MOLECULAR DYNAMICS GUIDED DESIGN OF COILED-COIL BASED THERAPEUTICS

by

Scott Smith Pendley

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of requirements for the degree of

Doctor of Philosophy

Department of Pharmaceutics and Pharmaceutical Chemistry

The University of Utah

December 2010

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Scott Smith Pendley**

has been approved by the following supervisory committee members:

Thomas E. Cheatham, III	, Chair	**8/5/2010** Date Approved
James Herron	, Member	**8/5/2010** Date Approved
Donald Blumenthal	, Member	**8/9/2010** Date Approved
Carol Lim	, Member	**8/5/2010** Date Approved
Julio Facelli	, Member	**8/10/2010** Date Approved

and by **David Granger**, Chair of the Department of **Pharmaceutics and Pharmaceutical Chemistry**

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

The coiled-coil is a common protein tertiary structural motif that is composed of two or more alpha helices intertwined together to formed a supercoil. In biological systems, the coiled-coil motif often forms the oligomerization domain of various proteins including DNA binding proteins, structural and transport proteins, and cellular transport and fusion proteins. It was first described by Crick in the 1950s while describing the structure of α-keratin and has since that time been the subject of numerous engineering and mutation studies. This versatile motif has been adapted to a number of nonbiological applications including environmentally responsive hydrogels, crosslinking agents, the construction of self-assembling fibers for tissue engineering, and biosensor surfaces.

In this dissertation, we test the applicability of computational methods to understand the underlying energetics in coiled-coils as we apply molecular modeling approaches in the development of pharmaceutics. Two studies are described which test the limits of modern molecular dynamic force fields to understand the structural dynamics of the motif and to use energy calculation methodologies to predict favorable mutations for heterodimer formation and specificity. The first study considers the increasingly common use of fluorinated residues in protein pharmaceutics with regard to their incorporation in coiled-coils. Many studies find that fluorinated residues in the hydrophobic core increase protein stability against chemical and thermal denaturants. Often their incorporation fails to consider structural, energetic, and geometrical differences between these fluorinated residues and their nonfluorinated counterparts. To consider these differences, several variants of Hodges' very stable parallel heterodimer coiled-coil were constructed to examine the effect of salt bridge lengths and geometries with mixed fluorinated and nonfluorinated packed hydrophobic cores. In the second study, we collaborated with an experimental laboratory in the development of a mutant Bcr monomer with designed

mutations to increase specificity and binding to the oncoprotein Bcr-Abl for use as an apoptosis inducing agent in chronic myelogenous leukemia (CML) cells.

The final chapters of this dissertation discuss challenges and limitations that were encountered using force fields and energetic methods in our attempts to use computational chemistry to model this protein motif.

TABLE OF CONTENTS

ABSTRACT iii

ACKNOWLEDGEMENTS viii

Chapter

1. INTRODUCTION 1

The Coiled-Coil Protein Motif 1
Structure and Function 1
Pharmaceutical and Bioengineering Applications 4
Computational Approaches 4
Molecular Modeling 4
Standard Protocols 8
Molecular Dynamics Successes with Protein Simulations 9
Molecular Modeling and Simulation of Coiled-Coils 9
Computer Resources 10
Aims and Objectives 11
References 13

2. MOLECULAR DYNAMICS GUIDED STUDY OF SALT BRIDGE LENGTH DEPENDENCE IN BOTH FLUORINATED AND NON-FLUORINATED PARALLEL DIMERIC COILED-COILS 24

Abstract 25
Introduction 25
Molecular Modeling and Molecular Dynamics Simulation of Coiled-Coils 26
Methods 28
Starting Geometries 28
Force Fields 28
Molecular Dynamics Simulations 29
Residue Parameterization 29
Molecular Dynamics Trajectory Analysis 30
Thermodynamic Integration Free Energy Calculations 30
Free Energy Decomposition 31
Circular Dichroism Calculations 31
Results 31
Observation of the Simulation Results 31
Hydration of Modeled Coiled-Coils 31
Helicity Measurements 34
Free Energy Calculations 34
Rotamer Preferences 35
Salt Bridges and Electrostatics 37

Conclusions 38
Acknowledgments 38
References 38
Supplemental Materials 43

3. COMPUTATIONAL GUIDED DESIGN OF AN APOPTOSIS INDUCING BCR MUTANT FOR THE TREATMENT OF CHRONIC MYELOGENOUS LEUKEMIA 69

Abstract 69
Introduction 70
Methods 73
Starting Geometries 73
Molecular Dynamics Simulations 73
Circular Dichroism Calculations 74
Percent Helicity Calculations 74
Intrahelical Hydrogen Bonds 75
MM-PBSA 75
Thermodynamic Integration Free Energy Calculations 76
Results and Discussion 78
Conclusions 87
References 89

4. ASSESSING MODEL STRUCTURES OF SMALL, PARALLEL IAAL-E3/K3 DIMERIC COILED-COILS WITH THE AMBER FORCE FIELDS 94

Abstract 94
Introduction 94
Methods 98
Starting Geometries 98
Molecular Dynamics Simulations 101
NMR Restraints 102
Analysis 102
Results and Discussion 104
Conclusions 128
References 131

5. EXPLICIT SOLVENT EFFECTS AND PARAMETERIZATION CORRECTIONS IN MM-PBSA CALCULATIONS 137

Abstract 137
Introduction 138
Methods 141
Molecule Parameterization 141
Molecular Dynamics Simulations 142
MM-PBSA 143
Solute-Solvent van der Waals Contributions 144
Polar Surface Area 144
Water Shell Calculations 145
Water Mediated Hydrogen Bonds 145
Configurational Entropy 145
Thermodynamic Integration Free Energy Calculations 146
Results and Discussion 146
Phase Specific Enthalpy and Entropy Differences 148
Direct Calculation of Solute-Solvent van der Waals 152

Alternative Surface Area Approximations 152
Potential Sources of Error 156
Conclusion 162
References 164

6. CONCLUSIONS AND FUTURE DIRECTIONS 170

References 175

ACKNOWLEDGEMENTS

I would like to thank Dr. Thomas Cheatham for allowing me to work in his research laboratory and his patience and latitude to allow me to pursue my research. His scholarship and influence often helped me as I made the significant scholastic transition from an education in molecular biology to the interesting and frequently challenging world of computational chemistry. I am grateful for the thoughtful advice from my committee members and the time they spared to review and guide the development of this dissertation. I appreciate the help from my fellow lab mates Jianyin Shao, Sophie Cang, Niel Henrickson, and Kiumar Shahrokh, as well as Insuk Joung and Shelly Rohlstad who have recently graduated. Stephen Sontum was a good friend and his advice in my research was invaluable while he spend his sabbatical in Utah. Jianyin and Steve were often accosted by my frequent questions when my research hit an obstacle or I was hunting some other 'white whale' in my computational efforts. Financial support from NIH grants R01-GM079383-01 and R21-EB002880-02 and computer time from NSF grant MCA01S027P and the University of Utah Center for High Performance Computing (NIH 1S10RR17214-01) is acknowledged and greatly appreciated. Finally, I am grateful for the support of my friends and family who have always believed in me and repeatedly drove me forward with their questions regarding my graduation date.

CHAPTER 1

INTRODUCTION

The objective of this dissertation is to determine if modern molecular dynamics (MD) and computational approaches are sufficiently developed to design coiled-coil proteins for use as pharmaceutical therapeutics and in delivery systems. Recent advances in the parameterization of protein force fields have shown significant improvements in their ability to predict structure and behavior of proteins when applied in MD simulation. We chose to evaluate these advances using one of the most well known protein motifs and determine if this approach is mature enough to understand underlying energetics and predict beneficial mutations when used in pharmaceutical applications.

The Coiled-Coil Protein Motif

Structure and Function

The coiled-coil is a very common tertiary structural motif found in both native and engineered proteins (1). Coiled-coils are composed of two to seven right-handed, amphiphatic α-helices wound around each other in a typically left-handed oriented supercoil (1-7). Helices can be aligned in either a parallel or an anti-parallel topology (8) and both homomeric and heteromeric coiled-coils are possible (1, 9, 10). In cells, coiled-coil motifs often act as an oligomerization domain, and coiled-coils have been found in a wide variety of proteins including cytoskeletal and signal-transduction proteins (11), enzyme complexes (12), proteins involved in vesicular trafficking (13), membrane proteins (14), transcription factors (15), motor proteins (16), chaperone proteins (17), tRNA synthetases (18), SNARE complexes (19), kinases (20), and DNA-binding proteins (21). The primary sequences of coiled-coil proteins are characterized by a heptad repeat denoted as **abcdefg** (see Figure 1.1). Positions **a** and **d** are commonly

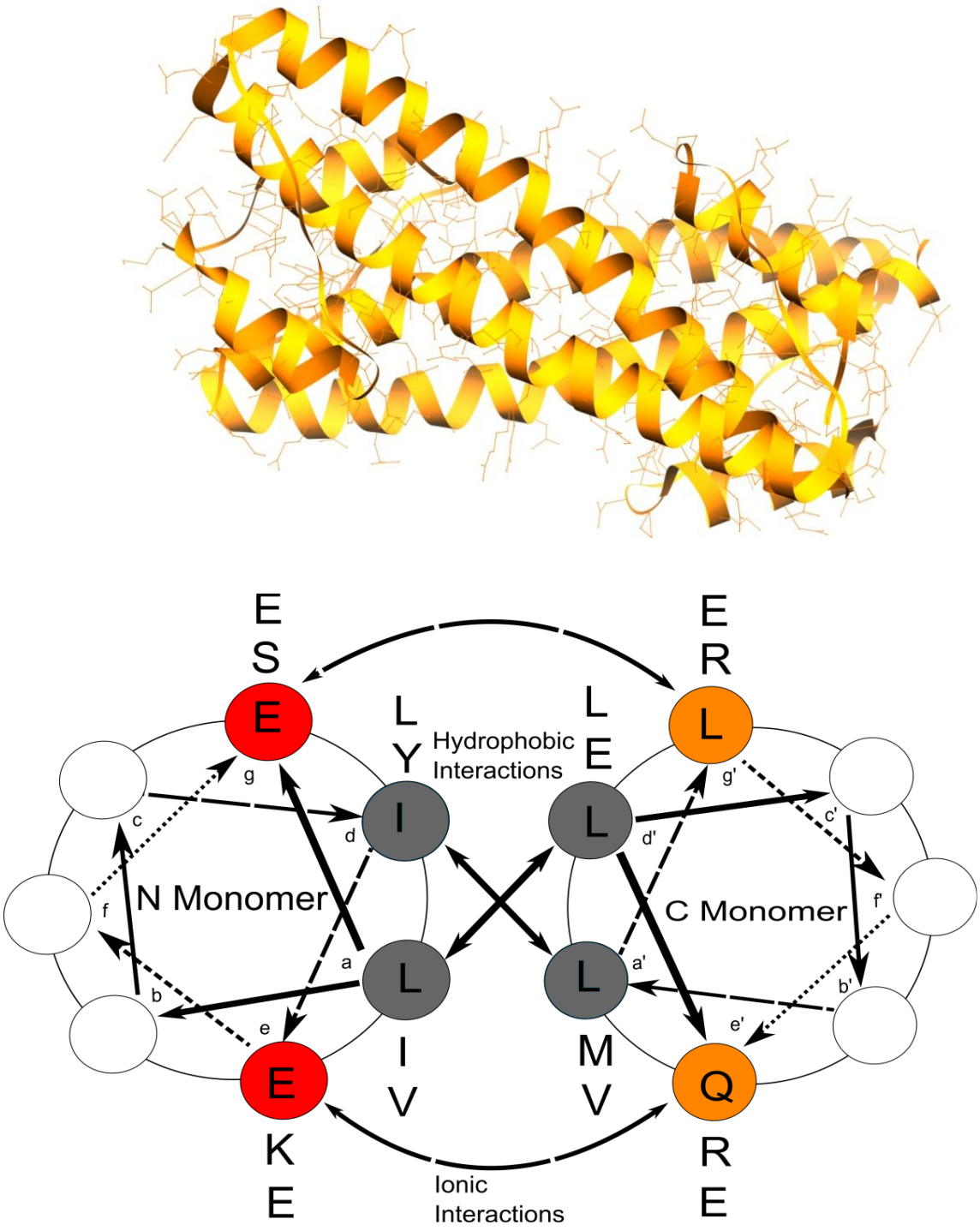


Figure 1.1 - A molecular graphics representation of the homotetramer antiparallel coiled-coil oligomerization domain of Bcr(29), and a helical wheel diagram showing amino acid interactions at the dimer interface, adapted from (30).

hydrophobic residues while positions **e** and **g** are charged residues that form an intermolecular salt bridge (22). The coiled-coil is produced when the individual helices align, bury their hydrophobic regions, and form additional stabilizing interactions such as salt bridges and hydrogen bonds. As the hydrophobic protein-protein interface gradually twists around the helix, the association of multiple helices results in a super-coil (23). Each turn of the α-helix results in the positional advancement of 3.6 residues, and the completion of each heptad repeat requires slightly less than two full twists around the helix (24). The interdigitations of the hydrophobic side chains that form the hydrophobic core is often described as “knobs into holes” packing, where a residue from one helix packs into a gap left behind by residues on an opposing helix (24). A description of this structural model was first proposed by Crick in the early 1950s from observations of α-keratin crystal structures (3, 4, 24, 25).

Two stranded coiled-coils can be found in either a parallel (aligned with peptide backbone) or anti-parallel (aligned opposite to the peptide backbone) orientation. Where coiled-coils are parallel, residues at position **a** will pack against **a’** residues of the opposing strand. Similarly, residues at position **d** will pack against **d’** residues of the opposing strand. Within the hydrophobic core, the packing requirement of parallel coiled-coils leads to two distinct layers with differing geometric requirements. In contrast, in anti-parallel coiled-coils **a** residues pack against **d’** residues on the opposing strand while **a’** residues pack against opposing **d** residues. This leads to the formation of a single hydrophobic layer in the core in anti-parallel coiled-coils (26). In both orientations, the hydrophobic side chains do not extend straight from the helix but are angled toward the amino terminus. In anti-parallel coiled-coils, optimal interactions are obtained when side-chains of opposing helices point toward each other and the C_α chains are not in register (3).

In some coiled-coils, intrahelical hydrogen bonds are formed by electrostatic interactions between positions **c** and **g** as well as **b** and **e** of the single helices. These interactions increase helicity in the coiled-coil monomers which can stabilize or destabilize the single helices and indirectly influence the stability of the α-helical coiled-coil dimer (27-29).

Helix formation (folding) and association in coiled-coils is correlated and primarily driven by hydrophobic forces, while polar residues are thought to assist in the correct aligning of

proximal helices (30, 31). Many two-stranded coiled-coils are thought to fold with a two-stage transition between the monomer and dimer oligomerization states (30). In these cases, a two stage transition may indicate that monomers are partially helical prior to association and formation of the supercoil occurs through an unstable and rate limiting transition state (30-32).

Pharmaceutical and Bioengineering Applications

The coiled-coil motif has been used in a number of applications including as a reversible dimerization domain in the development of biosensors (33-40), environmentally responsive hydrogel systems (35, 41), cross-linking agents in drug delivery applications (42), multifunctional delivery (43), targeting and imaging agents (38, 44), epitope display scaffolds (45), and for the construction of self-assembling fibers for tissue engineering applications (46). Functionally, coiled-coil motifs can act as levers, scaffolds, moving arms, and potentially as springs or nano-motors (47). Coiled-coils have also gained attention as potential pharmaceutical targets for altering protein-protein interactions for a large number of diseases (48). Strategies include a focus on the role of coiled-coils in viral infections (49, 50) and the therapeutic interference of protein-protein interactions (51, 52).

Computational Approaches

Molecular Modeling

Molecular modeling relates to the use of computational models and theoretical methods to study the structure, behavior, dynamics, and properties of a molecule or set of molecules (53-55). Accurate and useful modeling of a system requires the representation of the model to be at an appropriate level of granularity and proper sampling of the model at thermally accessible states (or conformations).

Molecular modeling can be applied at many levels, ranging from a complete *ab initio* quantum-mechanical (QM) representation of the structure and energetics to a “beads on a string” model with no implicit energy treatment (55). As each study has differing requirements and needs, various choices need to be made as to what level of molecular modeling to apply and what models to use to address the questions of the researcher. Quantum mechanical simulations

are extremely accurate but are also computationally expensive and may limit the number of atoms represented and time period sampled. At present QM is limited to small systems (<500 atoms) and short (picoseconds) time scales. Obviously, a quantum mechanical simulation would not be appropriate for sampling protein folding but may be the best choice for studying a single conformation or a small set of conformations for a molecule of <100 atoms (56). The choice in modeling can be thought of as a tradeoff between the size, the accuracy, the granularity of the system, and the time scale of the event to be modeled. Simulations of biological molecules may need to examine time scales at least as long as nano- to microsecond time periods. For these systems, all-atom molecular dynamics (MD) methods with an empirical potential may give reliable results; however, the movements of electrons are omitted or simplified (56). This simplification means that this simulation would not be appropriate for modeling chemical events involving the movement of electrons such as bond forming, bond breaking, or electron transfer, and highly polarizable metal ions are treated at a very approximate level. Despite this limitation MD simulations have proven capabilities for studying structural biology (57). Molecular modeling is not a black box. Critical evaluation of the model with experimental results is necessary to ensure the validity of the model. Higher level modeling treatments may be required if the model is unable to reproduce experimental measurements. The reader should note that approaches or models that are more precise and computational costly do not always provide more accurate insights and are not always needed when addressing a particular question or hypothesis (56).

For models that include a treatment of the underlying energetics of the system, an empirically derived or molecular mechanic (MM) potential function is the most commonly applied method for describing energy. This potential applies a simplified two-body function that has been parameterized to correctly model the system of interest. The potential energy function that models the molecular system as a function of positions and velocities of individual particles is commonly referred to as the "force field." This force field needs to represent the intramolecular interactions as well as the intermolecular interactions between all the atoms and molecules in the system. Intramolecular interactions ($U_{intramolecular}$) describe the covalent structure of the molecule. These interactions include the bonds, angles, dihedrals, and the connectivity and flexibility of the

model. In all-atom force fields, where each atom is treated independently, models of the covalent structure and intramolecular energetics are realized through energetic representations of bonds, angles and dihedrals. Often harmonic potentials are used to model bond lengths, r, and angles, θ, and Fourier terms represent rotations around dihedral bonds (or torsion angles):

$$U_{intramolecular} = \sum_{bonds} k_b(r-r_{eq})^2 + \sum_{angles} k_\theta(\theta-\theta_{eq})^2 + \sum_{dihedrals} \frac{V_\eta}{2}[1+\cos(\eta\phi-\gamma)] \quad [1.1]$$

where k_b, k_θ, and V_η are proportionality constants for the bond, angle, and dihedral terms and r_{eq} and θ_{eq} are the equilibrium bond distance and angle. The terms η, Φ, and γ are the periodicity, angle, and phase shift of the dihedral potential, respectively.

The pair-wise intermolecular energy ($U_{intermolecular}$) uses a Lennard-Jones potential (based on the distance, r_{ij}, between atoms i and j) to characterize the electron cloud repulsion (r_{ij}^{-12}), dispersion attraction (r_{ij}^{-6}) interactions (58), and a Coulombic term (with point charges q_i and q_j for atoms i and j, and the dielectric constant of the surrounding environment (ε) is set to unity in explicit solvent with a pre-factor for proper units of $4\pi\varepsilon_0$ where ε_0 is the permittivity of free space to represent the electrostatic interactions between all the atom pairs (55):

$$U_{intermolecular} = \frac{1}{2} \sum_{nonbonded(i,j)} \left(\frac{A_{ij}}{r_{ij}^{12}}\right) - \left(\frac{B_{ij}}{r_{ij}^{6}}\right) + \left(\frac{q_i q_j}{4\pi\varepsilon\varepsilon_0 r_{ij}}\right) \quad [1.2]$$

where A_{ij} and B_{ij} are Lennard-Jones coefficients for atoms i and j. Intermolecular interactions between atoms separated by three or less covalent bonds (1-1, 1-2, 1-3) are typically omitted from the calculation while atoms separated by four covalent bonds (1-4) are scaled.

For the simulations presented in this dissertation, molecular modeling will be considered at the all-atom molecular dynamics level of granularity. Molecular dynamics simulations follow Newton's equations of motion to unfold consecutive events as time progresses (57, 59-62). The classical equations of motion are integrated numerically for all atoms in the system (N). When an

MD simulation is started, random velocities, v_i, are assigned to each of the N particles (of mass m_i) as determined from a Boltzmann distribution at a given temperature, T, where k_B is the Boltzmann constant.

$$\frac{1}{2}\sum_{i}^{N} m_i v_i^2 = \frac{3Nk_BT}{2} \quad [1.3]$$

The dynamics are then initiated by integrating Newton's equations of motion using the pair-wise potential, U_i, to determine the acceleration, a_i, of the individual atoms based on the force experienced due to the proximity of the atoms to each other, $a_i = F_i/m_i$. This force is calculated by considering the differential of the potential energy of atom i, U_i, with respect to three-dimensional space by considering intramolecular and intermolecular interactions between atom i and all other potential atom pairs:

$$F_i = m_i \frac{d^2 r_i}{dt^2} = -\nabla_i \sum_{j}^{N} U_{ij} \quad [1.4]$$

where U_{ij} is the pair-wise potential energy between atoms i and j and ∇ is the vector differential operator. Molecular dynamics simulations require the calculation of the forces and velocities at each step in order to generate coordinates for the next time step. Since spatial and conformational sampling is based on the dynamic propagation of the molecular mechanical forces, it is requisite that the forces are accurate derivatives of the energy. For large systems, a simplification can be used to reduce the effective number of pair interactions to only those within a given range and a list of in-range pair interactions is maintained for each atom.

Dynamic properties and behaviors are very sensitive to the energy potential. Since the force field is primarily parameterized to represent structure, it is not always apparent that dynamic properties will be well reproduced (56). Often the gold standard of force field evaluation is how well force fields can model biological behavior and how well those representations correspond with the experimentally observed system (63-66). The accuracy of MD simulation is challenged

when modern simulation models fail to reproduce real properties of the molecular systems they are trying to represent (67, 68). With the advances of processor speed and parallelization, models that have previously shown high fidelity at modeling complex biological dynamics or structure may show errant behavior at longer simulation timescales where additional sampling is requisite to find the limitations of the original parameterization. When the reliability of these models is in doubt, a new implementation of the models or models with finer granularity are suggested to increase the accuracy.

Standard Protocols

Simulations in this dissertation used all-atom molecular dynamic approaches that considered the full energetic contribution in proteins and small organic molecules. All simulations and energetic determinations were performed in a fashion consistent with the AMBER suite of programs (69) using modern, established force fields and explicit solvent (70) with neutralizing salt in a periodic boundary system (71, 72). Proteins simulations used the AMBER ff03 (73) and ff99SB (68) force fields while small molecule simulations were performed using the general AMBER force field (GAFF) (74). Minimization was performed in two stages (restrained substrate and then unrestrained system) to reduce energetic clashes within the starting structure. During equilibration the substrate was once again restrained under constant volume as the system temperature was raised from 0 K to 300 K. An additional stage of equilibration removed the structural restraints once the target temperature was reached. Production MD simulation is then performed under constant temperature and pressure. Pressure in the periodic box was controlled using Berendsen's coupling algorithm (75) where the periodic box size is adjusted in order to maintain a constant pressure. Temperature regulation was handled in two ways. Older simulations used Berendsen's coupling algorithm (75) which linearly rescaled the velocities of individual atoms to maintain the velocity distribution and therefore the system temperature. In recent simulations, we have switched to the Langevin stochastic algorithm (76) which introduces random forces acting on atoms to maintain atomic velocities and system temperature.

Molecular Dynamics Successes with Protein Simulations

Approaches similar to those that we use in this dissertation have been successfully used to reproduce protein structure and behavior. Recent molecular dynamic simulations have been shown to be able to correctly fold and reproduce the structure of the β-hairpin trpzip2 (77), the trpcage (78, 79), the trypsin inhibitor CMTI-1 (80), the C-terminal hairpin of protein G (81), the acyl-carrier protein bound to FabI (82), and the villin headpiece subdomain (83, 84). Molecular models of hen egg white lysozyme and ubiquitin were able to reproduce NMR order parameters of the same structures showing that MD approaches were sufficient to capture both the structure and dynamic behavior of medium-sized proteins (85). Simulations of HIV-1 protease reported from the same research group were able to reproduce experimental observations of active site closure following docking with an inhibitor (85).

Force fields that were used in our simulations have also been show to be very accurate when used as a scoring function to discriminate between native structures and protein decoys (86) and to correctly predict the folding of loop regions inside of larger proteins (87-89).

Molecular Modeling and Simulation of Coiled-Coils

Initial modeling work of coiled-coils focused on extending a set of general formulas and principles developed by Crick to describe ideal parallel, coiled-coil dimer structure (24). More advanced modeling and MD simulation studies were later applied to reproduce experimental structure and dynamics, coiled-coil oligomerization states, and ultimately to estimate relative energies of coiled-coil association and folding. Early modeling studies used rigid backbones and side chain packing in the hydrophobic domain of parallel coiled-coils (90). As more structures were published and the regularity of the coiled-coil structure was confirmed, modeling approaches expanded and various methods and programs to build coiled-coil structures were developed. These approaches used combinations of means to pack the hydrophobic centers, optimize the electrostatics, investigate the flexibility of side chains, and/or determine preferential orientations and oligomeric states (91-122).

