. *Science* **2005,** *309,* 548–551.

[115] Meyer, T. "Solar Fuels and Next Generation Photovoltaics", website,.

[116] Case, D. *et al. University of California, San Francisco.* **2008,** .

[117] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006,** *65,* 712–725.

[118] Lee, C.; Yang, W.; Parr, R. *Phys Rev B* **1988,** *37,* 785–789.

[119] Hay, P. J.; Wadt, W. R. *Journal of Chemical Physics* **1985,** *82,* 270.

[120] Hay, P. J.; Wadt, W. R. *Journal of Chemical Physics* **1985,** *82,* 284.

[121] Hay, P. J.; Wadt, W. R. *Journal of Chemical Physics* **1985,** *82,* 299.

[122] Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *Journal of Physical Chemistry* **1993,** *97,* 10269–10280.

[123] LLC, S. **2009,** .

[124] Alder, B. J.; Wainwright, T. *Journal of Chemical Physics* **1959,** *31,* 459–466.

[125] Rahman, A. *Physical Review* **1964,** *136,* 405–411.

[126] MacKerell, A. D.; Banavali, N.; Foloppe, N. *Biopolymers* **2000,** *56,* 257–265.

[127] David Case, e. a. *University of California* **2008,** .

[128] Ewald, P. *Annalen der Physik* **1921,** *369,* 253-287.

[129] McCaldon, P.; Argos, P. *Proteins* **1988,** *4,* 99–122.

[130] Fogolari, F.; Brigo, A.; Molinari, H. *J Mol Recognit* **2002,** *15,* 377–392.

[131] Gouy, M. *J. Phys* **1910,** *9,* 457–468.

[132] Chapman, D. *Phil. Mag.* **1913,** *25,* 475–481.

[133] Debye, P.; Huckel, E. *Phys. Zeitschr* **1923,** *24,* 185–206.

[134] Onsager, L. *Chemical Reviews* **1933,** *13,* 73-89.

[135] Kirkwood, J. G. *Journal of Chemical Physics* **1934,** *2,* 767–781.

[136] Manning, G. S. *J Chem Phys* **1969,** *51,* 924–933.

[137] Brooks, C. L.; Onuchic, J. N.; Wales, D. J. *Science* **2001,** *293,* 612 - 613.

[138] et al., H. F. *J. Phys. Chem.* **1990,** *94(3),* 1024-1037.

[139] Frauenfelder, H.; Wolynes, P. G. *Physics Today* **1994,** *47,* 58-64.

[140] Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. *Annu Rev Phys Chem* **1997,** *48,* 545–600.

[141] Clarkson, M. W.; Gilmore, S. A.; Edgell, M. H.; Lee, A. L. *Biochemistry* **2006,** *45,* 7693–7699.

[142] Helmstaedt, K.; Krappmann, S.; Braus, G. H. *Microbiol Mol Biol Rev* **2001,** *65,* 404–21, table of contents.

[143] Levy, Y.; Onuchic, J. N. *Annu Rev Biophys Biomol Struct* **2006,** *35,* 389–415.

[144] Van der vaart, A.; Bursulaya, B. D.; Brooks, C. L.; Merz, K. M. *J. Phys. Chem. B* **2000,** *104,* 9554-9563.

[145] Hummer, G.; Garde, S.; Garcia, A.; Pratt, L. *Chemical Physics* **2000,** *258,* 349–370.

[146] Bizzarri, A.; Salvatore, C. *Journal of Physical Chemistry B* **2002,** *106,* 6617–6633.

[147] Falconi, M.; Brunelli, M.; Pesce, A.; Ferrario, M.; Bolognesi, M.; Desideri, A. *Proteins* **2003,** *51,* 607–615.

[148] Henchman, R. H.; McCammon, J. A. *Protein Sci* **2002,** *11,* 2080–2090.

[149] Russo, D.; Hura, G.; Head-Gordon, T. *Biophys J* **2004,** *86,* 1852–1862.

[150] Peng, J. W.; Wagner, G. *Biochemistry* **1992,** *31,* 8571–8586.

[151] Peng, J. W.; Wagner, G. *Biochemistry* **1995,** *34,* 16733–16752.

[152] Hu, H.; Clarkson, M. W.; Hermans, J.; Lee, A. L. *Biochemistry* **2003,** *42,* 13856–13868.

[153] Clarkson, M. W.; Lee, A. L. *Biochemistry* **2004,** *43,* 12448–12458.

[154] Hu, H.; Hermans, J.; Lee, A. L. *J Biomol NMR* **2005,** *32,* 151–162.

[155] Kitao, A.; Hayward, S.; Go, N. *Proteins* **1998,** *33,* 496–517.

[156] Tarus, B.; Straub, J. E.; Thirumalai, D. *J Am Chem Soc* **2006,** *128,* 16159–16168.

[157] Klimov, D.; Thirumalai, D. *Physical Review Letters* **1997,** *79,* 312–320.

[158] Bursulaya, B.; Brooks, C. *Journal of the American Chemical Society* **1999,** *121,* 9947-9951.

[159] Krivov, S. V.; Karplus, M. *Proc Natl Acad Sci U S A* **2004,** *101,* 14766–14770.

[160] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J Comput Chem* **2005,** *26,* 1781–1802.

[161] Hyberts, S. G.; Goldberg, M. S.; Havel, T. F.; Wagner, G. *Protein Science* **1992,** *1,* 736-751.

[162] Lipari, G.; Szabo, A. *Journal of the American Chemical Society* **1982,** *104,* 4546-4559.

[163] Hayward, S.; Kitao, A.; Hirata, F.; Go, N. *J Mol Biol* **1993,** *234,* 1207–1217.

[164] Hayward, S.; Kitao, A.; Go, N. *Protein Sci* **1994,** *3,* 936–943.

[165] Kitao, A.; Go, N. *Curr Opin Struct Biol* **1999,** *9,* 164–169.

[166] Berendsen, H. J.; Hayward, S. *Curr Opin Struct Biol* **2000,** *10,* 165–169.

[167] Lange, O.; Grubmuller, H. *Journal of Physical Chemistry B* **2006,** *110,* 22842-22852.

[168] Mu, Y.; Nguyen, P. H.; Stock, G. *Proteins* **2005,** *58,* 45–52.

[169] Raiteri, P.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2004,** *93,*.

[170] Ming, D.; Bruschweiler, R. *J Biomol NMR* **2004,** *29,* 363–368.

[171] Lindorff-Larsen, K.; Best, R. B.; Depristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005,** *433,* 128–132.

[172] Wu, S.; Zhuravlev, P. I.; Papoian, G. A. *Biophys J* **2008,** in press.

[173] Schiessel, H.; Gelbart, W. M.; Bruinsma, R. *Biophys J* **2001,** *80,* 1940–1956.

[174] Lund, M.; Jonsson, B. *Biophys J* **2003,** *85,* 2940–2947.

[175] Chodankar, S.; Aswal, V. K. *Phys Rev E Stat Nonlin Soft Matter Phys* **2005,** *72,* 041931.

[176] Koculi, E.; Hyeon, C.; Thirumalai, D.; Woodson, S. A. *J Am Chem Soc* **2007,** *129,* 2676–2682.

[177] Woodcock, C. L.; Grigoryev, S. A.; Horowitz, R. A.; Whitaker, N. *Proc Natl Acad Sci U S A* **1993,** *90,* 9021–9025.

[178] Langowski, J.; Heermann, D. W. *Semin Cell Dev Biol* **2007,** *18,* 659–667.

[179] Sun, J.; Zhang, Q.; Schlick, T. *Proc Natl Acad Sci U S A* **2005,** *102,* 8180–8185.