Simulation approaches beginning in the early 1990s using molecular dynamics with explicit representations did surprisingly well in reproducing coiled-coil structure. This includes

unrestrained molecular dynamics of the p1 coiled-coil region of the leucine zipper GCN4 starting from Cα atoms followed by automatic building with simulated annealing to produce a structure with a 1.25 Å backbone root mean square deviation (RMSd) from the crystal structure (123, 124). Protein engineering experiments based on the GCN4 structure would continue well into the next decade. Improvements to this model were made using Monte Carlo folding simulations followed by all-atom MD on the GCN4 dimer predicted structure to within 0.81 Å (backbone). The success of these calculations advanced future quaternary structure predictions (97, 102, 105) and led to MD guided rational design of coiled-coil protein sequences to qualitatively understand stability and predict stabilizing intrahelical hydrogen bonds (96).

Successful calculation of structure led to computational studies into coiled-coil energetics including predictions aimed at understanding contributions to oligomerization states (109, 122, 125-128), helical propensity (94), and stability (119). Notable application specific challenges included investigations into the influence of a membrane environment (111, 120, 129-131) and an external electrostatic field (129) on coiled-coil structure and dynamics and the use of targeted or force-induced MD simulations to investigate alterations in coiled-coil structure and dynamics between different conformations of seryl tRNA synthetase (132), influenza hemaglutinin (125), and engineered prion peptides (133).

Computer Resources

The simulations and energy calculations discussed in this dissertation were run on the computer clusters located in the Center for High Performance Computing (CHPC) at the University of Utah and the Ranger linux cluster located at the University of Texas. The Ranger cluster is composed of 16-way SMP 2.3 GHz AMD opteron processor nodes with Infiniband interconnects which allow normal (24 hour) and long (48 hour) simulation runs. Most of the simulations using the CHPC machines were run on two clusters: Sanddunearch and Updraft. The Sanddunearch cluster is used for more time intensive simulations (up to 36 hours) and is composed of two dual core 2.4 GHz AMD opteron processor nodes using Gigabit Ethernet interconnects. The Updraft cluster uses dual-quad core 2.8 GHz Intel Xeon processors with Infiniband DDR interconnects for simulation runs up to 24 hours. During the completion of our

studies, we have seen remarkable advances in computation power and simulation speed in the systems we have used. Initial studies, using small, parallel coiled-coils, were performed on a single node quad-core opteron machine and as simulations grew more intensive we moved to a heterogeneous Beowulf cluster of dual and quad-core 32 bit AMD Athlon processor machines (Icebox). With these early machines we could complete 2-5 ns of production simulation time with 32 dedicated processors running for 72 hours using the IAAL-E3/K3 coiled-coil. For comparison, using the same model on the Updraft cluster at the CHPC, we can complete 15 ns of production simulation in 24 hours using 32 dedicated processors. Similar simulations of the larger coiled-coil region of the Bcr protein using Updraft can complete approximately 6 ns of production simulation in 24 hours. Thermodynamic integration (TI) calculations are significantly more computationally expensive than production simulations. For the mutation studies of the Bcr coiled-coil reported in Chapter 3, a single 6ns thermodynamic integration conversion step required 230,400 processor hours on the Ranger cluster.

Aims and Objectives

The use of coiled-coils in pharmaceuticals and bioengineering is rapidly increasing (5, 10, 38, 134-145). The development of this motif in drug delivery and discovery requires a detailed knowledge of the underlying energetics in order to insure delivery of a system that provides a pharmaceutical response at nanomolar or picomolar concentrations. Molecular dynamics and other computational approaches can provide insights into dynamics as well as structure of proteins. Further, energetic analysis methods can allow us to gain insights into the stabilizing interactions of these bio-molecules and allow predictions of future structural modifications to improve protein behavior and desirable pharmaceutical traits.

The bulk of this dissertation and the studies comprising the following two chapters focus on two projects that we have pursued to test the applicability of computational methods in the design of coiled-coil pharmaceutics. In Chapter 2 we consider a fluorinated coiled-coil based on Hodges ultra-stable, parallel, heterodimer coiled-coil IAAL-E3/K3 (146). The use of fluorinated residues in coiled-coils has increased the thermostability of the engineered peptides but has in some case decreased secondary structure. While fluorinated residues are typically considered

isomorphic to their nonfluorinated pairs, the increased size of the fluorine atom relative to hydrogen may have important structure and energetic consequences. In this study, residues in salt bridge positions of the acidic monomer were mutated to find optimal salt bridge lengths and geometries. The structural and energetic analysis of these mutations were performed and described thoroughly in an article in Proteins: Function, Structure, and Bioinformatics of which Chapter 2 is a reprint (147). Importantly our findings help to bridge other experimental studies focusing on the incorporation of fluorinated molecules. Our work shows that while fluorinated molecules are not isomorphic to their nonfluorinated pairs, they adopt differing side-chain rotamers leading to a more rigid and packed core. This increase in core density allows a similar outward structural profile to nonfluorinated coiled-coils. This work also shows that the increased stability against chemical and thermal denaturant seen in fluorinated proteins can be attributed to a decrease in the free energy of solvation in fluorinated residues that drives association and folding.

In Chapter 3, results of a collaborative study with the Lim laboratory are presented. This study focused on engineering mutations to the wild type Bcr coiled-coil to improve the binding strength and specificity of the heterodimer form for the Bcr-Abl oncoprotein in order to be used as an apoptosis-inducing agent in the treatment of chronic myelogenous leukemia (CML). Our efforts in this study were focused on using molecular dynamics simulations to design coiled-coils incorporating ration designed mutations and protein engineering approaches that would increase the free energy of binding while maintaining heterodimer specificity. Initial course-grain predictions of favorable mutations were determined using percent helicity and circular dichroism calculations from stabilized simulations using the Dichrocalc program (148) with the Woody et al. semiempirical parameter set (149). For these approximations, the correlation of secondary structure and association was used to predict improvement to the free energy of binding. More fine-grain approaches followed using MM-PBSA energy calculations and umbrella sampling of the forced separation of the coiled-coil pairs to create a potential of mean force (PMF) to correctly rank homo- and hetero-dimer coiled-coils with respect to their free energy of binding. Using these energy calculations we were able to successfully design a mutant to the Bcr coiled-coil that

showed improved binding compared to the wild type protein while maintaining heterodimer specificity.

Research in the discipline of computational chemistry is focused not only on the application of the science but the further development of the techniques and science to overcome challenges. Chapters 4 and 5 discuss challenges, limitations, and application specific approaches of molecular dynamics to the study of coiled-coils. Chapter 4 focuses on the evaluation of modern AMBER force field models to correctly model structure and behavior of the IAAL-E3/K3 coiled-coil that is first introduced in Chapter 2. While none of the force fields studied were able to show long term stable reproduction of the NMR structure, it was found the short scale models using the ff99SB force field were fairly accurate at modeling structure and dynamics. Personal communication with the author of the ff99SB force field suggests that future force field improvements will address deficiencies found in our study. Chapter 5 comprises studies on improvements to the MM-PBSA energy calculation methodology. In coiled-coils, folding is correlated with association. Small errors in the MM-PBSA approach in calculating the nonpolar contribution to the free energy of solvation using a surface area approach lead to significant quantitative and qualitative errors when calculating the free energy of binding of coiled-coils. We found that improvement to this calculation can be made by directly calculating the solute-solvent van der Waals energetics and by the introduction of a combined surface area/ polar surface area term. While these corrections do improve the calculation of free energies of hydration they fail to correct for errors introduced from compounds containing two or more polar groups. Ultimately these errors arise due to complex interactions between the polar atoms and the solute molecules. Future directions to address the source of these errors will need to consider more solvent-based approaches.

References

1. Woolfson D. The design of coiled-coil structures and assemblies. Advances in Protein Chemistry 2005: 79-112.

2. Adamson JG, Zhou NE, Hodges RS. Structure, function and application of the coiled-coil protein folding motif. Current Opinions in Biotechnology 1993; 4: 428-437.

3. Lupas A. Coiled coils: new structures and new functions. Trends in Biochemical Sciences 1996; 21: 375-382.

4. Burkhard P, Stetefeld J, Strelkov SV. Coiled coils: a highly versatile protein folding motif. Trends in Cell Biology 2001; 11: 82-88.

5. Mason J, Arndt K. Coiled coil domains: stability, specificity, and biological implications. ChemBioChem 2004; 5: 170-176.

6. Lupas AN, Gruber M. The structure of alpha-helical coiled coils. Advances in Protein Chemistry 2005; 70: 37-78.

7. Liu J, Zheng Q, Deng Y, Cheng CS, Kallenbach NR, Lu M. A seven-helix coiled coil. Proceedings of the National Academy of Sciences, USA 2006; 103: 15457-15462.

8. Harbury P, Zhang T, Kim P, Alber T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science 1993; 262: 1401-1407.

9. Laage R, Rohde J, Brosig B, Langosch D. A conserved membrane-spanning amino acid motif drives homomeric and supports heteromeric assembly of presynaptic SNARE proteins. Journal of Biological Chemistry 2000; 275: 17481.

10. Boice J, Dieckmann G, DeGrado W, Fairman R. Thermodynamic analysis of a designed three-stranded coiled coil. Biochemistry 1996; 35: 14480-14485.

11. Stock JB, Surette MG, McCleary WR, Stock AM. Signal transduction in bacterial chemotaxis. Journal of Biological Chemistry 1992; 267: 19753-19756.

12. Peng H, Begg GE, Schultz DC, et al. Reconstitution of the KRAB-KAP-1 repressor complex: a model system for defining the molecular anatomy of RING-B box-coiled-coil domain-mediated protein-protein interactions. Journal of Molecular Biology 2000; 295: 1139-1162.

13. Raiborg C, Bremnes B, Mehlum A, et al. FYVE and coiled-coil domains determine the specific localisation of Hrs to early endosomes. Journal of Cell Science 2001; 114: 2255-2263.

14. Langosch D, Heringa J. Interaction of transmembrane helices by a knobs-into-holes packing characteristic of soluble coiled coils. Proteins: Structure, Function, and Genetics 1998; 31.

15. Yang J, Wang H. The central coiled-coil domain and carboxyl-terminal WD-repeat domain of Arabidopsis SPA1 are responsible for mediating repression of light signaling. Plant Journal 2006.

16. Hildebrandt ER, Gheber L, Kingsbury TJ, Hoyt MA. Homotetrameric form of CIN8P, an S. cerevisiae kinesin-5 motor, is essential for its in vivo function. Journal of Biological Chemistry 2006.

17. Watanabe YH, Takano M, Yoshida M. ATP binding to nucleotide binding domain (NBD)1 of the ClpB chaperone induces motion of the long coiled-coil, stabilizes the hexamer, and activates NBD2. Journal of Biological Chemistry 2005; 280: 24562-24567.

18. Fukai S, Nureki O, Sekine S, Shimada A, Vassylyev DG, Yokoyama S. Mechanism of molecular interactions for tRNA(Val) recognition by valyl-tRNA synthetase. RNA 2003; 9: 100-111.

19. Sorensen JB, Wiederhold K, Muller EM, et al. Sequential N- to C-terminal SNARE complex assembly drives priming and fusion of secretory vesicles. The EMBO Journal 2006; 25: 955-966.

20. Ishizaki T, Naito M, Fujisawa K, et al. p160ROCK, a Rho-associated coiled-coil forming protein kinase, works downstream of Rho and induces focal adhesions. FEBS Letters 1997; 404: 118-124.

21. Landschulz WH, Johnson PF, McKnight SL. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 1988; 240: 1759-1764.

22. Mason JM, Muller KM, Arndt KM. Considerations in the design and optimization of coiled coil structures. Methods in Molecular and Cellular Biology 2007; 352: 35-70.

23. Woolfson. An Introduction to Coiled Coils. 2000 [cited 2006 June 6 2006]; Available from: http://www.lifesci.sussex.ac.uk/research/Woolfson/html/coils.html

24. Crick F. The packing of alpha helices: simple coiled-coils. Acta Crystallographica 1953; 6: 689-697.

25. DeGrado W, Summa C, Pavone V, Nastri F, Lombardi A. De novo design and structural characterization of proteins and metalloproteins. Annual Review of Biochemistry 1999; 68: 779-819.

26. Oakley M, Hollenbeck J. The design of antiparallel coiled coils. Current Opinion in Structural Biology 2001; 11: 450-457.

27. Kohn W, Kay C, Hodges R. Salt effects on protein stability: two-stranded [alpha]-helical coiled-coils containing inter-or intrahelical ion pairs. Journal of Molecular Biology 1997; 267: 1039-1052.

28. Spek E, Bui A, Lu M, Kallenbach N. Surface salt bridges stabilize the GCN4 leucine zipper. Protein Science 1998; 7: 2431.

29. Kohn W, Kay C, Hodges R. Protein destabilization by electrostatic repulsions in the two-stranded alpha-helical coiled-coil/leucine zipper. Protein Science 1995; 4: 237.

30. Bosshard HR, Durr E, Hitz T, Jelesarov I. Energetics of coiled coil folding: the nature of the transition states. Biochemistry 2001; 40: 3544-3552.

31. Ibarra-Molero B, Zitzewitz J, Matthews C. Salt-bridges can stabilize but do not accelerate the folding of the homodimeric coiled-coil peptide GCN4-p1. Journal of Molecular Biology 2004; 336: 989-996.

32. Skolnick J, Kolinski A, Mohanty D. De novo predictions of the quaternary structure of leucine zippers and other coiled coils. International Journal of Quantum Chemistry 1999; 75: 165-176.

33. Gonzalez L, Jr., Plecs JJ, Alber T. An engineered allosteric switch in leucine-zipper oligomerization. Nature: Stuctural Biology 1996; 3: 510-514.

34. Chao H, Bautista DL, Litowski J, Irvin RT, Hodges RS. Use of a heterodimeric coiled-coil system for biosensor application and affinity purification. Journal of Chromatography B: Biomedical Sciences and Applications 1998; 715: 307-329.

35. Wang C, Stewart RJ, Kopecek J. Hybrid hydrogels assembled from synthetic polymers and coiled-coil protein domains. Nature 1999; 397: 417-420.

36. Tang A, Wang C, Stewart RJ, Kopecek J. The coiled coils in the design of protein-based constructs: Hybrid hydrogels and epitope displays. Journal of Controlled Release 2001; 72: 57-70.

37. Litowski JR, Hodges RS. Designing heterodimeric two-stranded alpha-helical coiled-coils. Effects of hydrophobicity and alpha-helical propensity on protein folding, stability, and specificity. Journal of Biological Chemistry 2002; 277: 37272-37279.

38. Yu YB. Coiled-coils: stability, specificity, and drug delivery potential. Adv Drug Deliv Rev 2002; 54: 1113-1129.

39. Ryan SJ, Kennan AJ. Variable stability heterodimeric coiled-coils from manipulation of electrostatic interface residue chain length. Journal of the American Chemical Society 2007; 129: 10255-10260.

40. Doerr AJ, McLendon GL. Design, folding, and activities of metal-assembled coiled coil proteins. Inorganic Chemistry 2004; 43: 7916-7925.

41. Qiu Y, Park K. Environment-sensitive hydrogels for drug delivery. Advanced Drug Delivery Reviews 2001; 53: 321-339.

42. Tomishige M, Vale R. Controlling kinesin by reversible disulfide cross-linking identifying the motility-producing conformational change. Journal of Cell Biology 2000; 151: 1081-1092.

43. Gariépy J, Kawamura K. Vectorial delivery of macromolecules into cells using peptide-based vehicles. Trends in Biotechnology 2001; 19: 21-28.

44. Yu YB. Fluorocarbon nanoparticles as multifunctional drug delivery vehicles. Journal of Drug Targeting 2006; 14: 663-669.

45. Nygren P, Uhlen M. Scaffolds for engineering novel binding sites in proteins. Current Opinion in Structural Biology 1997; 7: 463-469.

46. Ryadnov M, Woolfson D. Engineering the morphology of a self-assembling protein fibre. Nature Materials 2003; 2: 329-332.

47. Rose A, Meier I. Scaffolds, levers, rods and springs: Diverse cellular functions of long coiled-coil proteins. Cellular and Molecular Life Sciences 2004; 61: 1996-2009.

48. Cochran A. Antagonists of protein–protein interactions. Chemistry & Biology 2000; 7: 85-94.

49. Singh M, Berger B, Kim P. LearnCoil-VMF: computational evidence for coiled-coil-like motifs in many viral membrane-fusion proteins1. Journal of Molecular Biology 1999; 290: 1031-1041.

50. Chan D, Chutkowski C, Kim P. Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. Proceedings of the National Academy of Science 1998; 95: 15613.

51. Yamamoto Y, Gaynor R. Therapeutic potential of inhibition of the NF-kappaB pathway in the treatment of inflammation and cancer. Journal of Clinical Investigation 2001; 107: 135-142.

52. Ladner R, Sato A, Gorzelany J, de Souza M. Phage display-derived peptides as therapeutic alternatives to antibodies. Drug Discovery Today 2004; 9: 525-529.

53. Schlick T. Molecular modeling and simulation: an interdisciplinary guide. New York: Springer; 2002.

54. Burkert U, Allinger N. Molecular mechanics. Washington: American Chemical Society; 1982.

55. Leach A. Molecular modelling: principles and applications. Essex: Addison-Wesley Longman Ltd; 1996.

56. Cheatham T, Brooks B, Kollman P. Molecular Modeling of Nucleic Acid Structure: Energy and Sampling. Current Protocols in Nucleic Acid Chemistry. New York: John Wiley & Sons, Inc; 2001. p. 7.8.1-7.8.20.

57. Karplus M, McCammon J. Molecular dynamics simulations of biomolecules. Nature Structural & Molecular Biology 2002; 9: 646-652.

58. Wilhelm E, Battino R. Estimation of Lennard Jones (6, 12) pair potential parameters from gas solubility data. The Journal of Chemical Physics 1971; 55: 4012.

59. Haile J. Molecular dynamics simulation: elementary methods. New York: John Wiley & Sons, Inc.; 1992.

60. van Gunsteren W, Berendsen H. Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. Angewandte Chemie International Edition in English 1990; 29: 992-1023.

61. Hoover W. Molecular dynamics. Berlin: Springer-Verlag; 1986.

62. Karplus M, Petsko G. Molecular dynamics simulations in biology. Nature 1990; 347: 631 - 639.

63. Daura X, Jaun B, Seebach D, van Gunsteren W, Mark A. Reversible peptide folding in solution by molecular dynamics simulation. Journal of Molecular Biology 1998; 280: 925-932.

64. Chandrasekhar I, Kastenholz M, Lins R, et al. A consistent potential energy parameter set for lipids: dipalmitoylphosphatidylcholine as a benchmark of the GROMOS96 45A3 force field. European Biophysics Journal 2003; 32: 67-77.

65. Loncharich R, Brooks B. Temperature dependence of dynamics of hydrated myoglobin: comparison of force field calculations with neutron scattering data. Journal of Molecular Biology 1990; 215: 439-455.

66. Daggett V, Levitt M. Realistic simulations of native-protein dynamics in solution and beyond. Annual Review of Biophysics and Biomolecular Structure 1993; 22: 353-380.

67. Pérez A, Marchán I, Svozil D, et al. Refinement of the AMBER force field for nucleic acids: improving the description of / conformers. Biophysical Journal 2007; 92: 3817-3829.

68. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 2006; 65: 712-725.

69. Case DA, Cheatham TE, 3rd, Darden T, et al. The Amber biomolecular simulation programs. Journal of Computational Chemistry 2005; 26: 1668-1688.

70. Linge J, Williams M, Spronk C, Bonvin A, Nilges M. Refinement of protein structures in explicit solvent. Proteins: Structure, Function, and Bioinformatics 2003; 50: 496-506.

71. Essmann U, Perera L, Berkowitz M, Darden T, Lee H, Pedersen L. A smooth particle mesh Ewald method. Journal of Chemical Physics 1995; 103: 8577-8593.

72. Cheatham T, Miller J, Fox T, Darden T, Kollman P. Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. Journal of the American Chemical Society 1995; 117: 4193-4194.

73. Duan Y, Wu C, Chowdhury S, et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. Journal of Computational Chemistry 2003; 24: 1999-2012.

74. Wang J, Wolf R, Caldwell J, Kollman P, Case D. Development and testing of a general Amber force field. Journal of Computational Chemistry 2004; 25: 1157-1174.

75. Berendsen H, Postma J, Van Gunsteren W, DiNola A, Haak J. Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics 1984; 81: 3684.

76. Izaguirre J, Catarello D, Wozniak J, Skeel R. Langevin stabilization of molecular dynamics. The Journal of Chemical Physics 2001; 114: 2090.

77. Yang W, Pitera J, Swope W, Gruebele M. Heterogeneous folding of the trpzip hairpin: full atom simulation and experiment. Journal of Molecular Biology 2004; 336: 241-251.

78. Simmerling C, Strockbine B, Roitberg A. All-atom structure prediction and folding simulations of a stable protein. Journal of the American Chemical Society 2002; 124: 11258-11259.

79. Chowdhury S, Lee M, Xiong G, Duan Y. Ab initio folding simulation of the Trp-cage mini-protein approaches NMR resolution. Journal of Molecular Biology 2003; 327: 711-717.

80. Simmerling C, Lee M, Ortiz A, Kolinski A, Skolnick J, Kollman P. Combining MONSSTER and LES/PME to predict protein structure from amino acid sequence: application to the small protein CMTI-1. Proteins: Structure, Function, and Genetics 1997; 2: 6.

81. Zhou Y, Linhananta A. Role of hydrophilic and hydrophobic contacts in folding of the second-hairpin fragment of protein G: molecular dynamics simulation studies of an all-atom model. Proteins: Structure, Function, and Genetics 2002; 47: 154-162.

82. Rafi S, Novichenok P, Kolappan S, et al. Structure of acyl carrier protein bound to FabI, the FASII enoyl reductase from Escherichia coli. Journal of Biological Chemistry 2006; 281: 39285.

83. Duan Y, Kollman P. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 1998; 282: 740.

84. Zagrovic B, Snow C, Shirts M, Pande V. Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing. Journal of Molecular Biology 2002; 323: 927-937.

85. Hornak V, Okur A, Rizzo R, Simmerling C. HIV-1 protease flaps spontaneously close to the correct structure in simulations following manual placement of an inhibitor into the open state. Journal of the American Chemical Society 2006; 128: 2812.

86. Lee M, Duan Y. Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent model. Proteins: Structure, Function, and Bioinformatics 2004; 55: 620-634.

87. de Bakker P, DePristo M, Burke D, Blundell T. Ab initio construction of polypeptide fragments: accuracy of loop decoy discrimination by an all-atom statistical potential and the AMBER force field with the Generalized Born solvation model. Proteins: Structure, Function, and Genetics 2003; 51.

88. Jacobson M, Pincus D, Rapp C, et al. A hierarchical approach to all-atom protein loop prediction. Proteins: Structure, Function, and Bioinformatics 2004; 55.

89. Monnigmann M, Floudas C. Protein loop structure prediction with flexible stem geometries. Proteins: Structure, Function, and Bioinformatics 2005; 61.

90. Nishikawa K, Scheraga HA. Geometrical criteria for formation of coiled-coil structures of polypeptide chains. Macromolecules 1976; 9: 395-407.

91. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. Science 1991; 252: 1162-1164.

92. Tropsha A, Bowen JP, Brown FK, Kizer JS. Do interhelical side chain-backbone hydrogen bonds participate in formation of leucine zipper coiled coils? Proceedings of the National Academy of Science 1991; 88: 9488-9492.

93. Cregut D, Liautard JP, Heitz F, Chiche L. Molecular modeling of coiled-coil alpha-tropomyosin: analysis of staggered and in register helix-helix interactions. Protein Engineering 1993; 6: 51-58.

94. Zhang L, Hermans J. Molecular dynamics study of structure and stability of a model coiled coil. Proteins 1993; 16: 384-392.

95. Kerr ID, Sankararamakrishnan R, Smart OS, Sansom MS. Parallel helix bundles and ion channels: molecular modeling via simulated annealing and restrained molecular dynamics. Biophysical Journal 1994; 67: 1501-1515.

96. Rozzelle JE, Jr., Tropsha A, Erickson BW. Rational design of a three-heptad coiled-coil protein and comparison by molecular dynamics simulation with the GCN4 coiled coil: presence of interior three-center hydrogen bonds. Protein Science 1994; 3: 345-355.

97. Vieth M, Kolinski A, Brooks CL, 3rd, Skolnick J. Prediction of the folding pathways and structure of the GCN4 leucine zipper. Journal of Molecular Biology 1994; 237: 361-367.

98. Berger B, Wilson DB, Wolf E, Tonchev T, Milla M, Kim PS. Predicting coiled coils by use of pairwise residue correlations. Proceedings of the National Academy of Science 1995; 92: 8259-8263.

99. Harbury PB, Tidor B, Kim PS. Repacking protein cores with backbone freedom: structure prediction for coiled coils. Proceedings of the National Academy of Science 1995; 92: 8408-8412.

100. Offer G, Sessions R. Computer modelling of the alpha-helical coiled coil: packing of side-chains in the inner core. Journal of Molecular Biology 1995; 249: 967-987.

101. Sankararamakrishnan R, Sansom MS. Modelling packing interactions in parallel helix bundles: pentameric bundles of nicotinic receptor M2 helices. Biochim Biophys Acta 1995; 1239: 122-132.

102. Vieth M, Kolinski A, Brooks CL, 3rd, Skolnick J. Prediction of quaternary structure of coiled coils: application to mutants of the GCN4 leucine zipper. Journal of Molecular Biology 1995; 251: 448-467.

103. Woolfson DN, Alber T. Predicting oligomerization states of coiled coils. Protein Science 1995; 4: 1596-1607.

104. Hirst JD, Vieth M, Skolnick J, Brooks CL, 3rd. Predicting leucine zipper structures from sequence. Protein Engineering 1996; 9: 657-662.

105. Vieth M, Kolinski A, Skolnick J. Method for predicting the state of association of discretized protein models: application to leucine zippers. Biochemistry 1996; 35: 955-967.

106. Garnier N, Genest D, Duneau JP, Genest M. Molecular modeling of c-erbB2 receptor dimerization: coiled-coil structure of wild and oncogenic transmembrane domains--stabilization by interhelical hydrogen bonds in the oncogenic form. Biopolymers 1997; 42: 157-168.

107. Wolf E, Kim PS, Berger B. MultiCoil: a program for predicting two- and three-stranded coiled coils. Protein Science 1997; 6: 1179-1189.

108. Lupas A. Predicting coiled-coil regions in proteins. Current Opinion in Structural Biology 1997; 7: 388-393.

109. Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS. High-resolution protein design with backbone freedom. Science 1998; 282: 1462-1467.

110. Zhong Q, Husslein T, Moore PB, Newns DM, Pattnaik P, Klein ML. The M2 channel of influenza A virus: a molecular dynamics study. FEBS Letters 1998; 434: 265-271.

111. Zhong Q, Jiang Q, Moore PB, Newns DM, Klein ML. Molecular dynamics simulation of a synthetic ion channel. Biophysical Journal 1998; 74: 3-10.

112. Kukol A, Arkin IT. vpu transmembrane peptide structure obtained by site-specific fourier transform infrared dichroism and global molecular dynamics searching. Biophysical Journal 1999; 77: 1594-1601.

113. Yang PK, Tzou WS, Hwang MJ. Restraint-driven formation of alpha-helical coiled coils in molecular dynamics simulations. Biopolymers 1999; 50: 667-677.

114. Kukol A, Arkin IT. Structure of the influenza C virus CM2 protein transmembrane domain obtained by site-specific infrared dichroism and global molecular dynamics searching. Journal of Biological Chemistry 2000; 275: 4225-4229.

115. Walshaw J, Woolfson DN. Open-and-shut cases in coiled-coil assembly: alpha-sheets and alpha-cylinders. Protein Science 2001; 10: 668-673.

116. Walshaw J, Woolfson DN. Socket: a program for identifying and analysing coiled-coil motifs within protein structures. Journal of Molecular Biology 2001; 307: 1427-1450.

117. Briki F, Doucet J, Etchebest C. A procedure for refining a coiled coil protein structure using x-ray fiber diffraction and modeling. Biophysical Journal 2002; 83: 1774-1783.

118. Gorfe AA, Ferrara P, Caflisch A, Marti DN, Bosshard HR, Jelesarov I. Calculation of protein ionization equilibria with conformational sampling: pK(a) of a model leucine zipper, GCN4 and barnase. Proteins 2002; 46: 41-60.

119. Orzechoski M, Cieplak P, Piela L. Theoretical calculations of the coiled-coil stability in water in the context of its possible use as a molecular rack. Journal of Computational Chemistry 2002; 23: 106-110.

120. Ash WL, Stockner T, MacCallum JL, Tieleman DP. Computer modeling of polyleucine-based coiled coil dimers in a realistic membrane environment: insight into helix-helix interactions in membrane proteins. Biochemistry 2004; 43: 9050-9060.

121. Gruber M, Soding J, Lupas AN. Comparative analysis of coiled-coil prediction methods. Journal of Structural Biology 2006; 155: 140-145.

122. Ramos J, Lazaridis T. Energetic determinants of oligomeric state specificity in coiled coils. Journal of the American Chemical Society 2006; 128: 15499-15510.

123. Nilges M, Brunger AT. Automated modeling of coiled coils: Application to the GCN4 dimerization domain. Protein Engineering 1991; 4: 649-659.

124. Nilges M, Brunger AT. Successful prediction of the coiled coil geometry of the GCN4 leucine zipper domain by simulated annealing: comparison to the X-ray structure. Proteins 1993; 15: 133-146.

125. Madhusoodanan M, Lazaridis T. Investigation of pathways for the low-pH conformational transition in influenza hemagglutinin. Biophysical Journal 2003; 84: 1926-1939.

126. Pineiro A, Villa A, Vagt T, Koksch B, Mark AE. A molecular dynamics study of the formation, stability, and oligomerization state of two designed coiled coils: possibilities and limitations. Biophysical Journal 2005; 89: 3701-3713.

127. Yadav MK, Leman LJ, Price DJ, Brooks CL, 3rd, Stout CD, Ghadiri MR. Coiled coils at the edge of configurational heterogeneity. Structural analyses of parallel and antiparallel homotetrameric coiled coils reveal configurational sensitivity to a single solvent-exposed amino acid substitution. Biochemistry 2006; 45: 4463-4473.

128. Missimer JH, Steinmetz MO, Baron R, et al. Configurational entropy elucidates the role of salt-bridge networks in protein thermostability. Protein Science 2007; 16: 1349-1359.

129. Zhong Q, Moore PB, Newns DM, Klein ML. Molecular dynamics study of the LS3 voltage-gated ion channel. FEBS Letters 1998; 427: 267-270.

130. Mottamal M, Zhang J, Lazaridis T. Energetics of the native and non-native states of the glycophorin transmembrane helix dimer. Proteins 2006; 62: 996-1009.

131. Samna Soumana O, Garnier N, Genest M. Molecular dynamics simulation approach for the prediction of transmembrane helix-helix heterodimers assembly. European Biophysics Journal 2007; 36: 1071-1082.

132. El-Kettani MA, Smith JC. Pathways for conformational change in seryl-tRNA synthetase from Thermus thermophilus. The Academy of Science of the Royal Society of Canada 1996; 319: 161-169.

133. Ding F, LaRocque JJ, Dokholyan NV. Direct observation of protein folding, aggregation, and a prion-like conformational conversion. Journal of Biological Chemistry 2005; 280: 40235-40240.

134. Wang C, Stewart R. Hybrid hydrogels assembled from synthetic polymers and coiled-coil protein domains. Nature 1999; 397: 417-420.

135. Bilgicer B, Fichera A, Kumar K. A coiled coil with a fluorous core. Journal of the American Chemical Society 2001; 123: 4393-4399.

136. Wang C, Kopecek J, Stewart R. Hybrid hydrogels cross-linked by genetically engineered coiled-coil block proteins. Biomacromolecules 2001; 2: 912-920.

137. Kohn W, Kay C, Sykes B, Hodges R. Metal ion induced folding of a de novo designed coiled-coil peptide. Journal of the American Chemical Society 1998; 120: 1124-1132.

138. Pechar M, Kopecková P, Joss L, Kopecek J. Associative diblock copolymers of poly (ethylene glycol) and coiled-coil peptides. Macromolecular Bioscience 2002; 2: 199-206.

139. Tang A, Wang C, Stewart R, Kopecek J. The coiled coils in the design of protein-based constructs: hybrid hydrogels and epitope displays. Journal of Controlled Release 2001; 72: 57-70.

140. Woolfson DN. The design of coiled-coil structures and assemblies. Advances in Protein Chemistry 2005; 70: 79-112.

141. Potekhin S, Melnik T, Popov V, et al. De novo design of fibrils made of short α-helical coiled coil peptides. Chemistry & Biology 2001; 8: 1025-1032.

142. Xu C, Kopecek J. Genetically engineered block copolymers: influence of the length and structure of the coiled-coil blocks on hydrogel self-assembly. Pharmaceutical Research 2008; 25: 674-682.

143. Tang A, Wang C, Stewart R, Kopecek J. Self-assembled peptides exposing epitopes recognizable by human lymphoma cells. Bioconjugate Chemistry 2000; 11: 363-371.

144. Naik R, Kirkpatrick S, Stone M. The thermostability of an -helical coiled-coil protein and its potential use in sensor applications. Biosensors and Bioelectronics 2001; 16: 1051-1057.

145. Kopecek J. Hydrogel biomaterials: A smart future? Biomaterials 2007; 28: 5185-5192.

146. Lindhout DA, Litowski JR, Mercier P, Hodges RS, Sykes BD. NMR solution structure of a highly stable de novo heterodimeric coiled-coil. Biopolymers 2004; 75: 367 - 375.

147. Pendley S, Yu Y, Cheatham III T. Molecular dynamics guided study of salt bridge length dependence in both fluorinated and non-fluorinated parallel dimeric coiled-coils. Proteins: Structure, Function, and Bioinformatics 2009; 74: 612.

148. Bulheller BM, Hirst JD. DichroCalc--circular and linear dichroism online. Bioinformatics 2009; 25: 539-540.

149. Berova N, Nakanishi K, Woody R. Circular dichroism: principles and applications. New York: VCH Publishers; 1994.

CHAPTER 2

MOLECULAR DYNAMICS GUIDED STUDY OF SALT BRIDGE LENGTH DEPENDENCE IN BOTH FLUORINATED AND NON-FLUORINATED PARALLEL DIMERIC COILED-COILS

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Molecular dynamics guided study of salt bridge length dependence in both fluorinated and non-fluorinated parallel dimeric coiled-coils

Scott S. Pendley,[1] Yihua B. Yu,[1,2] and Thomas E. Cheatham III[1,3,4]*

[1] Departments of Pharmaceutics and Pharmaceutical Chemistry, University of Utah, Salt Lake City, Utah 84112

[2] Departments of Pharmaceutical Sciences and Bioengineering, University of Maryland, Baltimore, Maryland 21201

[3] Department of Medicinal Chemistry, University of Utah, Salt Lake City, Utah 84112

[4] Department of Bioengineering, University of Utah, Salt Lake City, Utah 84112

ABSTRACT

The α-helical coiled-coil is one of the most common oligomerization motifs found in both native and engineered proteins. To better understand the stability and dynamics of the coiled-coil motifs, including those modified by fluorination, several fluorinated and nonfluorinated parallel dimeric coiled-coil protein structures were designed and modeled. We also attempt to investigate how changing the length and geometry of the important stabilizing salt bridges influences the coiled-coil protein structure. Molecular dynamics (MD) and free energy simulations with AMBER used a particle mesh Ewald treatment of the electrostatics in explicit TIP3P solvent with balanced force field treatments. Preliminary studies with legacy force fields (ff94, ff96, and ff99) show a profound instability of the coiled-coil structures in short MD simulation. Significantly, better behavior is evident with the more balanced ff99SB and ff03 protein force fields. Overall, the results suggest that the coiled-coil structures can readily accommodate the larger acidic arginine or S-2,7-diaminoheptanedoic acid mutants in the salt bridge, whereas substitution of the smaller L-ornithine residue leads to rapid disruption of the coiled-coil structure on the MD simulation time scale. This structural distortion of the secondary structure allows both the formation of large hydration pockets proximal to the charged groups and within the hydrophobic core. Moreover, the increased structural fluctuations and movement lead to a decrease in the water occupancy lifetimes in the hydration pockets. In contrast, analysis of the hydration in the stable dimeric coiled-coils shows high occupancy water sites along the backbone residues with no water occupancy in the hydrophobic core, although transitory water interactions with the salt bridge residues are evident. The simulations of the fluorinated coiled-coils suggest that in some cases fluorination electrostatically stabilizes the intermolecular coiled-coil salt bridges. Structural analyses also reveal different side chain rotamer preferences for leucine when compared with 5,5,5,5′,5′,5′-hexafluoroleucine mutants. These observed differences in the side chain rotamer populations suggest differential changes in the side chain conformational entropy upon coiled-coil formation when the protein is fluorinated. The free energy of hydration of the isolated 5,5,5,5′,5′,5′-hexafluoroleucine amino acid is calculated to be 1.1 kcal/mol less stable than leucine; this hydrophobic penalty in the monomer may provide a driving force for coiled-coil dimer formation. Estimation of the ellipticity at 222 nm from a series of snapshots from the MD simulations with DicroCalc shows distinct increases in the ellipticity when the coiled-coil is fluorinated, which suggests that the helicity in the folded coiled-coils is greater when fluorinated.

Proteins 2009; 74:612–629.

Key words: computational chemistry; free energy of hydration; 5,5,5,5′,5′,5′-hexafluoroleucine; thermodynamic integration; rotamers.

INTRODUCTION

Coiled-coils are a very common and well studied protein motif composed of two or more α-helices wound around each other.[1–6] They can form both homo- and

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: NIH; Grant numbers: R01-GM079383-01, R21-EB002880-02; Grant sponsor: NSF LRAC; Grant number: MCA01S027P; Grant sponsor: University of Utah Center for High Performance Computing (NIH 1S10RR17214-01).
*Correspondence to: Thomas Cheatham, Department of Pharmaceutics and Pharmaceutical Chemistry, University of Utah, 2000 East, 30 South, Skaggs 201, Salt Lake City, Utah 84112. E-mail: tec3@utah.edu
Received 6 February 2008; Revised 8 May 2008; Accepted 23 May 2008
Published online 14 August 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22177

hetero- multimers with two to seven helices[7] and can have helix orientations either parallel or antiparallel. Their regular structure and ease of synthesis has led to their use as the basis for stimuli-sensitive hydrogels, as epitope display scaffolds, and as components of biosensors.[8–15] In cells, coiled-coil motifs can act as an oligomerization domain, and coiled-coils have been found in a wide variety of proteins including transcription factors,[16] motor proteins,[17] structural filaments, chaperone proteins,[18] tRNA synthetases,[19] SNARE complexes,[20] and cell or viral fusion proteins.[21] Functionally, coiled-coil motifs can act as levers, scaffolds, moving arms, and potentially as springs or nano-motors.[22] Potential pharmaceutical applications of coiled-coils include their use as multifunctional delivery, targeting, and imaging agents[13,23]). Imaging takes advantage of fluorination and ^{19}F NMR. Fluorinated amino acids can be easily incorporated into proteins both by solid state[24–26] and *in vivo*[27–29] protein synthesis, and fluorination leads to increased hydrophobicity,[30] increased stability,[31,32] and decreased drug metabolism.[33] Incorporation of fluorinated amino acids, such as by replacement of leucine with 5,5,5,5′,5′,5′-hexafluoroleucine (hFLeu) in the hydrophobic core of coiled-coils, leads to enhanced stability and resistance to both thermal and chemical denaturation.[28,29,31,34–38] Although fluorine incorporation in the hydrophobic core tends to increase protein stability, the influence of fluorine on the protein structure is not without controversy. In α-helices, the incorporation of fluorine has been reported to cause structural changes, structural distortions, lower helical propensity, and decreased ellipticity as measured in CD experiments.[39] This contrasts with other studies where incorporating fluorine leads to increased stability and secondary structure.[38] Measurements of the energetic contribution of fluorine to thermostability in α-helices vary; however, when fluorine is incorporated into the hydrophobic core it favorably contributes ~0.5–1.2 kcal/mol-residue to the stability.[28,36,39,40] In addition to differences in the free energetics, there are concerns related to the larger steric bulk of fluorinated compounds. Tirrell and coworkers incorporated fluorinated residues in α-helices by the introduction of 5,5,5-trifluoroleucine into the p1 region of the leucine zipper GCN4.[34] They proposed that these fluorinated residues were isomorphic to their nonfluorinated versions.[34,41] Although, the fluorinated amino acids maintain a shape very similar to their nonfluorinated counterparts, because the steric bulk of a trifluoromethyl group is closer in size to an isopropyl group,[42] and since hFLeu is ~37 Å^3 larger than leucine (consistent with the difference in van der Waals radii for hydrogen of 1.20 Å and 1.47 Å for fluorine),[37,43] fluorination leads to alterations in the core structure.[38]

The primary sequences of coiled-coil proteins are characterized by a highly recognizable heptad repeat denoted as **abcdefg** (see Fig. 1). Positions **a** and **d** are typically occupied by the hydrophobic residues whereas the positions **e** and **g** are typically the charged residues that form an intermolecular salt bridge.[44] As each turn of the α-helix results in the progression of 3.6 residues, each heptad repeat progresses slightly less than two full twists around the helix.[45] The coiled-coil structure is formed when the component helices come together, bury their hydrophobic regions, and form further stabilizing interactions such as salt bridges; the coiled-coil formation leads to a net-stabilization of each individual helix.[46] As the hydrophobic interface slowly twists around the helix, association of multiple helices at the hydrophobic domain results in a super-coil with classically defined inter-digitations of hydrophobic side chains between neighboring helices.[47]

As one of the simplest and most common tertiary structural motifs, coiled-coils are widely studied scaffolds for protein engineering and design.[5,44,48–59] Coiled-coils are also a commonly used model systems in protein folding and stability studies.[60–74] Although predicting the secondary and tertiary structure of coiled-coil proteins is relatively straightforward as compared to other protein motifs, 3D structure prediction is still difficult as coiled-coil proteins display a rich tapestry of structure and motion. Specifically small changes in sequence can switch the oligomerization state, orientation, or alter the coiled-coil structure and dynamics.[14,55,75–84] The structure and dynamics are also influenced by extrinsic factors, such as concentration, temperature, pH, ligand binding, and dielectric.[8,62,85–93] Even within well-behaved dimeric structures, distinct conformational substates in slow exchange may be populated.[73,94]

MOLECULAR MODELING AND MOLECULAR DYNAMICS SIMULATIONS OF COILED-COILS

The computational studies of coiled-coils to-date include molecular modeling, model building, and molecular dynamics (MD) simulation approaches aimed at better understanding the physical forces and chemical interactions that define the structure and dynamics of coiled-coil proteins. Initial modeling work focused on extending a set of principles and general formulas developed by Crick to describe ideal parallel coiled-coil dimer structure with knobs into holes packing of the hydrophobic core.[45] Later, more detailed modeling and MD simulation studies were applied to reproduce experimental structure and dynamics, coiled-coil oligomerization states, and ultimately to estimate relative energies of coiled-coil formation. The initial studies used rigid backbones and side chain packing in the hydrophobic domain of parallel coiled-coil.[95] As more structures emerged and the regularity of the coiled-coil structure was confirmed, a wide arsenal of different programs and methods to build coiled-coil structures emerged. These each use

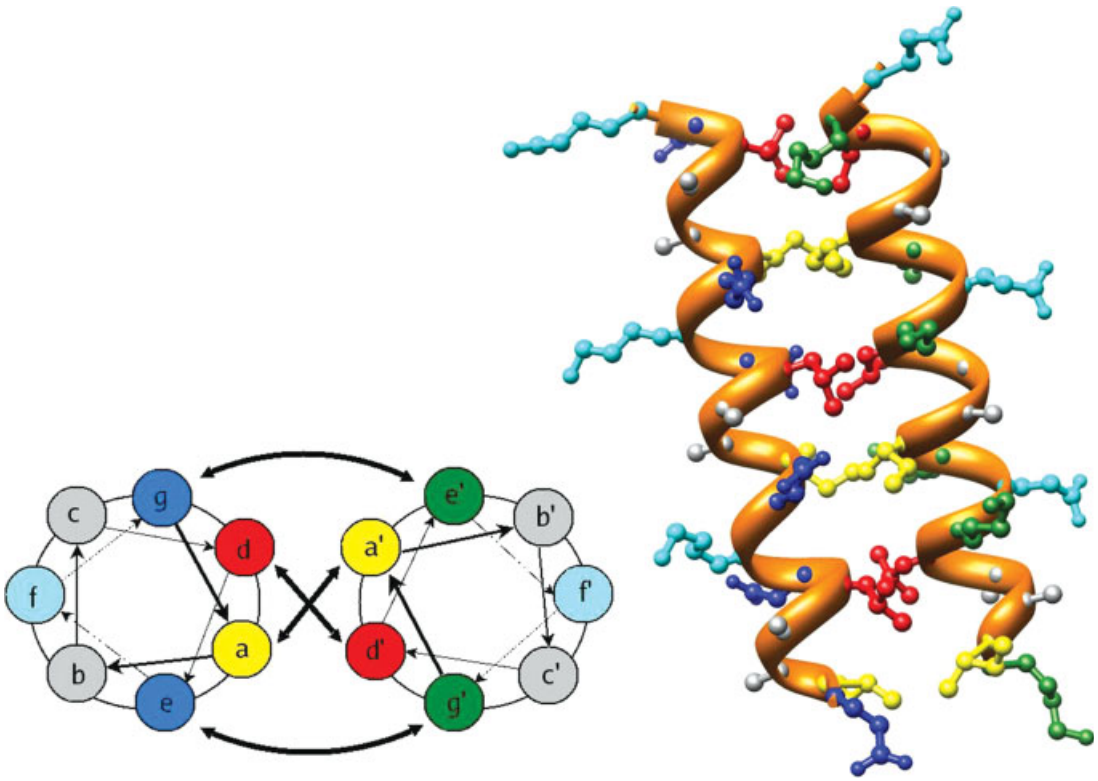


Figure 1

A heptad helical wheel and molecular graphics representation of a parallel three heptad repeat coiled-coil showing the heavy atoms and a ribbon representation of backbone secondary structure. The helical wheel shows a top-view looking down the α-helix of the two interacting helices that shows where each residue side chain is approximately located. The arrows within each circle show the residue connectivity. The crossed arrows at the center denote the interactions of the hydrophobic residues **a** and **a′** and **d** and **d′**. The arcing arrows at the top and bottom denote the classical salt bridge interactions between **g** and **e′** and **e** and **g′**. (The helical wheel representation shown is adapted from Ref. 12). The ribbon structure of two helices interacting as a coiled-coil is also shown including side-chain heavy atoms. The coiled-coil represented in the molecular graphics is the parallel three heptad repeat IAAL-E3/K3 dimer (PDB ID: 1U0I).[48]

various combinations of distinct means to pack the hydrophobic cores, explore the flexibility of side chains, optimize the electrostatics, and/or sample different orientations and oligomeric states.[60,87,96–125]

Highlights of the theoretical approaches towards the understanding of the coiled-coil structure and dynamics include:

- Early simulation approaches using MD simulation with explicit representations did remarkably well in reproducing coiled-coil structure. This includes unrestrained MD of the yeast transcriptional activator GCN4 starting from Cα atoms followed by automatic building with simulated annealing to produce a structure of 1.25 Å RMSd (backbone) from the crystal structure.[126,127] Additionally, Monte Carlo folding simulations followed by the atomistic MD on the GCN4 dimer predicted structure with an accuracy of 0.81 Å (backbone); these calculations facilitated quaternary structure predictions.[60,106,109] Other work performed comparative MD simulations on different coiled-coil protein sequences to qualitatively understand stability.[101] These early successes are notable in comparison with our recent work that shows considerable force field dependence in the simulation of the coiled-coils studied in this work (see the Methods and more data in the supplementary material).
- Investigations of the influence of a membrane environment on the coiled-coil structure and dynamics,[114,115,123,128,129] including a MD study showing formation of a well-defined coiled-coil structure in the presence of an external electrostatic field.[86]
- Investigations of coiled-coil proteins using various free energetic approaches to the better understand of helical propensity,[99] stability,[122] and the influence of sequence and the environment on oligomerization.[84,130,131]
- Detailed validation and prediction aimed at understanding the preferential oligomerization state of different coiled-coils.[88,113,125]
- The use of targeted or force-induced MD simulations to investigate alterations in coiled-coil structure and

dynamics, including calculation of a conformational exchange pathway connecting two different crystal structures of the anti-parallel coiled-coil of seryl tRNA synthetase,[132] transitions from the native to the putative fusogenic conformations of influenza hemagglutinin at low pH,[88] transitions among different GCN4 structures,[92] and investigation of putative conformational transitions in engineered prion peptides.[89]

Together, these results provide strong validation of the modern force fields and MD simulation protocols and their ability to model complex protein tertiary behavior and structure. Yet, very few simulations have investigated the effect of either fluorination or changes in the salt bridges on coiled-coil protein structure and dynamics. The only MD of fluorinated coiled-coils published to-date includes the studies by Tirrell's group where a combination of experiment and MD simulation with an implicit solvent model were applied to understand the effect of 5,5,5-trifluoroleucine substitution in the GCN4 coiled-coil.[28]

In this article, we present the results of structural and energetic studies of parallel dimeric coiled-coils, with and without fluorination, pursued through the use of biomolecular simulation methods. The aim of this study is to decompose contributions to the structural and energetics differences between the fluorinated and nonfluorinated coiled-coils and to give context to the variations seen in the early fluorination studies. To decompose the contributions of these differences, we have designed point mutations in the IAAL-E3/K3 coiled-coil[133] to both fluorinate the peptide and to vary the length and geometry of the salt bridge domain. As seen in Table I, the lysine residues in positions **e** and **g** of the basic monomer (labeled Monomer B in our tables) have been mutated to ornithine, arginine, and S-2,7-diaminoheptanoic acid. Fluorination mutations have also been designed at position **d** of both the monomers to substitute 5,5,5,5′,5′,5′-hexafluoroleucine (hFLeu) for leucine. Analysis of the MD simulation runs to investigate the influence of fluorination and salt bridge deformation were pursued through calculations of circular dichroism ellipicity values to characterize the helicity, studies of rotamer preferences in the hydrophobic domain, and calculations of the salt bridge length. The results of the energetic calculations to measure the free energy of hydration and the pair wise electrostatic contributions are also discussed.