[180] Savelyev, A.; Papoian, G. A. *Biophys J* **2009,** *96,* 4044-4052.

[181] Savelyev, A.; Papoian, G. A. *J Phys Chem B* **2009,** *113,* 7785-7793.

[182] Roccatano, D.; Barthel, A.; Zacharias, M. *Biopolymers* **2007,** *85,* 407–421.

[183] Ponomarev, S. Y.; Thayer, K. M.; Beveridge, D. L. *Proc Natl Acad Sci U S A* **2004,** *101,* 14771–14775.

[184] Varnai, P.; Zakrzewska, K. *Nucleic Acids Res* **2004,** *32,* 4269–4280.

[185] White, C. L.; Suto, R. K.; Luger, K. *EMBO J* **2001,** *20,* 5207–5218.

[186] Kohlbacher, O.; Lenhof, H. P. *Bioinformatics* **2000,** *16,* 815–824.

[187] Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc Natl Acad Sci U S A* **2001,** *98,* 10037–10041.

[188] Materese, C. K.; Goldmon, C. C.; Papoian, G. A. *Proc Natl Acad Sci U S A* **2008,** *105,* 10659–10664.

[189] Feig, M.; Pettitt, B. M. *Biophys J* **1999,** *77,* 1769–1781.

[190] Manning, G. S. *J Am Chem Soc* **2003,** *125,* 15087–15092.

[191] Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. *Biochim Biophys Acta* **2006,** *1764,* 1647–1676.

[192] Dobrynin, A. V.; Rubinstein, M. *Prog. Polym. Sci.* **2005,** *30,* 1049-1118.

[193] Deshkovski, A.; Obukhov, S.; Rubinstein, M. *Phys Rev Lett* **2001,** *86,* 2341–2344.

[194] Pruss, D.; Wolffe, A. P. *Biochemistry* **1993,** *32,* 6810–6814.

[195] Angelov, D.; Vitolo, J. M.; Mutskov, V.; Dimitrov, S.; Hayes, J. J. *Proc Natl Acad Sci U S A* **2001,** *98,* 6599–6604.

[196] Davey, C. A.; Sargent, D. F.; Luger, K.; Maeder, A. W.; Richmond, T. J. *J Mol Biol* **2002,** *319,* 1097–1113.

[197] Zhang, Z.; Berkowitz, M. L. *J Phys Chem B* **2009,** *113,* 7676–7680.

[198] Finkelstein, A. V.; Ptitsyn, O. B. *Protein physics;* Academic press: 2002.

[199] James H. Alstrum-Acevedo, M. K. B.; Meyer, T. J. *Inorganic Chemistry* **2005,** *44,* 6802-?6827.

[200] Demas, J. N.; Adamson, A. *Journal of the American Chemical Society* **1971,** *93,* 1800–1801.

[201] Gafney, H. D.; Adamson, A. W. *Journal of the American Chemical Society* **1972,** *94,* 8238–8239.

[202] C. R. Bock, T. J. M.; Whitten, D. G. *Journal of the American Chemical Society* **1974,** *96,* 4710–4712.

[203] Damrauer,; Cerullo,; Yeh,; Boussie,; Shank,; McCusker, *Science* **1997,** *275,* 54–57.

[204] Fleming, C. N.; Maxwell, K. A.; DeSimone, J. M.; Meyer, T. J.; Papanikolas, J. M. *J Am Chem Soc* **2001,** *123,* 10336–10347.

[205] Pandya, M. J.; Spooner, G. M.; Sunde, M.; Thorpe, J. R.; Rodger, A.; Woolfson, D. N. *Biochemistry* **2000,** *39,* 8728–8734.

[206] Ryadnov, M. G.; Woolfson, D. N. *Nat Mater* **2003,** *2,* 329–332.

[207] Brandt, P.; Norrby, T.; Akermark, B.; Norrby, P.-O. *Inorg Chem* **1998,** *37,* 4120–4127.