Table I
Summary of the Coiled Coil Models

Position in heptad	gabcdefgabcdefgabcdef
IAAL-E3/K3	
Monomer A	EIAALEKEIAALEKEIAALEK
Monomer B	**K**IAAL**K**E**K**IAAL**K**E**K**IAAL**K**E
IAA(hFLeu)-E3/K3	
Monomer A	EIAAĹEKEIAAĹEKEIAAĹEK
Monomer B	**K**IAAĹ**K**E**K**IAAĹ**K**E**K**IAAĹ**K**E
IAAL-E3/O3	
Monomer A	EIAALEKEIAALEKEIAALEK
Monomer B	**O**IAAL**O**E**O**IAAL**O**E**O**IAAL**O**E
IAA(hFLeu)-E3/O3	
Monomer A	EIAAĹEKEIAAĹEKEIAAĹEK
Monomer B	**O**IAAĹ**O**E**O**IAAĹ**O**E**O**IAAĹ**O**E
IAAL-E3/R3	
Monomer A	EIAALEKEIAALEKEIAALEK
Monomer B	**R**IAAL**R**E**R**IAAL**R**E**R**IAAL**R**E
IAA(hFLeu)-E3/R3	
Monomer A	EIAAĹEKEIAAĹEKEIAAĹEK
Monomer B	**R**IAAĹ**R**E**R**IAAĹ**R**E**R**IAAĹ**R**E
IAAL-E3/Ĥ3	
Monomer A	EIAALEKEIAALEKEIAALEK
Monomer B	**Ĥ**IAAL**Ĥ**E**Ĥ**IAAL**Ĥ**E**Ĥ**IAAL**Ĥ**E
IAA(hFLeu)-E3/Ĥ3	
Monomer A	EIAAĹEKEIAAĹEKEIAAĹEK
Monomer B	**Ĥ**IAAĹ**Ĥ**E**Ĥ**IAAĹ**Ĥ**E**Ĥ**IAAĹ**Ĥ**E

O = ornithine (Orn)
Ĺ = 5,5,5,5′,5′,5′-hexafluoroleucine (hFLeu)
Ĥ = (S)-2,7-diaminoheptanoic acid (DAH)

The sequence of each coiled-coil model is listed using single letter amino acid codes. The control sequence is listed at the top and changes in sequence are bolded and colored red. The monomer A is identical in each coiled-coil.

METHODS

Starting geometries

The initial structures were derived from the first NMR structure model of the IAAL-E3/K3 coiled-coil dimer (PDB ID: 1U0I).[134] Point mutations at positions e and g of the IAAL-E3/K3 basic monomer were made using Swiss PDB viewer[135] and the AMBER LEaP program[136] to vary salt bridge length and geometry.

Force fields

Simulations of 3- and 4- heptad repeat coiled-coils were validated and completed using the AMBER ff99SB modification[137] of the Cornell *et al.* force field (ff94).[138] Although not complete, validation of the force field was facilitated by visualization of the model structures and calculations of the RMSd values to the relevant experimental structures. Unless otherwise mentioned, the model structures were built as straight coordinate averages of best-fit structures after equilibration; the models were created from 2 ns windows from the trajectory (which was stored at 1 ps intervals), using ptraj. Earlier attempts to simulate these and related coiled-coil protein structures using the AMBER ff94,[138] ff96,[139] and ff99[140] force fields and equivalent simulation protocols resulted in the significant structural distortion. The coiled-coils became bent and unraveled in the short ns-scale simulations (see the supplementary material). The unwinding suggests instability of the helices comprising the coiled-coil; this was somewhat surprising as the ff94 and ff99 are known to overstabilize the α-helices.[141,142] Consistent with intuition and experiment, simulations of larger coiled-coil protein

structures with the ff96 and ff99 force fields showed less of a dependence on the force field parameters and comparatively less movement away from the experimental structure; however, the absence of movement away from a starting structure in the ns-scale simulation does not necessarily imply the greater stability. In fact, well-known force field biases can be hidden in the short MD simulations[143] or in simulations of the larger protein assemblies. As the best agreement with the experimental structures was observed with the ff99SB force field, this force field was adopted in the present studies.

Molecular dynamics simulations

All simulations of the parallel coiled-coil protein dimers and monomers were performed with the AMBER suite of programs[136,144] using the ff99SB force field.[141] Periodic boundary conditions were applied using the particle mesh Ewald method with a less than 1 Å charge grid and cubic B-spline interpolation.[145] Proteins were solvated by surrounding the protein with at least a 12 Å water layer in each direction within a truncated octahedron using TIP3P waters.[146] This amounts to on the order of 4500–6000 waters. After the equilibration protocol was simulated, MD simulations without restraints were continued until convergence—as determined by plateaus in RMSd plots—were observed. This typically required MD sampling times on the order of ~15–65 ns for each coiled-coil, with the longer convergence times necessary for the structures where larger amino acid mutations were made, such as the replacement of lysine by arginine in the salt bridge. Note that these simulations are significantly longer than the <2 ns MD sampled in simulations with ff94, ff96, and ff99 where significant structural distortions were evident (see supplementary material). All of the MD simulations were performed with a 2 fs time step and a direct space nonbonded cutoff of 10 Å with the pair list of atomic interactions built out to 11 Å and heuristic update of the pair list triggered when any atom moved more than 0.5 Å since the previous update. During MD, bond lengths involving hydrogen atoms were constrained with SHAKE[147,148] with a geometric tolerance for the constraint of 0.00001 Å during the coordinate resetting. Initial minimization was followed by heating to 300 K at constant volume over a period of 10 ps using harmonic restraints of 2 kcal/(mol Å) on the protein atoms. During production runs, the center of mass translational motion of the entire system was removed after the initial velocity assignments and subsequently every 5000 MD steps. Constant temperature was maintained using the weak coupling algorithm and heat bath coupling with a 2 ps time constant.[149] Pressure (1 atm) was maintained using isotropic position scaling with the Berendsen weak coupling algorithm with a 1.0 ps pressure relaxation time.[149]

Residue parameterization

The development of new force field parameters for the residues ornithine, 5,5,5,5′,5′,5′-hexafluoroleucine (hFLeu), and S-2,7-diaminoheptanoic acid [Fig. 2(d)] was completed consistent with the original Cornell *et al.* force field[138] and recent ff99SB modifications.[141] The ff99SB modifications to the torsional potential of the peptide backbones were necessary to improve the conformational ensemble sampled by the glycine residues and also, through modifications to the ϕ/ψ torsional potentials, to provide a better energetic balance between the α-helix and β-sheet peptide geometries. The existing force field parameters, along with RESP derived charges at a consistent level of theory, have proven effective not only for modeling both peptides and nucleic acids,[150,151] but

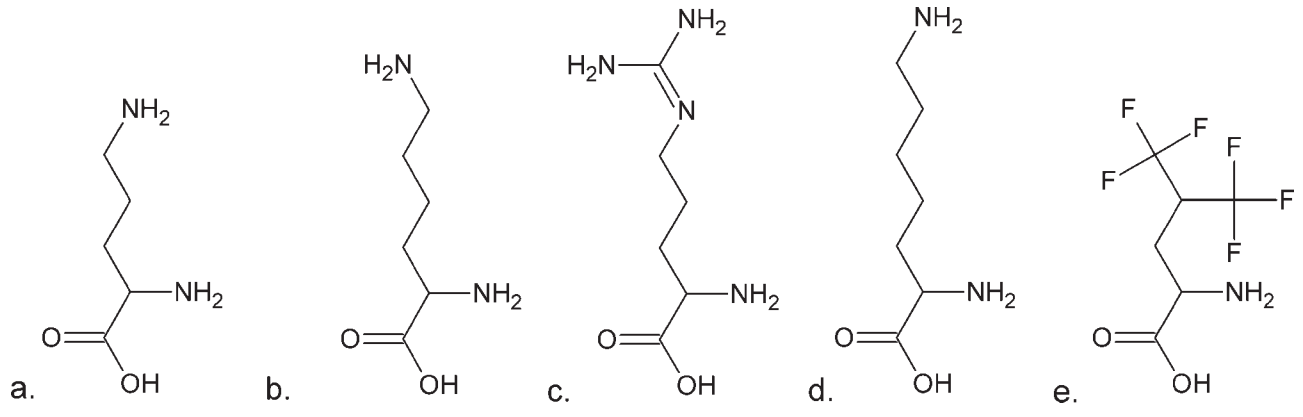


Figure 2

Shown are the structures of the various amino acid substitutions used at the **e** and **g** positions of the basic monomer, monomer B. (**a**) *L*-ornithine, (**b**) *L*-lysine, (**c**) *L*-arginine and (**d**) S-2,7-diaminoheptanoic acid. Substitutions at the **d** position of both monomers replace L-leucine with (**e**) 5,5,5,5′,5′,5′-hexafluoro-leucine.

Table II
Standard Dihedral Angle Definitions for Leucine and Isoleucine[156]

Residue	Angle	Rotamer	Angle definitions
Ile	χ1	r1	N-Cα-Cβ-Cγ1
Ile	χ2	r2	Cα-Cβ-Cγ1-Cδ
Leu	χ1	r1	N-Cα-Cβ-Cγ
Leu	χ2	r2	Cα-Cβ-Cγ-Cδ1

also for small organic molecules.[152,153] To build the new residues, atom types were chosen consistent with the Cornell *et al.* and ff99SB force fields. Torsional and angle parameters were assigned consistent with the existing parameters from the ff99SB force field. No new angle or dihedral force field parameters were required beyond those already available in ff99SB. Although these residues are new, the fluorine parameters in the ff99SB force field have been previously validated, albeit with a less accurate charge model, through estimations of fluorophilicity for a series of small molecules[154] and simulation of aggregation behavior of fluoroalkanes when compared with alkanes.[155,156] More rigorous tests of solvation free energies for small fluorine containing molecules using the AMBER fluorine parameters and a RESP treatment of the charges suggest that the applied model is appropriate for the fluorine containing compounds.[152,153]

Di-peptide analogues of the amino acids were created by capping the N- and C- terminal ends of the amino acids with acetyl (ACE) and N-methyl (NME) groups, respectively. These structures were then optimized using the Gaussian 98 software[157] at the HF/6-31+G* level. An SCF convergence criterion of 10^{-8} with tight optimization was used to ensure a fully minimized structure. This minimized structure was then used to calculate a molecular electrostatic potential (MEP) on a three dimensional grid using the GAMESS quantum mechanics software package[158] (again at the HF/6-31+G* level). Six distinct orientations of the structure were calculated and exported to the AMBER RESP program,[159] which was used to fit atom centered RESP charges to the MEP. Charge constraints were placed on the capping groups of the di-peptides such that the sum of charges of the ACE and NME groups was neutral. The parameterization was greatly facilitated by the RED II program, which provides an automated method to create the MEP and fit the RESP charges.[160]

AMBER off libraries containing these parameters have been deposited in the AMBER Parameter Database maintained by Richard Bryce[161] and are summarized in the supplementary material.

Molecular dynamics trajectory analysis

Rotamer preferences and salt bridge distances were calculated using the AMBER ptraj module.[162] Rotamer definitions are defined in Tables II and III consistent with Dunbrack and Cohen's work.[163] Rotamers were calculated over a 10–12 ns window post equilibration of the trajectories with sampling every 10 ps using the dihedral function of ptraj. The dihedral measurements for the leucine and isoleucine rotamers were taken from the d and a position, respectively, of the middle heptad of both the monomers. Salt bridge distances were calculated using atoms as described in Table IX. Distances were also calculated over a 10–12 ns window, sampling at 10 ps intervals, using the distance function of ptraj.

Hydration analyses of the coiled-coil structures were performed using the ptraj H-bond and grid utilities. A maximum cutoff angle of 120.0° and cutoff length of 3.4 Å were used in the hydrogen bond definitions. In the fluorinated coiled-coil dimer models, the carbon bonded fluorine was considered a hydrogen bond acceptor (electron donor). Measured hydrogen bonds and calculated water densities used in the hydration calculations were taken during the last 1 ns (1000 frames) of MD simulation. Hydrations sites were determined using solvent distributions calculated by binning atom positions from RMS coordinate fitting over all protein atoms at 1 ps intervals into $(0.5\ \text{Å})^3$ grids.[164] These grids were contoured using the volume visualization module of UCSF Chimera.[165]

Thermodynamic integration free energy calculations

Calculations of the relative free energy of hydration were completed using thermodynamic integration.[166] The use and accuracy of the thermodynamic integration to calculate the free energy of hydration of the amino acid residues has been shown to possibly be as accurate as sub kcal/mol measurements.[167] On the basis of a thermodynamic cycle, leucine ($\lambda = 1$) was perturbed in both gas (vacuum) and aqueous environments to 5,5,5,5′,5′,5′-hexafluoroleucine ($\lambda = 0$). Relative free energy of hydration values were determined by subtracting the gas phase perturbation measurements from the aqueous phase measurements. Aqueous simulations were performed using a particle mesh Ewald treatment of electrostatics, as previously described. These simulations were performed using the thermodynamic integration module in AMBER.[162] Residues were modeled as a di-peptide, using the original residue capped with an acetyl group (ACE) on the N-terminus and an *N*-methyl group (NME) on the C-terminus. Minimization and equilibration of di-peptides followed the procedures as described

Table III
Angle Limits and Conformation Definitions for Leucine and Isoleucine[156]

Conformation	X range
gauche+, g+	0–120 degrees
trans, t	120–240 degrees
gauche−, g−	−120–0 degrees

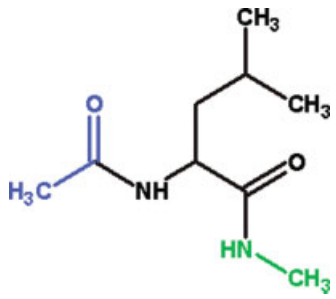


Figure 3

A representative capped amino acid dipeptide analog used in charge derivation and thermodynamic integration experiments. Acetyl (ACE) groups in blue are covalently bonded to the N-terminus of the amino acid, *N*-methyl (NME) groups in green are covalently bonded to the C-terminus of the amino acid. ACE-Leucine-NME is shown here.

previously. The structures were allowed to relax during 6 ns of MD simulation followed by the thermodynamic integration sampling. Accurate and sufficient sampling was ensured by calculating the perturbation energy by the reverse pathway and by observation of $\Delta V/\Delta\lambda$ measurement, as the simulation progressed. For our calculations, we used twelve sampling points of λ, based on Gaussian quadrature (0.00922, 0.04794, 0.11505, 0.20634, 0.31608, 0.43738, 0.56262, 0.68392, 0.79366, 0.88495, 0.95206, and 0.99078), sampled over 3 ns.

Free energy decomposition

In some studies presented below, the interactions of the specific residues and their relative contributions to the free energy of binding were studied by performing component analysis. As described elsewhere,[168] each residue was split into its component atoms. Internal energies were calculated if all atoms contributed to the bond, angle, or torsion angle energy terms. The van der Waals interactions were calculated as one half of the pair wise energy calculations for atoms composing a single residue as well as the calculations of inter-residue contacts. The electrostatic potential of the pair-wise interactions was calculated using the General Born (GB) equation, to model a given charge distribution for a solute embedded in a uniform (high) dielectric solvent. A dielectric constant of 80.0 was applied for GB based electrostatic potential calculations.

Circular dichroism calculations

Quantification and comparison of structural helical content was measured by calculating mean residue ellipticities representing the CD spectra of 20 individual structures spanning the final 2 ns of simulation of each coiled-coil dimer using the DichroCalc program.[169] A Gaussian curve type was assumed with a bandwidth at half maximum of 12.5 nm and two backbone transitions.

RESULTS

Observation of the simulation results

With the exception of the "short" salt bridged, or ornithine substituted, (E3/O3) coiled-coil models, all of the MD simulations with the ff99SB force field maintained the characteristic parallel dimeric coiled-coil structure over the course of 15–65 ns of MD simulation. In Figure 4, the root mean square deviations (RMSD) for the simulation of the IAAL-E3/K3, IAA(hFLeu)-E3/K3, IAAL-E3/O3, and IAA(hFLeu)-E3/O3 coiled-coils relative to their initial starting structures are shown. Comparison of the final IAAL-E3/K3 structure to published NMR structures[134] showed a RMSD of 0.84 Å (using the backbone atoms and omitting the terminal three residues). Whereas the standard-length salt bridged E3/K3 coiled-coils are stable on the MD simulation time scale, the shorter salt bridged IAAL-E3/O3 and IAA(hFLeu)-E3/O3 coiled-coils both undergo rapid loss of secondary structure during the MD simulation. With the IAAL-E3/O3 coiled-coil, this loss of secondary structure occurs between 10 and 20 ns of simulation time and once the structure moves away, it never comes back (see Fig. 4). Initially, the unraveling of the secondary structure in the terminal regions of both the helices precedes exposure to the solvent of the interior hydrophobic interface of the central heptad at ~10 ns. At 14 ns, the monomers begin to separate at the terminal regions, bending inward in opposing directions. Bending in the basic monomer persists until the folded regions resemble a horseshoe with a face perpendicular to the acidic monomer (15 ns). Final rearrangement occurs as the acidic monomer folds around the face of the horseshoe maximizing hydrophobic interactions. This is shown in greater detail in the supplemental material, Figure S6. Similar disruption of structure was seen in the hFLeu substituted E3/O3 coiled-coil.

The arginine (IAAL-E3/R3, IAA(hFLeu)-E3/R3) and diaminoheptanoic acid (IAAL-E3/Ĥ3, and IAA(hFLeu)-E3/Ĥ3) salt bridged coiled-coils showed similar RMSD plots compared to those observed for IAAL-E3/K3 with the exception of slower equilibration time for IAA(hFLeu)-E3/R3. We expect that the slight delay in equilibration is due to the requisite side chain reorientation and expansion in an initially constrained starting structure (all dimers were created using homology modeling to a nonfluorinated dimer). The observed remodeling of the salt bridges and coiled-coil structure are consistent with the experiment.[14,63,170]

Hydration of modeled coiled-coils

Investigation of the solvation of the folded coiled-coils was enabled by analysis and visualization of the average water occupancy in grids surrounding the protein. These grids were contoured at a level corresponding to ~2–3x bulk water density and revealed specific hydration sites

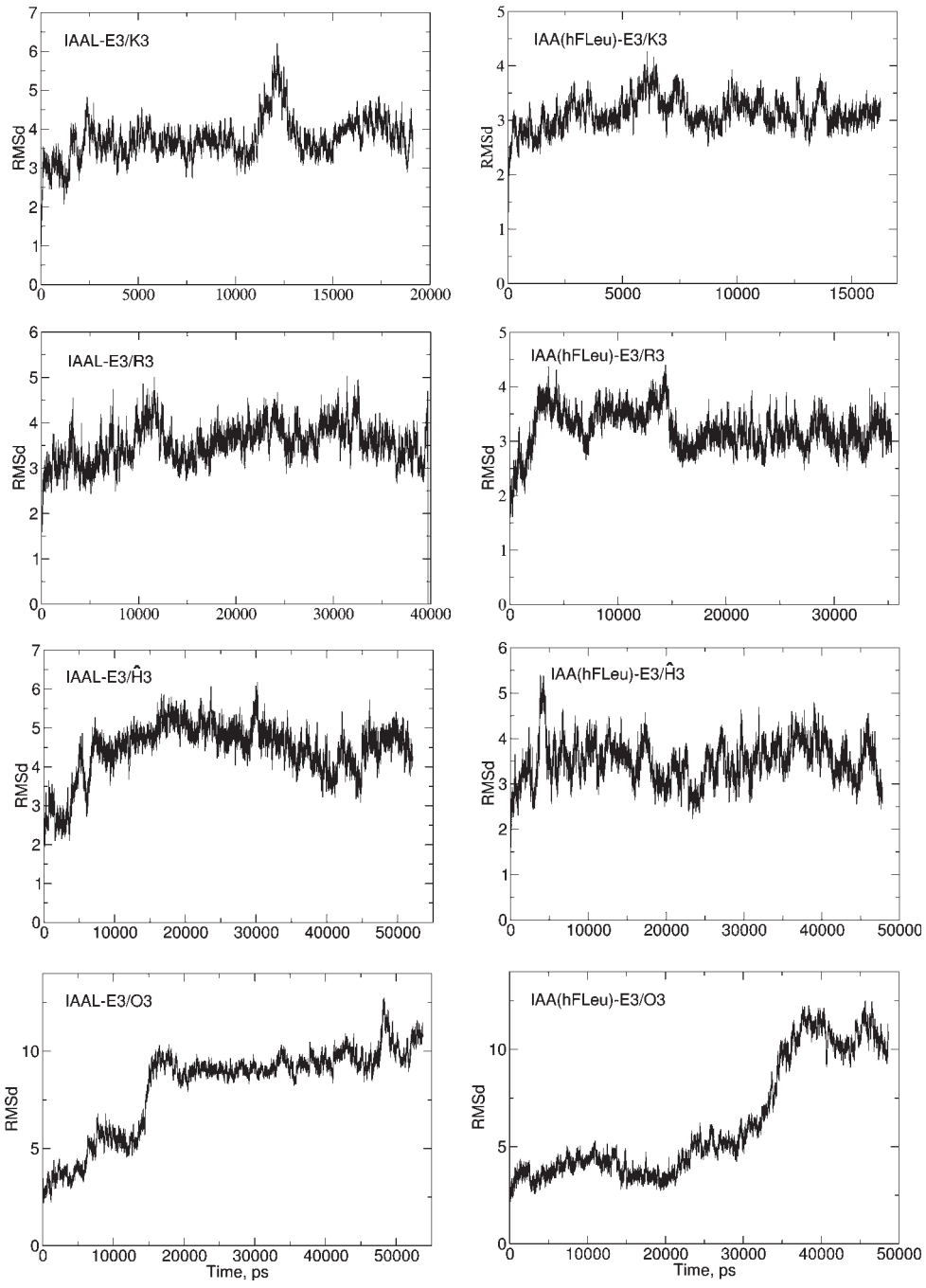


Figure 4

Root mean square deviation plots of selected coiled-coil dimers. Shown from left to right and top to bottom are IAAL-E3/K3, IAA(hFLeu)-E3/K3, IAAL-E3/R3, IAA(hFLeu)-E3/R3, IAAL-E3/Ĥ3, IAAL-E3/Ĥ3, IAAL-E3/O3, and IAA(hFLeu)-E3/O3. Note that the scales in each RMSD plot are different and the units are Å versus time (ps). Large fluctuations seen in IAAL-E3/O3 and IAA(hFLeu)-E3/O3 dimers reveal large conformation changes and loss of secondary structure.

along the coiled-coil backbone. Clear hydration, evidenced by high occupancy water sites, was found at the distal hydrophilic face of the coiled-coil near the b and c backbone oxygen and nitrogen atoms and also near the extended f and f′ residues. Distinct hydration around the residues participating in the salt bridges (positions e/g′ or e′/g) was observed peripheral to the dimer interface, although very few high occupancy hydration sites were observed directly at the salt bridge interface. Hydrogen bond analysis shows that the side chain atoms of the salt

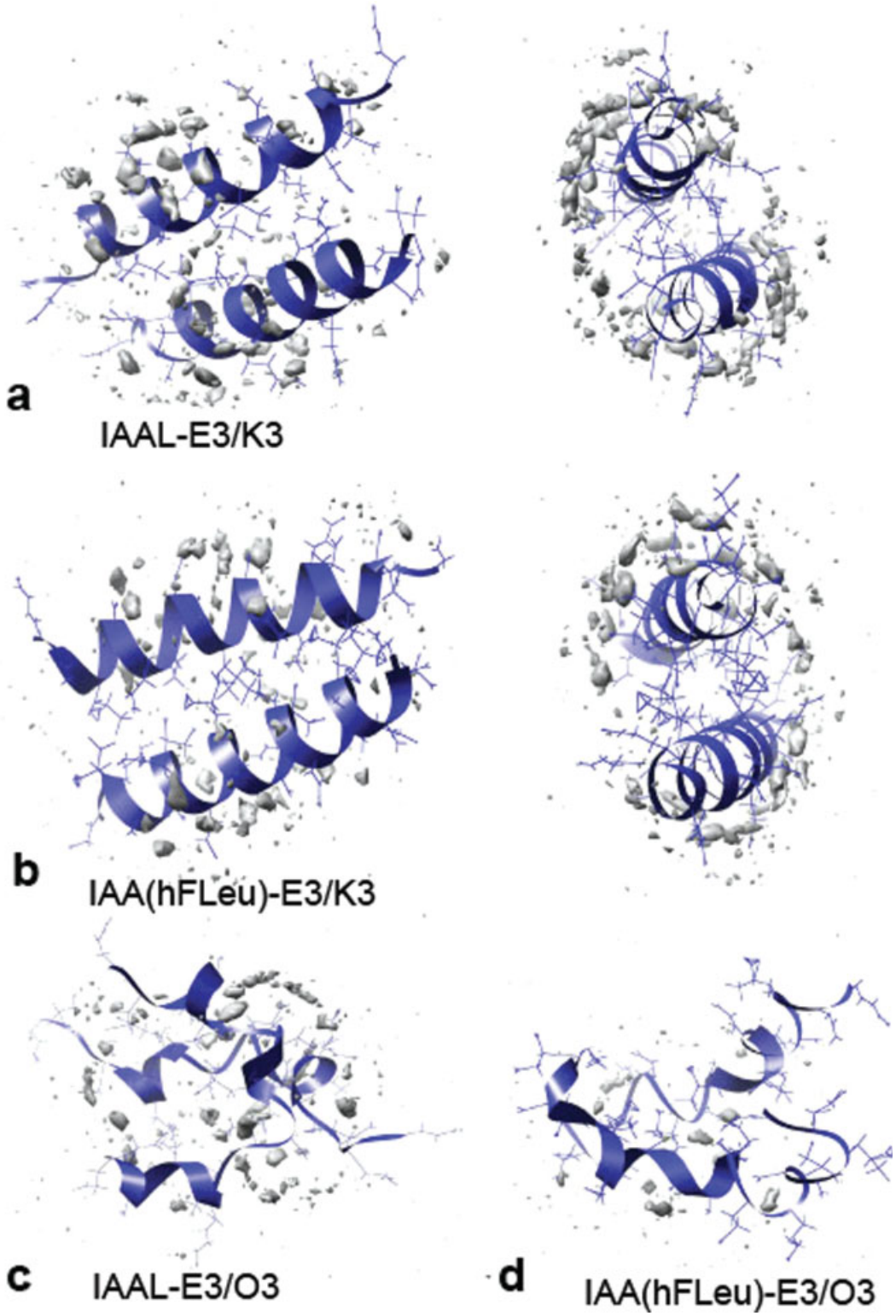


Figure 5

Hydration sites of selected coiled-coil dimers. High occupancy water hydration sites are displayed in gray in both side view and top down views of the (**a**) IAAL-E3/K3 and (**b**) IAA(hFLeu)-E3/K3 coiled-coils. Also shown are the side views of the (**c**) IAAL-E3/O3 and (**d**) IAA(hFLeu)-E3/O3 coiled-coils. The latter figures display distinct hydration within the hydrophobic core. The contouring of the water occupancy grids is shown at 12.0 hits per (0.5 Å)^3 grid from a 1 ns portion of the trajectory which is an occupancy ~2.87 times greater than bulk solvation.[157] The water density is superimposed on an average structure calculated over the same interval.

bridge residues form multiple transitory solvent-peptide hydrogen bonds and transient salt bridges. As the salt bridge interactions are rather dynamic, this tends to inhibit the formation of long term, high occupancy water interactions at the salt bridge. For all of the coiled-coils, with the exception of the ornithine substituted coiled-coils, essentially no ordered hydration sites are evident at the helix–helix hydrophobic interface made up of residues a and d. Different hydration patterns where observed for the relatively unstable ornithine coiled-coil dimers (IAAL-E3/O3 and IAA(hFLeu)-E3/O3). The loss of secondary structure in the ornithine coiled-coils effectively allows on the order of 2–4 waters into the hydrophobic core. One of the three largest hydration pockets shows contributory interactions from six separate hydrogen bond forming atoms (Ala11 at N, Glu13 at OE1,OE2, Leu12 at N, Ala10 at N, and Ile37 at O). Smaller hydration pockets can be found along backbone atoms in structured domains with hydration pockets in the unfolded regions where water interacts with multiple proximal ionic groups. In Figure 5, distinctly less order hydration is seen around the backbone atoms in the IAAL-E3/O3 and IAA(hFLeu)-E3/O3 dimers; this is likely a result of the larger fluctuations in the structure and larger RMSD values, which smear the time average.

Helicity measurements

Fluorinating α-helices introduces the structural changes that have been reported to both the increase and decrease helicity.[34,37] To better understand and quantify the changes in helicity as the core coiled-coil is mutated or fluorinated, CD spectra were calculated using the DichroCalc program using equivalent time samplings from the individual MD calculations. In Table IV, we present the calculated mean residue ellipticity for the simulated coiled-coil dimer pairs. In each of the coiled-coil structures, the MD simulations suggest that the fluorinated coiled-coils are more helical-as inferred from the calculated ellipticities—than their nonfluorinated counterparts. As seen in Table II, Figure 5, and also in the supplementary material, it is clear that the ornithine substituted coiled-coils (IAAL-E3/O3 and IAA(hFLeu)-E3/O3) lose secondary and tertiary structure in the nanosecond time scales sampled in MD simulation. Individual snapshots of these dimers from the MD trajectory do show regions of the α-helical secondary structure, however, less than is observed with the other coiled-coil structures. Moreover, the hydrophobic interface is largely disrupted. Both effects are likely the result of the steric strain induced by the shortened salt bridge.

For all of the other models where the coiled-coil structure was largely maintained, the measured ellipticity measurements are comparable. The standard lysine bridged coiled-coil (IAAL-E3/K3) was calculated to be the most helical of the nonfluorinated coiled-coil structures, suggesting that the lysine-glutamate salt bridge length to be the most ideal for nonfluorinated coiled-coils. A similar trend is seen in the fluorinated dimers, through the degree to which salt bridge length and orientation increase ellipticity is less resolvable because the fluorinated monomers are essentially fully helical. Our results contrast somewhat with the findings from Kennan and coworkers[14] who used variable length nonnatural amino acids in the **e** and **g** positions of the acid monomer (our work focused on the basic monomer). They found the optimal salt bridge length to be ~7–8 methylene units in the pHHGlu-pLys and pHGlu-pLys coiled-coil dimers (where H here represents a single methylene unit).[14] However, our simulations agree with their experiment in suggesting that the smaller salt bridges (such as with the E3/O3 coiled-coils) are less stable. Together these differences suggest that the salt bridge interactions are very subtle and contextual, consistent with previous work.[14,54,57,77,78,85,131,134,170]

Table IV
Calculated Mean Residue Ellipticity

Coiled coil dimer	$[\theta]_{222}$	Stdev
IAAL-E3/K3	−17,229.4	1176.1
IAA(hFLeu)-E3/K3	−17,516.1	1154.2
IAAL-E3/O3	−6573.1	688.8
IAA(hFLeu)-E3/O3	−8903.8	1540.2
IAAL-E3/R3	−14,831.7	1485.9
IAA(hFLeu)-E3/R3	−17,100.9	912.2
IAAL-E3/Ĥ3	−14,714.0	777.9
IAA(hFLeu)-E3/Ĥ3	−16,546.6	975.8

Shown are the calculated average ellipicities (degree cm^2 $dmol^{-1}$) at 222 nm and their standard deviations calculated from twenty representative snapshots from stable portions of the MD trajectories. The ellipicities were calculated using the DichroCalc program.[161]

Free energy calculations

Thermodynamic integration calculations were conducted to better understand the free energy consequences of replacing leucine with 5,5,5,5′,5′,5′-hexafluoroleucine (hFLeu). These were performed on very simple di-peptide model systems (see Fig. 3) to allow better understanding of the solvation free energy costs of fluorination when the side chain is exposed. Twelve Gaussian quadrature weighted sampling points were used to continuously transform the di-peptide models of leucine to hFLeu in the aqueous and gas phase environments. These thermodynamic integration results show that the hFLeu has a less favorable free energy of hydration by 1.1 kcal/mol (std. dev. 0.35 kcal/mol) when compared with leucine. This less favorable free energy of hydration would lead to energetic penalties if the residue is exposed to an aqueous environment. If fluorinated residues are incorporated at binding domain interfaces or positions in the α-helices

that undergo higher level interactions (such as coiled-coils or α-helical bundles), this decreased free energy of hydration, or destabilization of the exposed side chain, would drive higher order association leading to the hydrophobic burial. In coiled-coils where secondary structure in the monomers is driven by association, fluorination in the hydrophobic heptad positions leads to both increased association and helicity. Conversely, incorporation of fluorinated residues in the α-helices that do not undergo higher order interactions (or solvent exposed positions in coiled-coils) would lead to structural distortions and decreased helicity in an attempt to bury the hydrophobic atoms. Additional discussion of structural distortional in α-helices from incorporation of fluorinated residues can be found in related work.[39]

Rotamer preferences

Using the AMBER program ptraj, we analyzed the rotamers for leucine (hFLeu in the fluorinated molecules) and isoleucine in the coiled-coil dimers that we simulated. We found that incorporating fluorine leads to rotamer changes in residue side chains. Using the rotamer definitions in Tables II and III, we found a strong preference in hFLeu for the rotamer r1 as gauche− in the χ1 dihedral angle (−60°) and r2 as trans in the χ2 dihedral angle (180°).

This preference for *gauche*−, *trans* rotamers in 5,5,5,5′,5′,5′-hexafluoroleucine is likely a result of steric hindrance. Fluorine has a larger radius and is more electronegative than hydrogen. In the *trans*, *gauche*+ rotamers, significant charge repulsion and steric hindrance are likely to occur as shown schematically in the molecular graphics cartoon of Figure 6.

The two rotamer pairs, (r1 = *gauche*−, r2 = *trans*) and (r1 = *trans*, r2 = *gauche*+), seem to be the dominant rotamer choices seen in leucine residues of the coiled-coils dimers modeled. The choice of the rotamer pair appears to correlate with salt bridge distance and is likely to be part of a compensatory mechanism where the side chains in the hydrophobic domain expand or compress subject to the constraints imposed by the salt bridge length and fluorination (see Table V). We speculate that increased salt bridge length allows an expansion of the hydrophobic side chains, while still maintaining the hydrophobic core, and that fluorination increases the steric size of fluorinated residues and leads to side chain compression in the hydrophobic domain.) The choice of *gauche*-, *trans* rotamer pairs is found in both the coiled-coil dimers with normal or compressed dimerization domains. This is consistent with what is seen, with few exceptions, for the structures of coiled-coil dimers found in the RCSB Protein Data Bank (see supplementary material). When the salt bridge length is increased (for example with mutations of lysine to arginine or (S)-2,7-diaminoheptanoic acid), the rotamer preference is shifted towards expanded side chain geometries with *trans*, *gauche*+ rotamers.

Our simulation results show greater variability and less predictability among isoleucine rotamers when compared with leucine rotamers. Rotamer choices for isoleucine residues are more influenced by the salt bridge length than fluorination. In Table VI, we summarize the rotamer pairs for both sets of monomers. A trend from

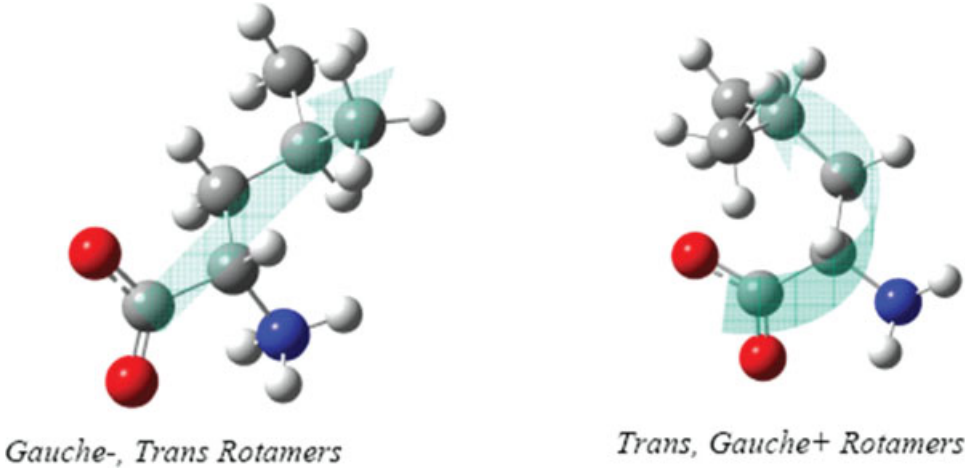


Figure 6

Selected leucine/hFLeu side chain rotamers. These two side chain leucine rotamers are the most common choices (for leucine and hexafluoroleucine residues) seen in our simulations of three heptad repeat coiled-coils. Green arrows added to emphasize differences in overall length and linear persistence of the two rotamer presented here. For hFLeu, the *trans*, *gauche*+ rotamer is disfavored due to the unfavorable steric interaction of the larger trifluoromethyl groups with the backbone and unfavorable electrostatic interactions of the electronegative fluorines with oxygen.

Table V
Leucine/hFLeu Sidechain Rotamer Distributions in Coiled-Coil Dimers Calculated from the Central Heptads in Both Monomers

Dimer	Monomer1 rotamers (r1,r2)	Monomer2 rotamers (r1,r2)	Occupied (%)
IAAL-E3K3	gauche−, trans	gauche−, trans	95.7
	trans, gauche+	gauche−, trans	4.3
IAA(hFLeu)-E3/K3	gauche−, trans	gauche−, trans	100.0
IAAL-E3/R3	trans, gauche+	trans, gauche+	48.5
	trans, gauche+	gauche−, trans	35.8
	trans, gauche+	gauche−, gauche+	15.8
IAA(hFLeu)-E3/R3	gauche−, trans	gauche−, trans	99.6
IAAL-E3/Ĥ3	trans, gauche+	trans, gauche+	49.9
	trans, gauche+	gauche−, trans	43.0
	trans, gauche+	gauche−, gauche−	4.2
	gauche−, trans	gauche−, trans	2.9
IAA(hFLeu)-E3/Ĥ3	gauche−, trans	gauche−, trans	90.7
	trans, gauche+	gauche−, trans	9.3

Note that occupancy calculations were measured as specific pairwise rotamer contributions of both monomer rotamer sets; calculated occupancy reflects rotamer interactions across the entire hydrophobic domain rather than within a single monomer.

r1 = *gauche*−, r2 = *gauche*− to r1 = *trans*, r2 = *gauche*+ rotamers is seen as the protein–protein interaction domain is first compressed then expanded (see Table VII and Fig. 7). This pattern strongly correlates to the salt bridge length with the sole exception of IAA(hFLeu)-E3/R3, which is likely influenced by the favorable electrostatic interactions with fluorine in the core (see below). Barring this exception, the choice of isoleucine rotamers can be described as a tradeoff between the total expansion, or compression, of the side chain necessary to satisfy the salt bridge while also providing the space filling bulk necessary to maintain the hydrophobic interface.

In general, the fluorinated coiled-coils show a decrease in the total number of side chain rotamer states accessed by the fluorinated residues. This decrease in populated rotamer states reflects a decrease in the hydrophobic side chain conformation entropy ($\mathbf{S}_\chi$) in fluorinated coiled-coils.[171] The larger size of the fluorinated residues rigidifies the folded structures and leads to a loss of conformational entropy. Such a loss should be compensated by other terms if the net effect of fluorination is stabilizing.[171] In Table VIII, the conformational entropy of the hydrophobic side chain rotamers ($\mathbf{S}_\chi$) was calculated by considering relative distributions of rotamers in fluorinated and nonfluorinated coiled-coils (as measured by percentage contribution, p_i) using Eq. (1). For this calculation, fluorinated and nonfluorinated dimer measurements were pooled respective to fluorine incorporation to allow for complete sampling of all potential rotamers irrespective of preferences derived from the salt bridge lengths

Table VI
Sidechain Rotamer Distributions for Isoleucine Residues in Coiled-Coil Dimers

Dimer	Monomer1 (r1,r2)	Monomer2 (r1,r2)	Occupied (%)
IAA(hFLeu)-E3K3	gauche−, gauche−	gauche+, trans	24.6
	gauche−, trans	gauche+, trans	75.4
IAAL-E3K3	gauche−, gauche−	gauche+, trans	2.9
	gauche−, trans	gauche−, trans	7.8
	gauche+, trans	gauche−, trans	89.2
IAA(hFLeu)-E3/Ĥ3	gauche−, gauche−	trans, trans	2.1
	gauche−, gauche+	trans, trans	1.8
	gauche−, trans	trans, trans	84.9
	trans, trans	trans, trans	2.1
	trans, gauche+	gauche−, trans	9.2
IAAL-E3/ Ĥ3	gauche−, trans	trans, trans	22.2
	gauche−, trans	trans, gauche+	1.8
	trans, trans	trans, trans	71.8
	trans, gauche+	trans, trans	4.2
IAA(hFLeu)-E3/R3	gauche−, trans	trans, trans	30.6
	trans, trans	gauche−, trans	9.3
	trans, trans	trans, trans	4.1
	trans, gauche+	gauche−, gauche−	19.2
	trans, gauche+	gauche−, trans	36.7
IAAL-E3/R3	gauche−, trans	trans, trans	7.7
	gauche−, trans	trans, gauche+	4.5
	trans, trans	gauche−, gauche−	1.6
	trans, trans	Trans, trans	70.8
	trans, gauche+	trans, trans	13.3
	trans, gauche+	trans, gauche+	2.1

Coiled-coil dimers are ordered from the most compressed to the most expanded dimer interface in descending order as determined by the isoleucine rotamer choices (see Figure 7). Rotamer measurements were taken from the isoleucine residues (position **a**) of the central heptad in both monomers. In cases where calculated rotamers are not identical for both monomers, the acid monomer isoleucine rotamer tends to be more expanded in the nonfluorinated dimers and more compressed in the fluorinated dimers.

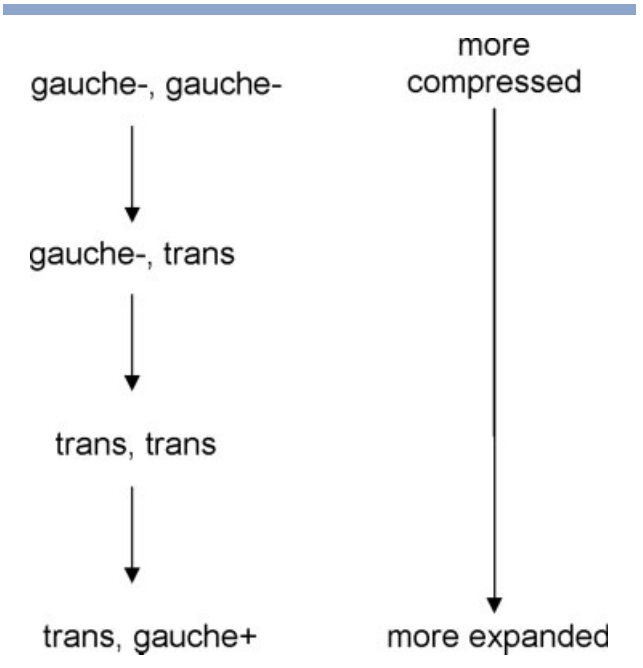


Figure 7
Isoleucine rotamer choices at varying coiled-coil compression levels. Trends in isoleucine rotamer choices (on left) seen as a function of coiled-coil compression. Rotamer choices are often complex adjustments to compensate for salt bridge lengths and to optimize for interior space filling and optimizing hydrophobic interactions.

Table VII
Intrahelical Cα-Cα Distances Between Leucine/hFLeu Residues as a Measurement of the Hydrophobic Interface

Coiled coil dimer	Leu/hFLeu Cα-Cα distance (Å)	Stdev.
IAAL-E3/K3	6.49	0.27
IAA(hFLeu)-E3/K3	6.76	0.30
IAAL-E3/R3	7.48	0.30
IAA(hFLeu)-E3/R3	7.11	0.30
IAAL-E3/Ĥ3	7.34	0.27
IAA(hFLeu)-E3/ Ĥ3	7.23	0.36

Measurements taken from snapshots spanning the final 2 ns of stable simulation time between residues 9 and 30 in the central heptad. Expansion and compression in coiled-coils can be seen to be influenced strongly by both fluorination and salt bridge length.

Table IX
Salt Bridge Distance Summary

Dimer pair	Average salt bridge distance	Participating atoms
IAAL-E3K3	8.4 Å (stdev 2.1)	Glu13 at Cδ Lys29 at Nζ
IAA(hFLeu)-E3/K3	6.8 Å (stdev 1.9)	Glu13 at Cδ Lys29 at Nζ
IAAL-E3O3	21.3 Å (stdev 1.5)	Glu13 at Cδ Orn29 at Nε
IAA(hFLeu)-E3/O3	15.6 Å (stdev 1.7)	Glu13 at Cδ Orn29 at Nε
IAAL-E3/R3	6.2 Å (stdev 2.3)	Glu13 at Cδ Arg29 at Nη1,Nη2
IAA(hFLeu)-E3/R3	4.1 Å (stdev 1.0)	Glu13 at Cδ Arg29 at Nη1,Nη2
IAAL-E3/Ĥ3	6.1 Å (stdev 2.7)	Glu13 at Cδ DAH29 at Nη
IAA(hFLeu)-E3/Ĥ3	5.4 Å (stdev 2.2)	Glu13 at Cδ DAH29 at Nη

and geometries. Direct interactions of hydrophobic residues require that conformational entropy be calculated as a function of rotamer occupancy across the dimer interphase (both monomers) producing a set of 81 unique rotamers (3 conformations per rotamer, 2 rotamers per side-chain, 2 residues = 3^4 unique rotamer sets). From Table VIII, a ten fold difference in side chain conformational entropy is seen between leucine and hexafluoroleucine while the isoleucine rotameric conformational entropies are nearly equivalent comparing fluorinated and nonfluorinated. The simulations also show that multiple rotameric sub-states are sampled over the course of the MD simulations. Time courses for the rotamer angles for leucine and isoleucine from the IAAL-E3/R3 coiled-coil are shown in the supplementary material.

$$S_\chi = -R\sum_{i=1}^{81} p_i \ln p_i \qquad (1)$$

Salt bridges and electrostatics

Coiled-coils are a flexible, dynamic protein–protein binding motif.[172] For most of the structural mutations, small conformational and rotameric changes are observed. These compensatory changes in side chain rotations and salt bridge length do have limits. The trajectories for both IAAL-E3/O3 and IAA(hFLeu)-E3/O3 dimers show a rapid loss of secondary structure during simulation. This suggests that the salt bridge formed by the glutamate–ornithine interaction was too short to span the hydrophobic interface. The resulting dimer maintained association, however much of the secondary structure and binding interface was lost or distorted. Similar instability due to shorted salt bridge lengths was seen in related work.[14] Table IX displays the distance of the salt bridge for many of the coiled-coil dimers that were successfully modeled.

Table VIII
Calculated Conformational Entropy for Hydrophobic Sidechain Rotamers

	Leucine/hFLeu	Isoleucine	Combined
Fluorinated	0.12 cal/(mol K)	1.41 cal/(mol K)	1.53 cal/(mol K)
Nonfluorinated	1.16 cal/(mol K)	1.20 cal/(mol K)	2.36 cal/(mol K)

Measurements taken from pooled fluorinated and nonfluorinated coiled-coil dimers using measurement of the rotamers sampled by the central heptad repeat. Individual rotamer contributions to overall conformational entropy calculations may be found in the supplemental materials table S3.

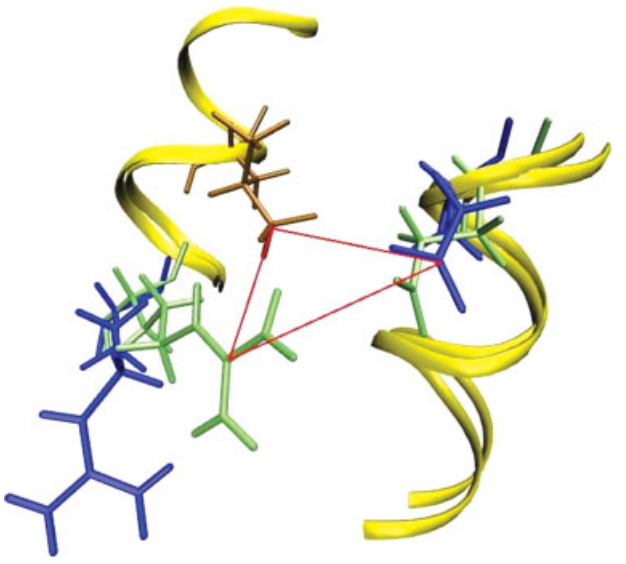


Figure 8
Increased electrostatic interactions between Arg29, Glu13, and hFLeu33 allow the formation of an unusually strong salt bridge. Superimposed IAAL-E3/R3 and IAA(hFLeu)-E3/R3 structures are represented by secondary structure ribbon with explicit tube representation of residues Arg29 and Glu13 in blue for IAAL-E3/R3, Arg29 and Glu13 in light green for IAA(hFLeu)-E3/R3. hFLeu33 is shown in brown and red lines are drawn for emphasis. Due to averaging artifacts, this figure was generated from a single representative snapshot at equivalent stable simulation times from both IAAL-E3/K3 and IAA(hFLeu)-E3/K3 trajectories. As discussed in the text, salt bridge formation is a transitory event and the increased displacement of Arg29 from Glu13 in the IAAL-E3/R3 representation reflects salt bridge separation.

The relatively small standard deviation and short average distance seen in the salt bridge formed by the IAA(hFLeu)-E3/R3 coiled-coil is distinctive. Residue by residue pair-wise free energy decomposition shows a stabilization of −1.8 kcal/mol for combined van der Waals and electrostatic energetics among the Arg, Glu, and hFLeu interacting residues (see Fig. 8) when compared with IAAL-E3/R3. The driving force for this stabilization derives from an increased average electrostatic interaction of −2.8 kcal/mol between Arg and hFLeu pairs. This suggests that the orientation of the hFLeu residue allows for additional favorable electrostatic interactions to stabilize the positioning of the arginine residue close to the glutamate residue. This increased favorable interaction allows for a stronger and more stable glutamate–arginine salt bridge. Additionally, fluctuations in the MD simulation RMSd (after convergence) appear lower with the hFLeu substituted E3/R3 coiled-coil. Note also that the salt bridges are transient over the course of the MD simulations as shown through the time-sources of salt bridge distances in Figure S9 in the supplementary material.

CONCLUSIONS

In this study, we have used simulation methods to explore the structural and energetic differences between fluorinated and nonfluorinated coiled-coils with varying salt bridge lengths and geometries. We found that independent of salt bridge length, fluorinated coiled-coils showed a higher ellipticity than their complementary nonfluorinated pairs.

Using thermodynamic integration, we were able to calculate the free energy of hydration of 5,5,5,5′,5′,5′-hexafluoroleucine (hFLeu) to be 1.1 kcal/mol less stable than leucine. We speculate that this destabilizing characteristic of fluorinated residues drives higher order association in coiled-coils to facilitate the burial of hydrophobic atoms exposed to an aqueous environment when the fluorinated atoms are found in the binding domain. This hydrophobic driving force in coiled-coils results in favorable ellipticity, suggestive of increased helicity, due to the coupling of association and secondary structure in coiled-coils.[173] In contrast, helicity is often reduced when incorporation of fluorinated residues occurs in isolated α-helices because the unfavorable free energy of hydration of hFLeu drives burial of the hydrophobic atoms at the expense of helicity.[39]

In structural studies, we found that hFLeu and leucine have different sidechain rotamer preferences. Leucine sidechain rotamers progress from *gauche*−, *trans* in compressed and normal trajectories to *trans*, *gauche*+ in expanded trajectories. 5,5,5,5′,5′,5′-hexafluoroleucine shows a strong preference for the single rotamer *gauche*−, *trans*. This difference in rotamer preferences is likely a result of both structural and electrostatic contributions. Fluorine atoms have a larger radius than hydrogen, so that the increased atomic size could lead to steric clashes between the fluorine atoms and the backbone oxygen atoms in the *trans*, *gauche*− rotamer state (see Fig. 6). Flourine covalently attached to carbon atoms is more negative than carbon bound hydrogens. In the *trans*, *gauche*− rotamer the negatively charged fluorine atoms come in close contact to the negatively charged backbone oxygen which could result in electrostatic repulsion. Sidechain conformation entropy in fluorinated coiled-coil dimers was calculated (see Table VIII) and found to be significantly different for leucine and hexalfuorleucine sidechains due to size and geometric differences in fluorinated residues. These rotamer preference differences combined with earlier studies on steric mass[37,43] suggest that fluorinated residues cannot be considered isomorphic with their nonfluorinated compliments.

Observation of the average salt bridge lengths revealed that electrostatic differences in fluorinated residues may also play a role in stabilizing the salt bridge interactions that reinforce coiled-coil association. Pairwise energetic decomposition of IAA(hFLeu)-E3/R3 trajectories show an increased favorable energetic interaction of −1.8 kcal/mol between Arg and hFLeu pairs when compared with IAAL-E3/R3. Analysis of the coiled-coil structural interactions (see Fig. 7), suggests that the orientation of the electronegative fluorine atoms of the hFLeu residue stabilizes the positioning of Arg proximal to the Glu residue.

Finally, the differences that we have shown between the fluorinated and nonfluorinated coiled-coil dimers suggest that additional engineering and optimization of key residues in fluorinated dimers may lead to additional stabilization of the coiled-coil. This could prove useful for planned applications of coiled-coils as targeting agents, biosensors, and nanomanipulators. Extension of these findings to antiparallel coiled-coils could be applicable in principle, however the results are likely to be context specific. Although coiled-coil ellipticity, expansion, hydration, energetics, and salt-bridge behavior will likely remain similar, antiparallel coiled-coils differ in the residue and position-specific interactions in the hydrophobic binding domain. Trends seen in isoleucine and potentially in leucine and 5,5,5,5′,5′,5′-hexafluoroleucine rotamer preferences and behavior may vary significantly from those reported here.

ACKNOWLEDGMENTS

We wish to thank Seonah Kim for the useful discussion of methods to calculate mean residue ellipticity.

REFERENCES

1. Adamson JG, Zhou NE, Hodges RS. Structure, function and application of the coiled-coil protein folding motif. Curr Opin Biotechnol 1993;4:428–437.
2. Lupas AN. Coiled coils: new structures and new functions. Trends Biochem Sci 1996;21:375–382.

3. Burkhard P, Stetefeld J, Strelkov SV. Coiled coils: a highly versatile protein folding motif. Trends Cell Biol 2001;11:82–88.
4. Mason JM, Arndt KM. Coiled coil domains: Stability, specificity, and biological implications. Chembiochem 2004;6:170–176.
5. Woolfson DN. The design of coiled-coil structures and assemblies. Adv Protein Chem 2005;70:79–112.
6. Lupas AN, Gruber M. The structure of alpha-helical coiled coils. Adv Protein Chem 2005;70:37–78.
7. Liu J, Zheng Q, Deng Y, Cheng CS, Kallenbach NR, Lu M. A seven-helix coiled coil. Proc Natl Acad Sci USA 2006;103:15457–15462.
8. Gonzalez L, Jr, Plecs JJ, Alber T. An engineered allosteric switch in leucine-zipper oligomerization. Nat Struct Biol 1996;3:510–514.
9. Chao H, Bautista DL, Litowski J, Irvin RT, Hodges RS. Use of a heterodimeric coiled-coil system for biosensor application and affinity purification. J Chromatogr B Biomed Sci Appl 1998;715:307–329.
10. Wang C, Stewart RJ, Kopecek J. Hybrid hydrogels assembled from synthetic polymers and coiled-coil protein domains. Nature 1999;397:417–420.
11. Tang A, Wang C, Stewart RJ, Kopecek J. The coiled coils in the design of protein-based constructs: hybrid hydrogels and epitope displays. J Control Release 2001;72:57–70.
12. Litowski JR, Hodges RS. Designing heterodimeric two-stranded alpha-helical coiled-coils. Effects of hydrophobicity and alpha-helical propensity on protein folding, stability, and specificity. J Biol Chem 2002;277:37272–37279.
13. Yu YB. Coiled-coils: stability, specificity, and drug delivery potential. Adv Drug Deliv Rev 2002;54:1113–1129.
14. Ryan SJ, Kennan AJ. Variable stability heterodimeric coiled-coils from manipulation of electrostatic interface residue chain length. J Am Chem Soc 2007;129:10255–10260.
15. Doerr AJ, McLendon GL. Design, folding, and activities of metal-assembled coiled coil proteins. Inorg Chem 2004;43:7916–7925.
16. Yang J, Wang H. The central coiled-coil domain and carboxyl-terminal WD-repeat domain of Arabidopsis SPA1 are responsible for mediating repression of light signaling. Plant J 2006;47:564–576.
17. Hildebrandt ER, Gheber L, Kingsbury TJ, Hoyt MA. Homotetrameric form of CIN8P, an S. cerevisiae kinesin-5 motor, is essential for its in vivo function. J Biol Chem 2006;281:26004–26013.
18. Watanabe YH, Takano M, Yoshida M. ATP binding to nucleotide binding domain (NBD)1 of the ClpB chaperone induces motion of the long coiled-coil, stabilizes the hexamer, and activates NBD2. J Biol Chem 2005;280:24562–24567.
19. Fukai S, Nureki O, Sekine S, Shimada A, Vassylyev DG, Yokoyama S. Mechanism of molecular interactions for tRNA(Val) recognition by valyl-tRNA synthetase. RNA 2003;9:100–111.
20. Sorensen JB, Wiederhold K, Muller EM, Milosevic I, Nagy G, de Groot BL, Grubmuller H, Fasshauer D. Sequential N- to C-terminal SNARE complex assembly drives priming and fusion of secretory vesicles. EMBO J 2006;25:955–966.
21. Heldwein EE, Lou H, Bender FC, Cohen GH, Eisenberg RJ, Harrison SC. Crystal structure of glycoprotein B from herpes simplex virus 1. Science 2006;313:217–220.
22. Rose A, Meier I. Scaffolds, levers, rods and springs: diverse cellular functions of long coiled-coil proteins. Cell Mol Life Sci 2004;61:1996–2009.
23. Yu YB. Fluorocarbon nanoparticles as multifunctional drug delivery vehicles. J Drug Target 2006;14:663–669.
24. Kukhar VP, Soloshonok VA. Fluorine-containing amino acids: synthesis and properties. Chichester, New York: Wiley; 1995. p 422.
25. Xing X, Fichera A, Kumar K. A novel synthesis of enantiomerically pure 5,5,5,5′,5′,5′-hexafluoroleucine. Org Lett 2001;3:1285–1286.
26. Jiang ZX, Yu YB. The synthesis of a geminally perfluoro-tert-butylated beta-amino acid and its protected forms as a potential pharmacokinetic modulator and reporter for peptide-based pharmaceuticals. J Org Chem 2007;72:1464–1467.
27. Rennert OM, Anker HS. On the incorporation of 5′,5′,5′-trifluoroleucine into proteins of *E. coli*. Biochemistry 1963;2:471–476.
28. Tang Y, Ghirlanda G, Vaidehi N, Kua J, Mainz DT, Goddard IiiWA, DeGrado WF, Tirrell DA. Stabilization of coiled-coil peptide domains by introduction of trifluoroleucine. Biochemistry 2001;40:2790–2796.
29. Wang P, Tang Y, Tirrell DA. Incorporation of trifluoroisoleucine into proteins in vivo. J Am Chem Soc 2003;125:6900–6906.
30. Doyon JB, Jain A. The pattern of fluorine substitution affects binding affinity in a small library of fluoroaromatic inhibitors for carbonic anhydrase. Org Lett 1999;1:183–185.
31. Bilgicer B, Fichera A, Kumar K. A coiled coil with a fluorous core. J Am Chem Soc 2001;123:4393–4399.
32. Meng H, Kumar K. Antimicrobial activity and protease stability of peptides containing fluorinated amino acids. J Am Chem Soc 2007;129:15615–15622.
33. Lin JH, Lu AY. Role of pharmacokinetics and metabolism in drug discovery and development. Pharmacol Rev 1997;49:403–449.
34. Tang Y, Ghirlanda G, Petka WA, Nakajima T, DeGrado WF, Tirrell DA. Fluorinated coiled-coil proteins prepared in vivo display enhanced thermal and chemical stability. Angew Chem Int 2001; 40:1494–1496.
35. Bilgicer B, Kumar K. Synthesis and thermodynamic characterization of self-sorting coiled coils. Tetrahedron 2002;58:4105–4112.
36. Yoder NC, Kumar K. Fluorinated amino acids in protein design and engineering. Chem Soc Rev 2002;31:335–341.
37. Lee K-H, Lee H-Y, Slutsky MM, Anderson JT, Marsh ENG. Fluorous effect in proteins: de novo design and characterization of a four-alpha-helix bundle protein containing hexafluoroleucine. Biochemistry 2004;43:16277–16284.
38. Lee H-Y, Lee K-H, Al-Hashimi HM, Marsh ENG. Modulating protein structure with fluorous amino acids: increased stability and native-like structure conferred on a 4-helix bundle protein by hexafluoroleucine. J Am Chem Soc 2006;128:337–343.
39. Chiu HP, Suzuki Y, Gullickson D, Ahmad R, Kokona B, Fairman R, Cheng RP. Helix propensity of highly fluorinated amino acids. J Am Chem Soc 2006;128:15556–15557.
40. Bilgicer B, Xing X, Kumar K. Programmed self-sorting of coiled coils with leucine and hexafluoroleucine cores. J Am Chem Soc 2001;123:11815–11816.
41. Pal PP, Bae JH, Azim MK, Hess P, Friedrich R, Huber R, Moroder L, Budisa N. Structural and spectral response of *Aequorea victoria* green fluorescent proteins to chromophore fluorination. Biochemistry 2005;44:3663–3672.
42. Nagai T, Nishioka G, Koyama M, Ando A, Miki T, Kumadaki I. Reactions of trifluoromethyl ketones. IX. Investigation of the steric effect of a trifluoromethyl group based on the stereochemistry on the dehydration of trifluoromethyl homoallyl alcohols J Fluorine Chem 1992;57:229–237.
43. Bondi A. van der Waals volumes and radii. J Phys Chem 1964;68:441–451.
44. Mason JM, Muller KM, Arndt KM. Considerations in the design and optimization of coiled coil structures. Methods Mol Biol 2007;352:35–70.
45. Crick F. The packing of α-helices: simple coiled-coils. Acta Crystallogr 1953;6:689–697.
46. Bosshard HR, Durr E, Hitz T, Jelesarov I. Energetics of coiled coil folding: the nature of the transition states. Biochemistry 2001;40:3544–3552.
47. Phillips GN, Jr. What is the pitch of the alpha-helical coiled coil. Proteins 1992;14:425–429.
48. Lumb KJ, Kim PS. A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. Biochemistry 1995;34:8642–8648.
49. Nautiyal S, Woolfson DN, King DS, Alber T. A designed heterotrimeric coiled coil. Biochemistry 1995;34:11645–11651.
50. Nautiyal S, Alber T. Crystal structure of a designed, thermostable heterotrimeric coiled coil. Protein Sci 1999;8:84–90.

51. McClain DL, Woods HL, Oakley MG. Design and characterization of a heterodimeric coiled coil that forms exclusively with an antiparallel relative helix orientation. J Am Chem Soc 2001;123:3151–3152.
52. van Hest JC, Tirrell DA. Protein-based materials, toward a new level of structural control. Chem Commun 2001;19:1897–1904.
53. Arndt KM, Pelletier JN, Muller KM, Pluckthun A, Alber T. Comparison of in vivo selection and rational design of heterodimeric coiled coils. Structure 2002;10:1235–1248.
54. Campbell KM, Lumb KJ. Complementation of buried lysine and surface polar residues in a designed heterodimeric coiled coil. Biochemistry 2002;41:7169–7175.
55. Schnarr NA, Kennan AJ. Peptide tic-tac-toe: heterotrimeric coiled-coil specificity from steric matching of multiple hydrophobic side chains. J Am Chem Soc 2002;124:9779–9783.
56. Gurnon DG, Whitaker JA, Oakley MG. Design and characterization of a homodimeric antiparallel coiled coil. J Am Chem Soc 2003;125:7518–7519.
57. Schnarr NA, Kennan AJ. Sequential and specific exchange of multiple coiled-coil components. J Am Chem Soc 2003;125:13046–13051.
58. Schnarr NA, Kennan AJ. Strand orientation by steric matching: a designed antiparallel coiled-coil trimer. J Am Chem Soc 2004;126:14447–14451.
59. Yadav M, Redman JE, Leman LJ, Alvarez-Gutierrez JM, Zhang Y, Stout CD, Ghadiri MR. Structure-based engineering of internal cavities in coiled-coil peptides. Biochemistry 2005;44:9723–9732.
60. Vieth M, Kolinski A, Brooks CL, III, Skolnick J. Prediction of the folding pathways and structure of the GCN4 leucine zipper. J Mol Biol 1994;237:361–367.
61. Lumb KJ, Carr CM, Kim PS. Subdomain folding of the coiled coil leucine zipper from the bZIP transcriptional activator GCN4. Biochemistry 1994;33:7361–7367.
62. Gonzalez L, Jr, Brown RA, Richardson D, Alber T. Crystal structures of a single coiled-coil peptide in two oligomeric states reveal the basis for structural polymorphism. Nat Struct Biol 1996;3:1002–1010.
63. Burkhard P, Ivaninskii S, Lustig A. Improving coiled-coil stability by optimizing ionic interactions. J Mol Biol 2002;318:901–910.
64. Dragan AI, Potekhin SA, Sivolob A, Lu M, Privalov PL. Kinetics and thermodynamics of the unfolding and refolding of the three-stranded alpha-helical coiled coil. Lpp-56. Biochemistry 2004;43:14891–14900.
65. Meisner WK, Sosnick TR. Fast folding of a helical protein initiated by the collision of unstructured chains. Proc Natl Acad Sci USA 2004;101:13478–13482.
66. Meisner WK, Sosnick TR. Barrier-limited, microsecond folding of a stable protein measured with hydrogen exchange: implications for downhill folding. Proc Natl Acad Sci USA 2004;101:15639–15644.
67. Missimer JH, Steinmetz MO, Jahnke W, Winkler FK, van Gunsteren WF, Daura X. Molecular-dynamics simulations of C- and N-terminal peptide derivatives of GCN4-p1 in aqueous solution. Chem Biodivers 2005;2:1086–1104.
68. Bjelic S, Karshikoff A, Jelesarov I. Stability and folding/unfolding kinetics of the homotrimeric coiled coil Lpp-56. Biochemistry 2006;45:8931–8939.
69. Bunagan MR, Cristian L, DeGrado WF, Gai F. Truncation of a cross-linked GCN4-p1 coiled coil leads to ultrafast folding. Biochemistry 2006;12:10981–10986.
70. d'Avignon DA, Bretthorst GL, Holtzer ME, Schwarz KA, Angeletti RH, Mints L, Holtzer A. Site-specific experiments on folding/unfolding of Jun coiled coils: thermodynamic and kinetic parameters from spin inversion transfer nuclear magnetic resonance at leucine-18. Biopolymers 2006;83:255–267.
71. Steinmetz MO, Jelesarov I, Matousek WM, Honnappa S, Jahnke W, Missimer JH, Frank S, Alexandrescu AT, Kammerer RA. Molecular basis of coiled-coil formation. Proc Natl Acad Sci USA 2007;104:7062–7067.
72. Mason JM, Hagemann UB, Arndt KM. Improved stability of the Jun-Fos activator protein-1 coiled coil motif: a stopped-flow circular dichroism kinetic analysis. J Biol Chem 2007;282:23015–23024.
73. Nikolaev Y, Pervushin K. NMR spin state exchange spectroscopy reveals equilibrium of two distinct conformations of leucine zipper GCN4 in solution. J Am Chem Soc 2007;129:6461–6469.
74. Bornschlogl T, Reif M. Single-molecular dynamics of mechanical coiled-coil unzipping. Langmuir 2008;24:1338–1342.
75. O'Shea EK, Klemm JD, Kim PS, Alber T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science 1991;254:539–544.
76. Harbury PB, Zhang T, Kim PS, Alber T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science 1993;262:1401–1407.
77. Junius FK, Mackay JP, Bubb WA, Jensen SA, Weiss AS, King GF. Nuclear magnetic resonance characterization of the Jun leucine zipper domain: unusual properties of coiled-coil interfacial polar residues. Biochemistry 1995;34:6164–6174.
78. Yu Y, Monera OD, Hodges RS, Privalov PL. Investigation of electrostatic interactions in two-stranded coiled-coils through residue shuffling. Biophys Chem 1996;59:299–314.
79. Gonzalez L, Jr, Woolfson DN, Alber T. Buried polar residues and structural specificity in the GCN4 leucine zipper. Nat Struct Biol 1996;3:1011–1018.
80. Caballero-Herrera A, Nilsson L. Molecular dynamics simulations of the E1/E2 transmembrane domain of the Semliki Forest virus. Biophys J 2003;85:3646–3658.
81. Danciulescu C, Nick B, Wortmann FJ. Structural stability of wild type and mutated alpha-keratin fragments: molecular dynamics and free energy calculations. Biomacromolecules 2004;5:2165–2175.
82. Smith TA, Steinert PM, Parry DA. Modeling effects of mutations in coiled-coil structures: case study using epidermolysis bullosa simplex mutations in segment 1a of K5/K14 intermediate filaments. Proteins 2004;55:1043–1052.
83. Ambroggio XI, Kuhlman B. Computational design of a single amino acid sequence that can switch between two distinct protein folds. J Am Chem Soc 2006;128:1154–1161.
84. Yadav MK, Leman LJ, Price DJ, Brooks CL, III, Stout CD, Ghadiri MR. Coiled coils at the edge of configurational heterogeneity. Structural analyses of parallel and antiparallel homotetrameric coiled coils reveal configurational sensitivity to a single solvent-exposed amino acid substitution. Biochemistry 2006;45:4463–4473.
85. Yu Y, Monera OD, Hodges RS, Privalov PL. Ion pairs significantly stabilize coiled-coils in the absence of electrolyte. J Mol Biol 1996;255:367–372.
86. Zhong Q, Moore PB, Newns DM, Klein ML. Molecular dynamics study of the LS3 voltage-gated ion channel. FEBS Lett 1998;427:267–270.
87. Gorfe AA, Ferrara P, Caflisch A, Marti DN, Bosshard HR, Jelesarov I. Calculation of protein ionization equilibria with conformational sampling: pK(a) of a model leucine zipper, GCN4 and barnase. Proteins 2002;46:41–60.
88. Madhusoodanan M, Lazaridis T. Investigation of pathways for the low-pH conformational transition in influenza hemagglutinin. Biophys J 2003;84:1926–1939.
89. Ding F, LaRocque JJ, Dokholyan NV. Direct observation of protein folding, aggregation, and a prion-like conformational conversion. J Biol Chem 2005;280:40235–40240.
90. Hawkins RJ, McLeish TC. Dynamic allostery of protein alpha helical coiled-coils. J R Soc Interface 2006;3:125–138.
91. Vagt T, Zschornig O, Huster D, Koksch B. Membrane binding and structure of de novo designed alpha-helical cationic coiled-coil-forming peptides. Chemphyschem 2006;7:1361–1371.
92. Laughton CA, Luisi BF, Pratap JV, Calladine CR. A potential molecular switch in an alpha-helical coiled coil. Proteins 2007;70:25–30.

93. Palaiomylitou M, Tartas A, Vlachakis D, Tzamarias D, Vlassi M. Investigating the structural stability of the Tup1-interaction domain of Ssn6: evidence for a conformational change on the complex. Proteins 2007;70:72–82.
94. Holtzer ME, Bretthorst GL, d'Avignon DA, Angeletti RH, Mints L, Holtzer A. Temperature dependence of the folding and unfolding kinetics of the GCN4 leucine zipper via 13C(alpha)-NMR. Biophys J 2001;80:939–951.
95. Nishikawa K, Scheraga HA. Geometrical criteria for formation of coiled-coil structures of polypeptide chains. Macromolecules 1976;9:395–407.
96. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. Science 1991;252:1162–1164.
97. Tropsha A, Bowen JP, Brown FK, Kizer JS. Do interhelical side chain-backbone hydrogen bonds participate in formation of leucine zipper coiled coils? Proc Natl Acad Sci USA 1991;88:9488–9492.
98. Cregut D, Liautard JP, Heitz F, Chiche L. Molecular modeling of coiled-coil alpha-tropomyosin: analysis of staggered and in register helix-helix interactions. Protein Eng 1993;6:51–58.
99. Zhang L, Hermans J. Molecular dynamics study of structure and stability of a model coiled coil. Proteins 1993;16:384–392.
100. Kerr ID, Sankararamakrishnan R, Smart OS, Sansom MS. Parallel helix bundles and ion channels: molecular modeling via simulated annealing and restrained molecular dynamics. Biophys J 1994;67: 1501–1515.
101. Rozzelle JE, Jr, Tropsha A, Erickson BW. Rational design of a three-heptad coiled-coil protein and comparison by molecular dynamics simulation with the GCN4 coiled coil: presence of interior three-center hydrogen bonds. Protein Sci 1994;3:345–355.
102. Berger B, Wilson DB, Wolf E, Tonchev T, Milla M, Kim PS. Predicting coiled coils by use of pairwise residue correlations. Proc Natl Acad Sci USA 1995;92:8259–8263.
103. Harbury PB, Tidor B, Kim PS. Repacking protein cores with backbone freedom: structure prediction for coiled coils. Proc Natl Acad Sci USA 1995;92:8408–8412.
104. Offer G, Sessions R. Computer modelling of the alpha-helical coiled coil: packing of side-chains in the inner core. J Mol Biol 1995;249:967–987.
105. Sankararamakrishnan R, Sansom MS. Modelling packing interactions in parallel helix bundles: pentameric bundles of nicotinic receptor M2 helices. Biochim Biophys Acta 1995;1239:122–132.
106. Vieth M, Kolinski A, Brooks CL, III, Skolnick J. Prediction of quaternary structure of coiled coils. Application to mutants of the GCN4 leucine zipper. J Mol Biol 1995;251:448–467.
107. Woolfson DN, Alber T. Predicting oligomerization states of coiled coils. Protein Sci 1995;4:1596–1607.
108. Hirst JD, Vieth M, Skolnick J, Brooks CL III. Predicting leucine zipper structures from sequence. Protein Eng 1996;9:657–662.
109. Vieth M, Kolinski A, Skolnick J. Method for predicting the state of association of discretized protein models. Application to leucine zippers. Biochemistry 1996;35:955–967.
110. Garnier N, Genest D, Duneau JP, Genest M. Molecular modeling of c-erbB2 receptor dimerization: coiled-coil structure of wild and oncogenic transmembrane domains–stabilization by interhelical hydrogen bonds in the oncogenic form. Biopolymers 1997;42:157–168.
111. Wolf E, Kim PS, Berger B. MultiCoil: a program for predicting two- and three-stranded coiled coils. Protein Sci 1997;6:1179–1189.
112. Lupas A. Predicting coiled-coil regions in proteins. Curr Opin Struct Biol 1997;7:388–393.
113. Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS. High-resolution protein design with backbone freedom. Science 1998;282:1462–1467.
114. Zhong Q, Jiang Q, Moore PB, Newns DM, Klein ML. Molecular dynamics simulation of a synthetic ion channel. Biophys J 1998; 74:3–10.
115. Zhong Q, Husslein T, Moore PB, Newns DM, Pattnaik P, Klein ML. The M2 channel of influenza A virus: a molecular dynamics study. FEBS Lett 1998;434:265–271.
116. Kukol A, Arkin IT. vpu transmembrane peptide structure obtained by site-specific fourier transform infrared dichroism and global molecular dynamics searching. Biophys J 1999;77:1594–1601.
117. Yang PK, Tzou WS, Hwang MJ. Restraint-driven formation of alpha-helical coiled coils in molecular dynamics simulations. Biopolymers 1999;50:667–677.
118. Kukol A, Arkin IT. Structure of the influenza C virus CM2 protein transmembrane domain obtained by site-specific infrared dichroism and global molecular dynamics searching. J Biol Chem 2000; 275:4225–4229.
119. Walshaw J, Woolfson DN. Open-and-shut cases in coiled-coil assembly: alpha-sheets and alpha-cylinders. Protein Sci 2001;10:668–673.
120. Walshaw J, Woolfson DN. Socket: a program for identifying and analysing coiled-coil motifs within protein structures. J Mol Biol 2001;307:1427–1450.
121. Briki F, Doucet J, Etchebest C. A procedure for refining a coiled coil protein structure using x-ray fiber diffraction and modeling. Biophys J 2002;83:1774–1783.
122. Orzechoski M, Cieplak P, Piela L. Theoretical calculations of the coiled-coil stability in water in the context of its possible use as a molecular rack. J Comput Chem 2002;23:106–110.
123. Ash WL, Stockner T, MacCallum JL, Tieleman DP. Computer modeling of polyleucine-based coiled coil dimers in a realistic membrane environment: insight into helix-helix interactions in membrane proteins. Biochemistry 2004;43:9050–9060.
124. Gruber M, Soding J, Lupas AN. Comparative analysis of coiled-coil prediction methods. J Struct Biol 2006;155:140–145.
125. Ramos J, Lazaridis T. Energetic determinants of oligomeric state specificity in coiled coils. J Am Chem Soc 2006;128:15499–15510.
126. Nilges M, Brunger AT. Automated modeling of coiled coils: application to the GCN4 dimerization domain. Protein Eng 1991;4:649–659.
127. Nilges M, Brunger AT. Successful prediction of the coiled coil geometry of the GCN4 leucine zipper domain by simulated annealing: comparison to the X-ray structure. Proteins 1993;15:133–146.
128. Mottamal M, Zhang J, Lazaridis T. Energetics of the native and non-native states of the glycophorin transmembrane helix dimer. Proteins 2006;62:996–1009.
129. Samna Soumana O, Garnier N, Genest M. Molecular dynamics simulation approach for the prediction of transmembrane helix-helix heterodimers assembly. Eur Biophys J 2007;36:1071–1082.
130. Pineiro A, Villa A, Vagt T, Koksch B, Mark AE. A molecular dynamics study of the formation, stability, and oligomerization state of two designed coiled coils: possibilities and limitations. Biophys J 2005;89:3701–3713.
131. Missimer JH, Steinmetz MO, Baron R, Winkler FK, Kammerer RA, Daura X, van Gunsteren WF. Configurational entropy elucidates the role of salt-bridge networks in protein thermostability. Protein Sci 2007;16:1349–1359.
132. El-Kettani MA, Smith JC. Pathways for conformational change in seryl-tRNA synthetase from *Thermus thermophilus*. C R Acad Sci III 1996;319:161–169.
133. Litowski JR, Hodges RS. Designing heterodimeric two-stranded alpha-helical coiled-coils: the effect of chain length on protein folding, stability and specificity. J Pept Res 2001;58:477–492.
134. Lindhout DA, Litowski JR, Mercier P, Hodges RS, Sykes BD. NMR solution structure of a highly stable de novo heterodimeric coiled-coil. Biopolymers 2004;75:367–375.
135. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 1997;18:2714–2723.
136. Case DA, Cheatham TE, III, Darden TA, Gohlker H, Luo R, Merz KM Jr, Onufriev AV, Simmerling C, Wang B, Woods R. The AMBER biomolecular simulation programs. J Comput Chem 2005; 26:1668–1688.
137. Simmerling C, Strockbine B, Roitberg AE. All-atom structure prediction and folding simulations of a stable protein. J Am Chem Soc 2002;124:11258–11259.

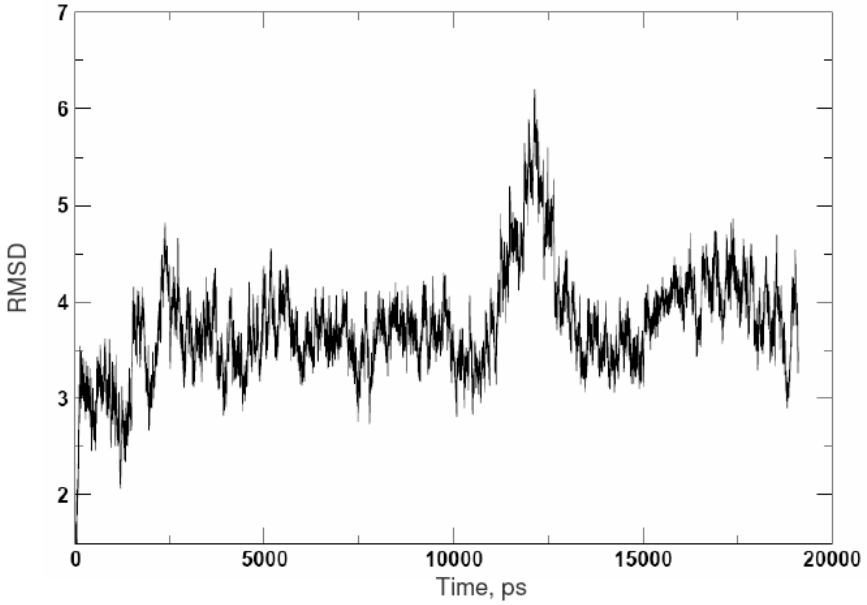
138. Cornell WD, Cieplak P, Baylyl CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins. Nucleic acids, and organic molecules. J Am Chem Soc 1995;117:5179–5197.
139. Kollman PA, Dixon R, Cornell W, Fox T, Chipot C, Pohorille A. The development/application of a 'minimalist' organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data. In: van Gunsteren WF, Weiner PK, Wilkinson AJ, editors. Computer simulation of biomolecular systems, Vol. 3. Norwell MA, USA: Kluwer Academic Press; 1997. p 83–96.
140. Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J Comput Chem 2000;21:1049–1074.
141. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 2006;65:712–725.
142. Roe DR, Okur A, Wickstrom L, Hornak V, Simmerling C. Secondary structure bias in generalized born solvent models: comparison of conformational ensembles and free energy of solvent polarization from explicit and implicit solvation. J Phys Chem B 2007;111: 1846–1857.
143. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, III, Laughton CA, Orozco M. Refinement of the AMBER force field for nucleic acids. Improving the description of alpha/gamma conformers. Biophys J 2007;11:3817–3829.
144. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, Debolt S, Ferguson D, Seibel G, Kollman P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structure and energetic properties of molecules. Comput Phys Commun 1995;91:1–41.
145. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys 1995; 103:8577–8593.
146. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparisons of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
147. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 1977;23:327–341.
148. Barth E, Kuczera K, Leimkuhler B, Skeel RD. Algorithms for constrained molecular dynamics. J Comput Chem 1995;16:1192–1209.
149. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Comput Phys 1984;81:3684–3690.
150. Hobza P, Kabelac M, Sponer J, Mejzlik P, Vondrasek J. Performance of empirical potentials (AMBER. CFF95, CVFF, CHARMM, OPLS, POLTEV), semiempirical quantum chemical methods (AM1, MNDO/M, PM3), and *ab initio* Hartree-Fock method for interaction of DNA bases: comparison with nonempirical beyond Hartree-Fock results. J Comput Chem 1997;18:1136–1150.
151. Showalter SA, Bruschweiler R. Quantitative molecular ensemble interpretation of NMR dipolar couplings without restraints. J Am Chem Soc 2007;129:4158–4159.
152. Mobley DL, Dumont E, Chodera JD, Dill KA. Comparison of charge models for fixed-charge force fields: small-molecule hydration free energies in explicit solvent. J Phys Chem B 2007;111: 2242–2254.
153. Nicholls A, Mobley DL, Guthrie JP, Chodera JD, Bayly CI, Cooper MD, Pande VS. Predicting small-molecule solvation free energies: an informal blind test for computational chemistry. J Med Chem 2008;51:769–779.
154. Daniels SM, Saunders RA, Platts JA. Prediction of fluorophilicity of organic and transition metal compounds using molecular surface areas. J Fluorine Chem 2004;125:1291–1298.
155. Friedemann R, Naumann S, Brickmann J. Aggregation of alkane and fluoroalkane clusters: molecular dynamics simulation results. Phys Chem Chem Phys 2001;3:4195–4199.
156. Friedemann R, Naumann S, Brickmann J. Molecular dynamics studies on the aggregation of Y-shaped fluoroalkanes. J Mol Model 2002;8:266–271.
157. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Tokoyo K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Sannenberg JJ, Zakrewski VG, Dapprich S, Daniels S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA. Gaussian 2003. Pittsburgh, PA: Gaussian, Inc.; 2003.
158. Granovsky AA. PC GAMESS version 7.0. 2006.
159. Bayly CI, Cieplak P, Cornell WD, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges- the RESP model. J Phys Chem 1993;97: 10269–10280.
160. Pigache A, Cieplak P, Dupradeau F-Y. Automatic and highly reproducible RESP and ESP charge derivation: Application to the development of programs RED and X RED, 227th ACS National Meeting, Anaheim, CA, USA, March 28–April 1, 2004.
161. Bryce R. Amber parameter database. 2006. http://www.pharmacy.manchester.ac.uk/bryce/amber
162. Case DA, Darden TA, Cheatham ITE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak B, Cui G, Beroza P, Scafmeister C, Caldwell JW, Ross WS, Kollman PA. AMBER8. San Francisco: University of California; 2004.
163. Dunbrack RL Jr, Cohen FE. Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci 1997;6:1661–1681.
164. Cheatham TE, Kollman PA. Molecular dynamics simulations highlight the structural differences among DNA:DNA, RNA:RNA, and DNA:RNA hybrid duplexes. J Am Chem Soc 1997;119:4805–4825.
165. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera–a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–1612.
166. Mezei M. The finite difference thermodynamic integration, tested on calculating the hydration free energy difference between acetone and dimethylamine in water. J Chem Phys 1987;86:7084–7088.
167. Shirts MR, Pitera JW, Swope WC, Pande VS. Extremely precise free energy calculations of amino acid side chain analogs: comparison of common molecular mechanics force fields for proteins. J Chem Phys 2003;119:5740–5761.
168. Gohlke H, Kiel C, Case DA. Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. J Mol Biol 2003;330: 891–913.
169. Bulheller BM, Rodger A, Hirst JD. Circular and linear dichroism of proteins. Phys Chem Chem Phys 2007;9:2020–2035.
170. Arndt KM, Pelletier JN, Muller KM, Alber T, Michnick SW, Pluckthun A. A heterodimeric coiled-coil peptide pair selected in vivo from a designed library-versus-library ensemble. J Mol Biol 2000; 295:627–639.
171. Yu YB, Lavigne P, Privalov PL, Hodges RS. The measure of interior disorder in a folded protein and its contribution to stability. J Am Chem Soc 1999;121:8443–8449.
172. Mucke N, Kreplak L, Kirmse R, Wedig T, Herrmann H, Aebi U, Langowski J. Assessing the flexibility of intermediate filaments by atomic force microscopy. J Mol Biol 2004;335:1241–1250.
173. Jelesarov I, Bosshard HR. Thermodynamic characterization of the coupled folding and association of heterodimeric coiled coils (leucine zippers). J Mol Biol 1996;263:344–358.

Supplemental Material for “Molecular dynamics guided study of salt bridge length dependence in both fluorinated and non-fluorinated parallel dimeric coiled coils” Pendley, S.S.; Yu, Y.B.; Cheatham, T. E. *Proteins* (2008).

Figure S1 – RMSD plots of coiled coil dimers. RMSD (Å) versus time (ps) with reference to the initial frame of each MD trajectory of the coiled coil dimers. Dimer name and sequence presented above each RMSD plot.

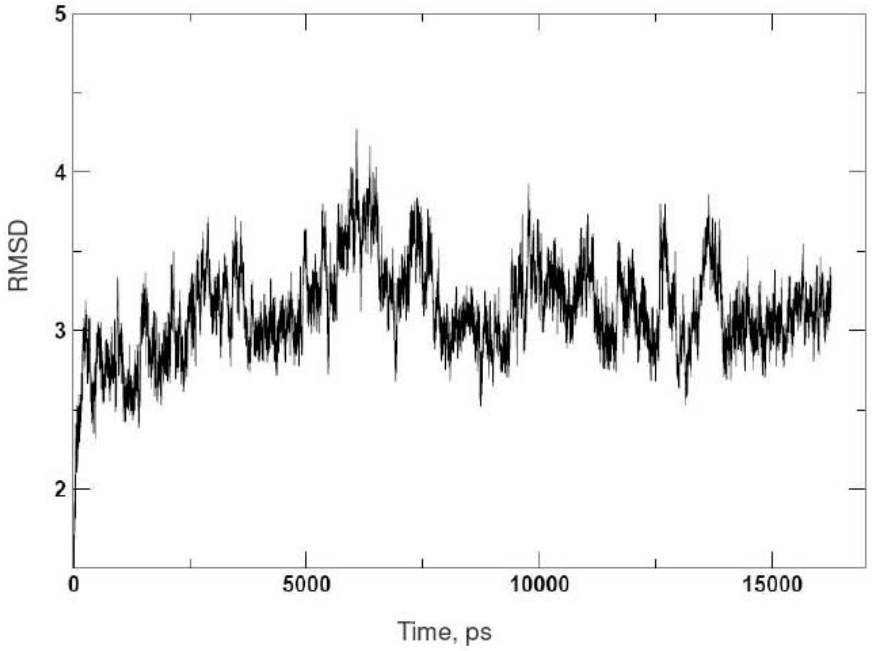
IAAL-E3/K3

Monomer A EIAALEKEIAALEKEIAALEK
Monomer B KIAALKEKIAALKEKIAALKE



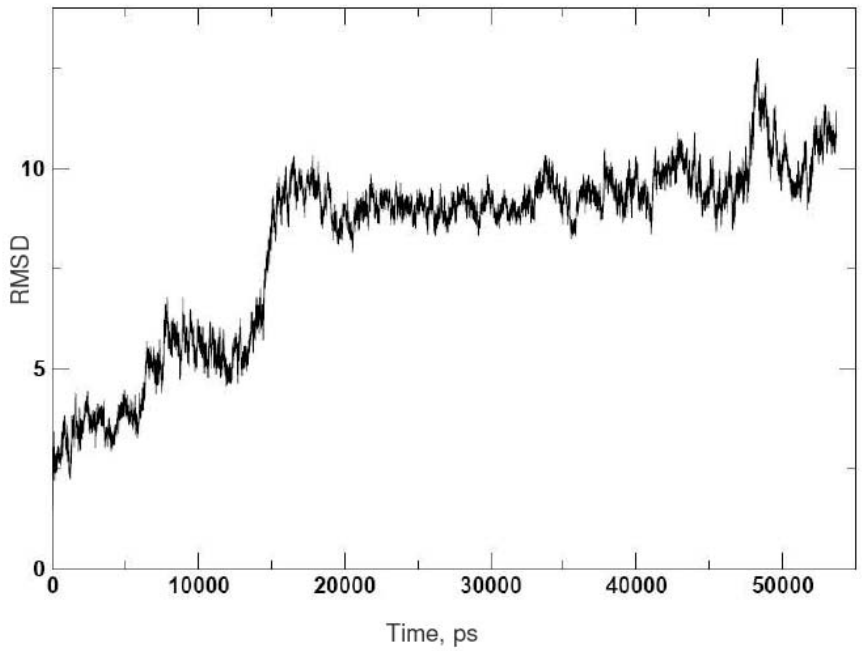
IAA(hFLeu)E3/K3

Monomer A EIAAĹEKEIAAĹEKEIAAĹEK
Monomer B KIAAĹKEKIAAĹKEKIAAĹKE



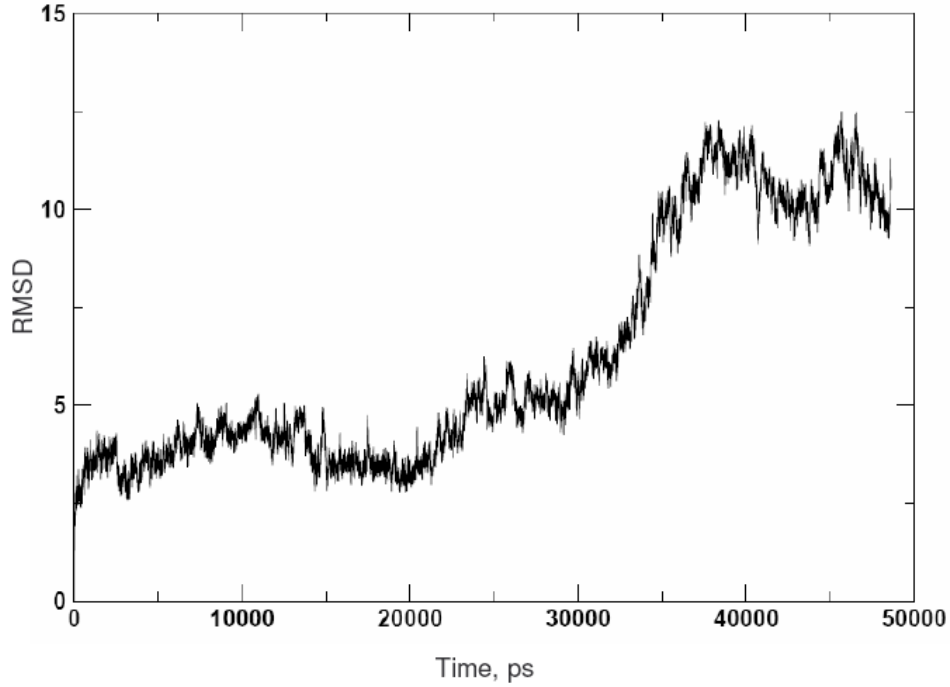
IAAL-E3/O3

Monomer A EIAALEKEIAALEKEIAALEK
Monomer B OIAALOEOIAALOEOIAALOE



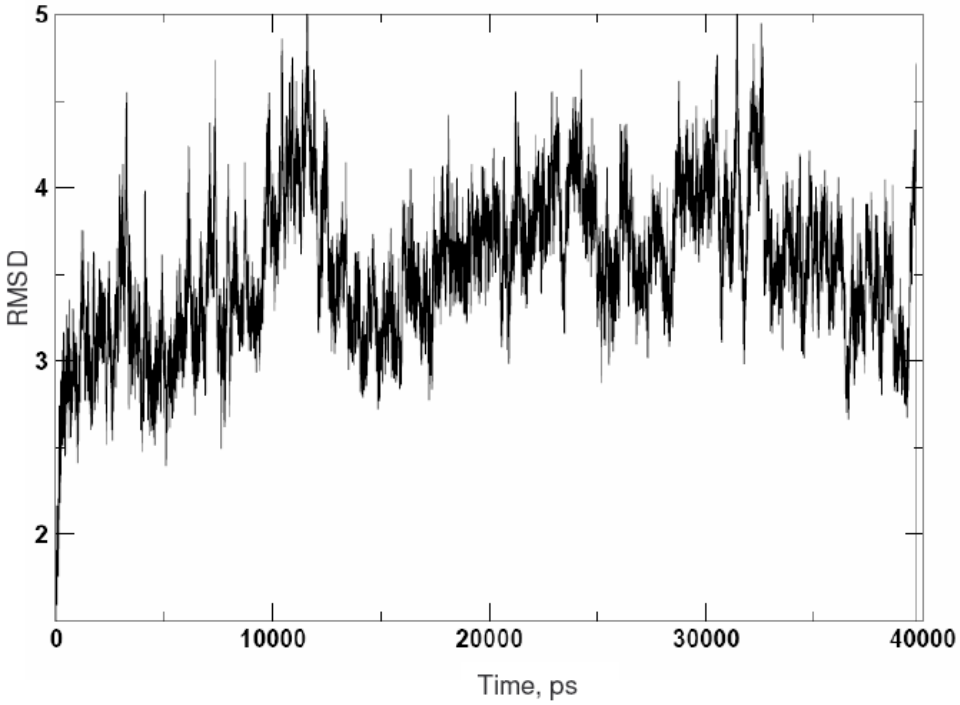
IAA(hFLeu)-E3/O3

Monomer A EIAAĹEKEIAAĹEKEIAAĹEK
Monomer B OIAAĹOEOIAAĹOEOIAAĹOE



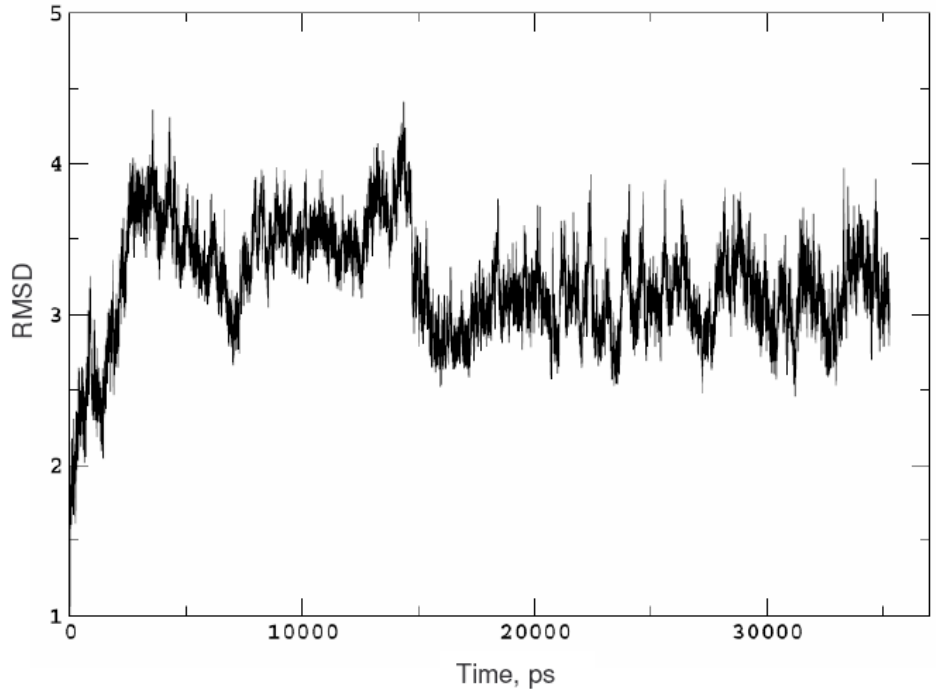
IAAL-E3/R3

Monomer A EIAALEKEIAALEKEIAALEK
Monomer B RIAALRERIAALRERIAALRE



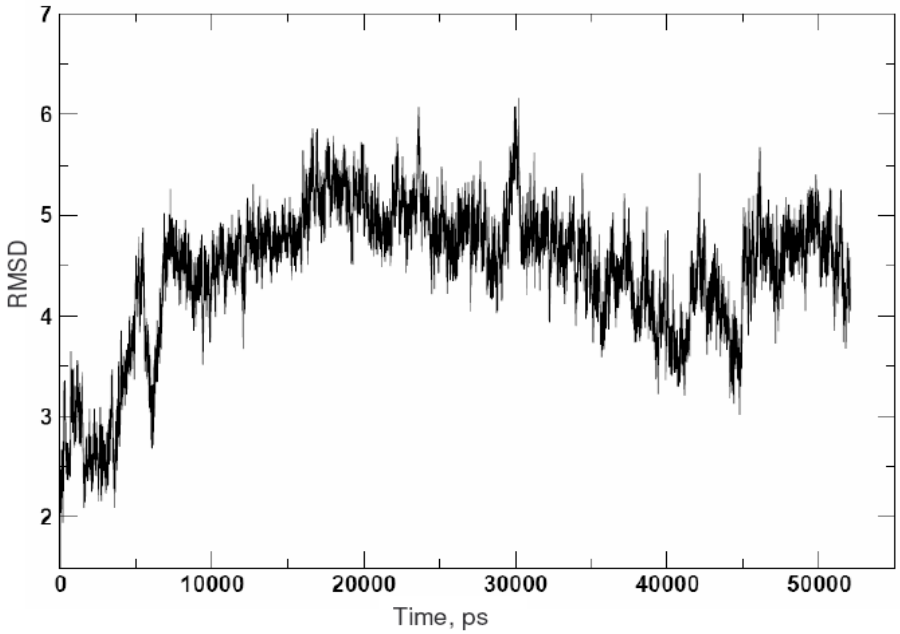
IAA(hFLeu)-E3/R3

Monomer A EIAAĹEKEIAAĹEKEIAAĹEK
Monomer B RIAAĹRERIAAĹRERIAAĹRE



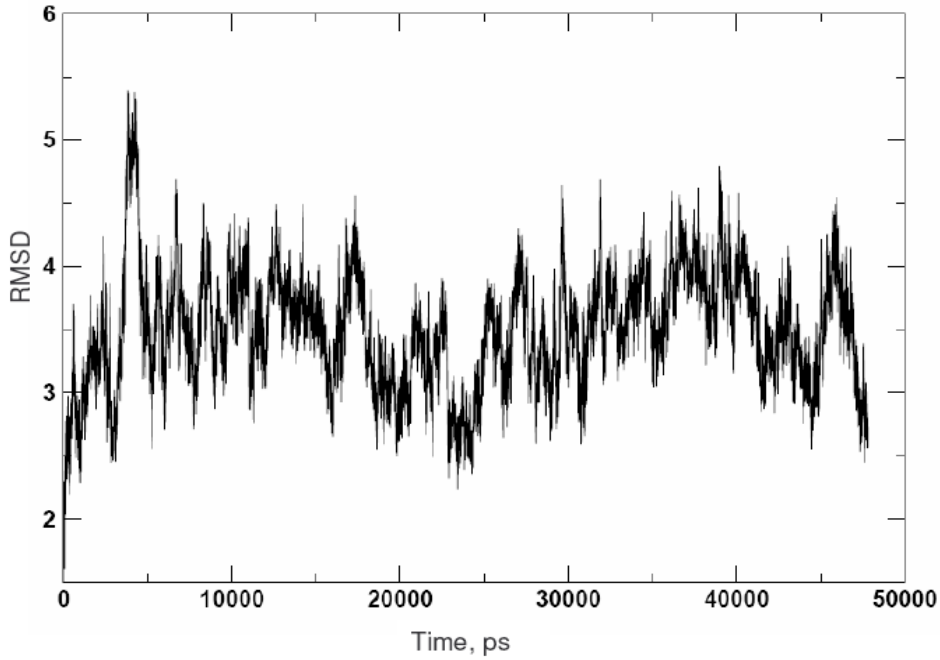
IAAL- E3/Ĥ3

Monomer A EIAALEKEIAALEKEIAALEK
Monomer B ĤIAALĤEĤIAALĤEĤIAALĤE



IAA(hFLeu)- E3/Ĥ3

Monomer A EIAAĹEKEIAAĹEKEIAAĹEK
Monomer B ĤIAAĹĤEĤIAAĹĤEĤIAAĹĤE



O = ornithine
Ĺ = 5,5,5,5’,5’,5’-hexafluoroleucine
Ĥ = (S)-2,7-diaminoheptanoic acid

Figure S2 – Force field comparison of simulations of coiled coils using modern nonpolarizable AMBER force fields. Stick and secondary structure representations of IAAL-E3/K3 as generated by NMR and MD simulations. The NMR structure is generated as an average of the published NMR structures (pdb: 1UOI). MD structures were generated as an average of 5-6 ns sampling taken from the final stable trajectory. Forcefields represented here show the original ff99 forcefield[1] and four variants that differ in their treatment of phi and psi backbone torsion angle parameterization. Mod1 is an early attempt to improve the treatment of torsional backbone parameters[2]. Mod 2 was developed by Junmei Wang and Ray Luo to correct for poor secondary structure behavior seen in the ff99 forcefield. ff99SB is the "Stony Brook" modification of the ff99 forcefield by Simmerling et al (a further refinement of the Mod 1 force field)[3]. Finally, ff03 adjusts the phi and psi backbone torsion angles as well as charge determination methodologies using quantum mechanics with a simulated continuum dielectic [4].

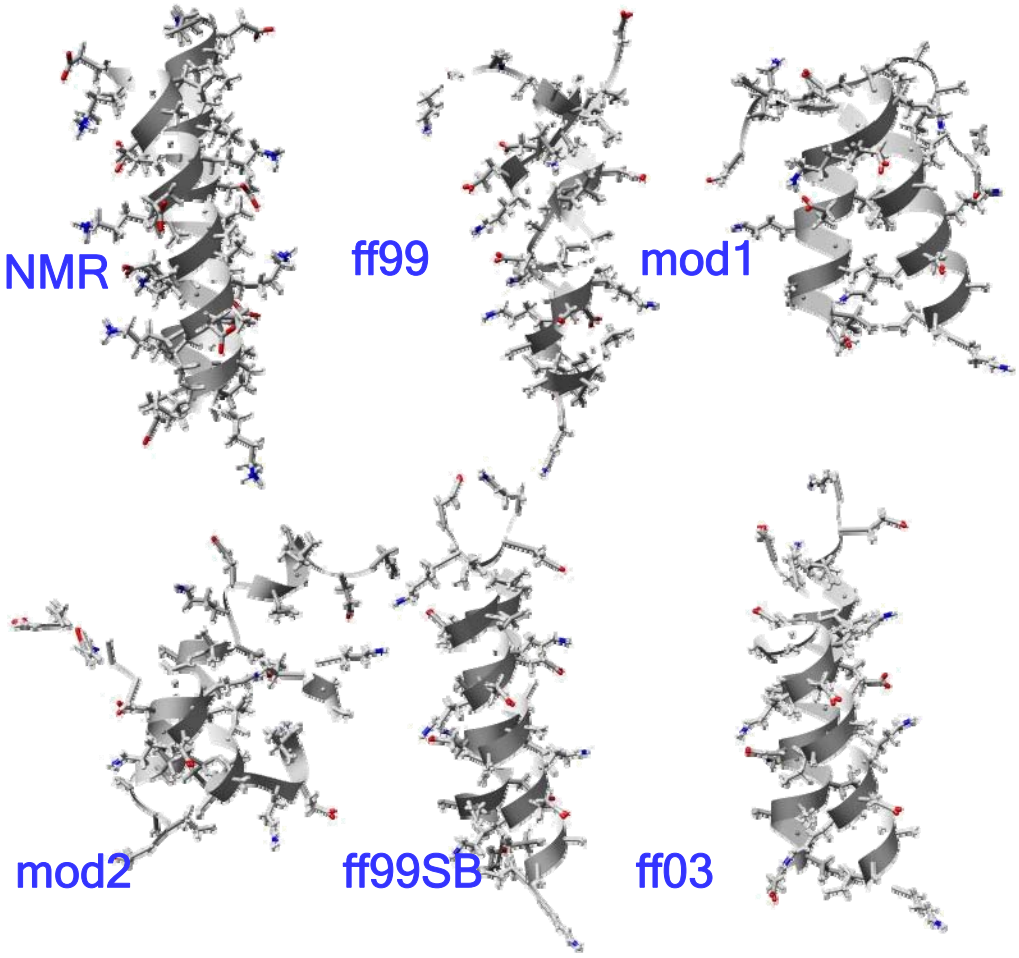


Figure S3 – Superimposition of IAAL-E3/K3 structures generated from AMBER forcefields ff99SB and ff03 with the average structures of IAAL-E3/K3 from NMR snapshots. All atom stick and secondary structure representations of IAAL-E3/K3 showing superimposition of NMR average with MD calculated structure. NMR average structure is shown in yellow, MD average structure is shown white. MD generated structure taken from 5-6 ns average of final stable trajectory.

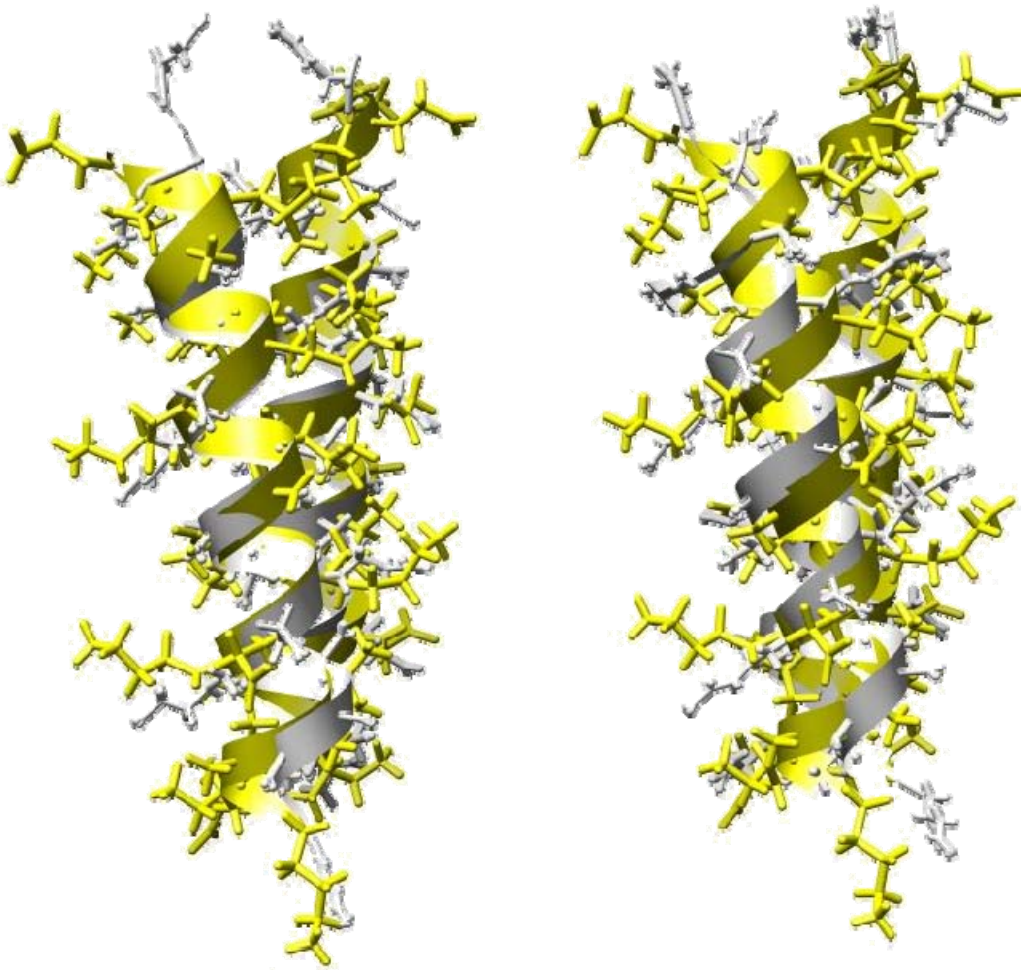
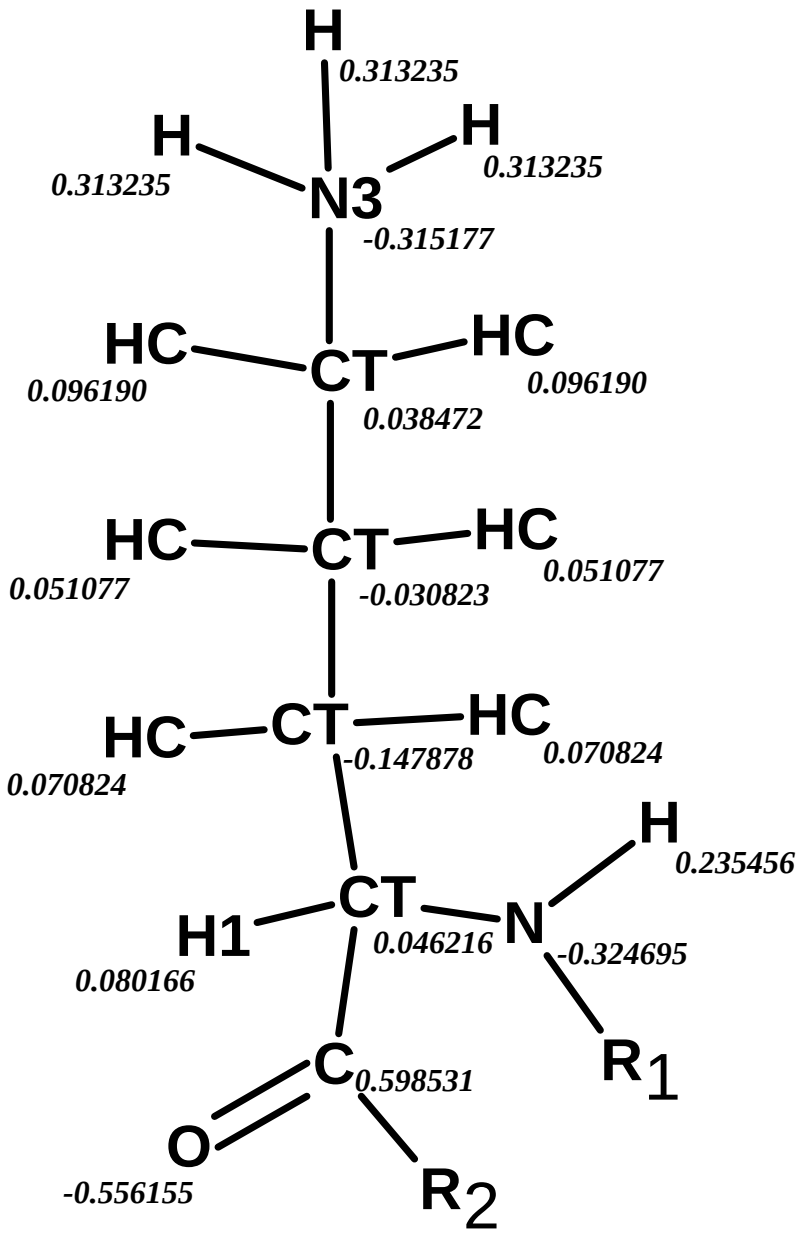
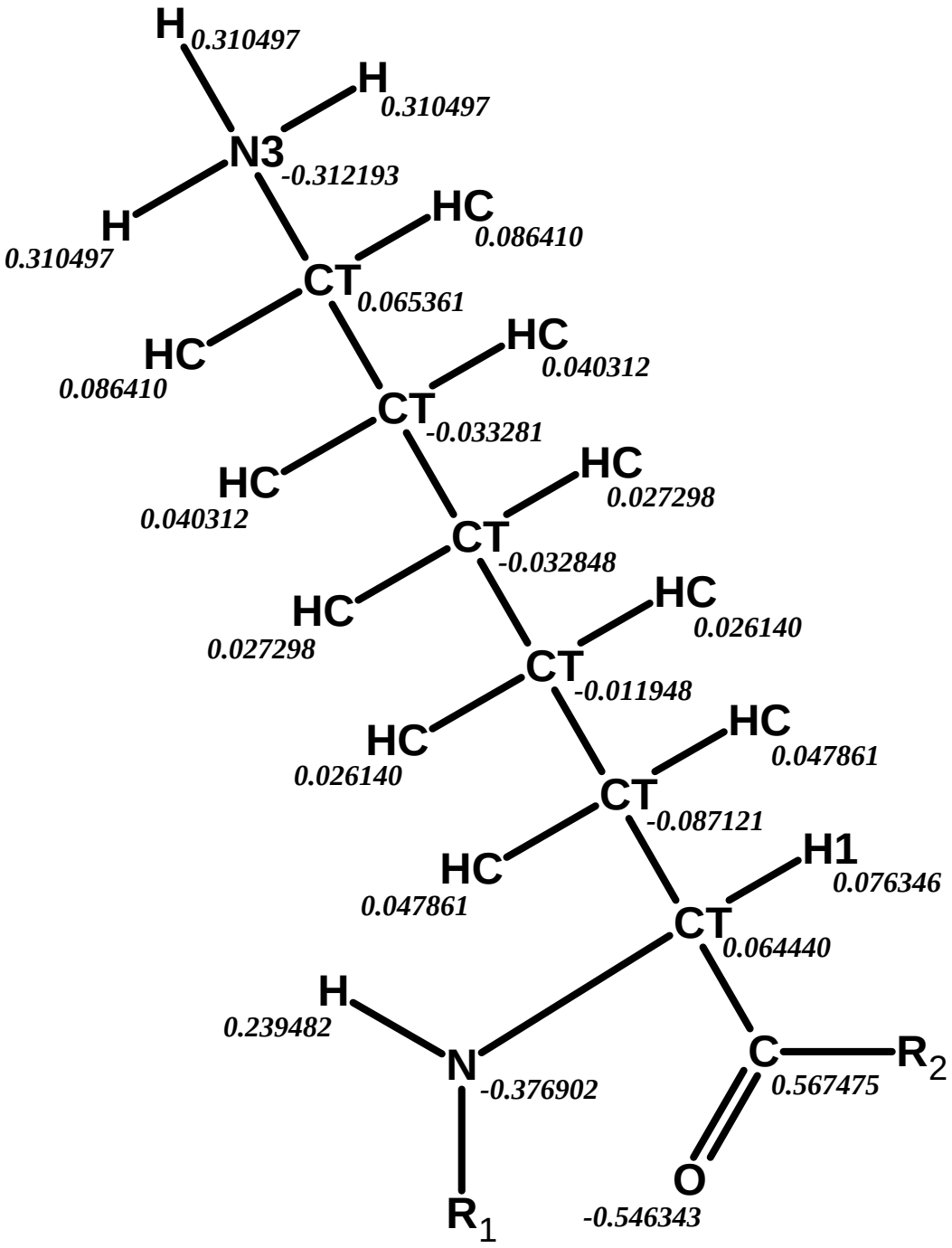


Figure S4 – Molecular structure, atom types, and charges for ornithine, (S)-2,7-diaminoheptanoic acid, and 5,5,5,5',5',5'-hexafluoroleucine. AMBER atom types are given at the atom positions, charges are displayed at the subscript position. R1 and R2 represent covalent bonds to the N-terminal and C-terminal attached residues respectively.

Ornithine



(S)-2,7-Diaminoheptanoic Acid



5,5,5,5’,5’,5’-Hexafluoroleucine

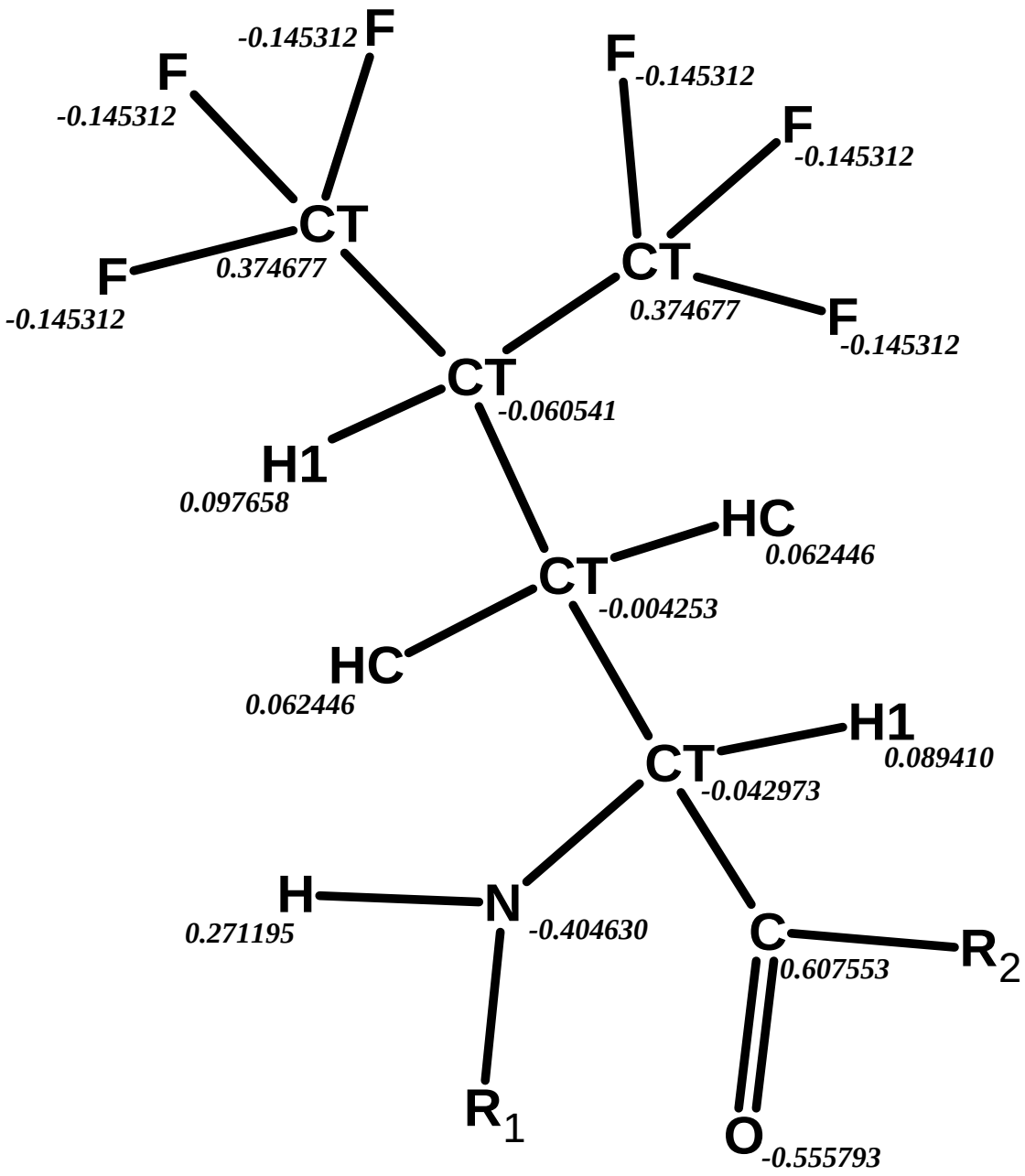
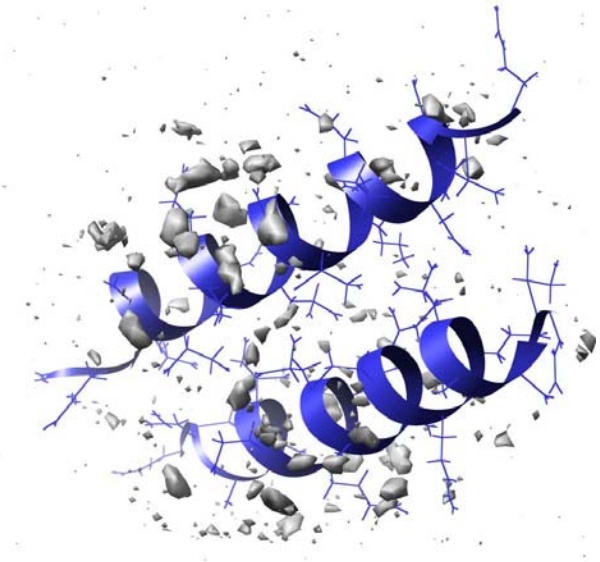


Figure S5 - Hydration sites of selected coiled coil dimers. Water binding sites were calculated using the grid function of ptraj over a 5 ns trajectory. Hydration sites shown were contoured at 2.87 times bulk water density using the volume visualization module of Chimera[5].

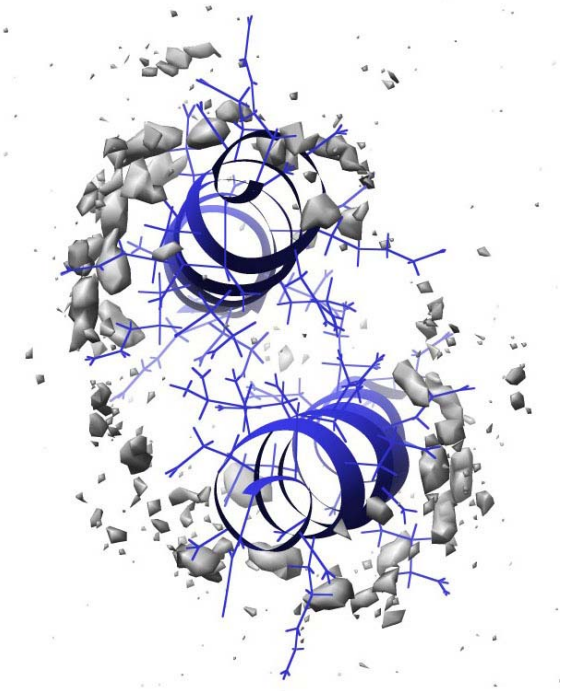
IAAL-E3/K3

Monomer A	EIAALEKEIAALEKEIAALEK
Monomer B	KIAALKEKIAALKEKIAALKE

Side View



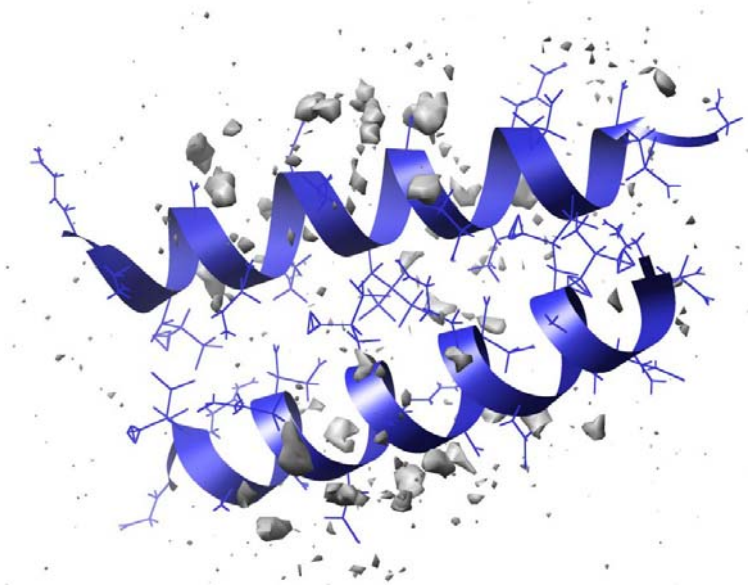
Top-Down View



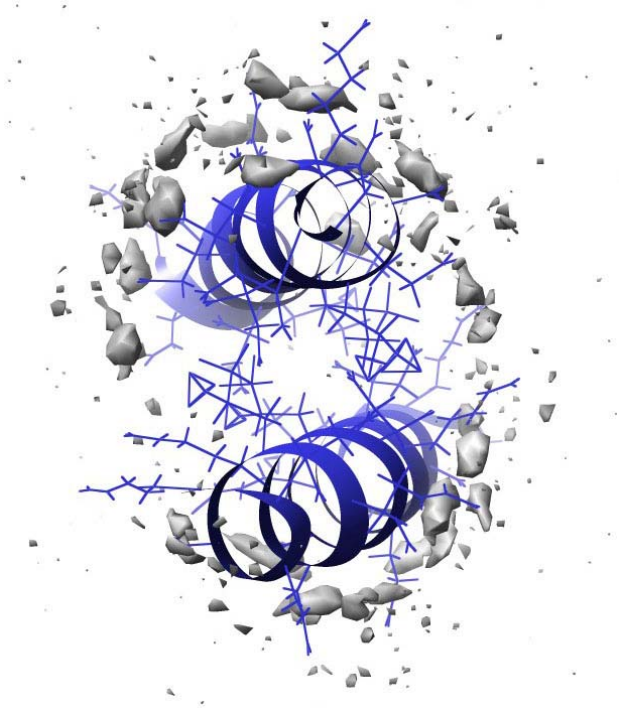
IAA(hFLeu)-E3/K3

Monomer A EIAAĹEKEIAAĹEKEIAAĹEK
Monomer B KIAAĹKEKIAAĹKEKIAAĹKE

Side View



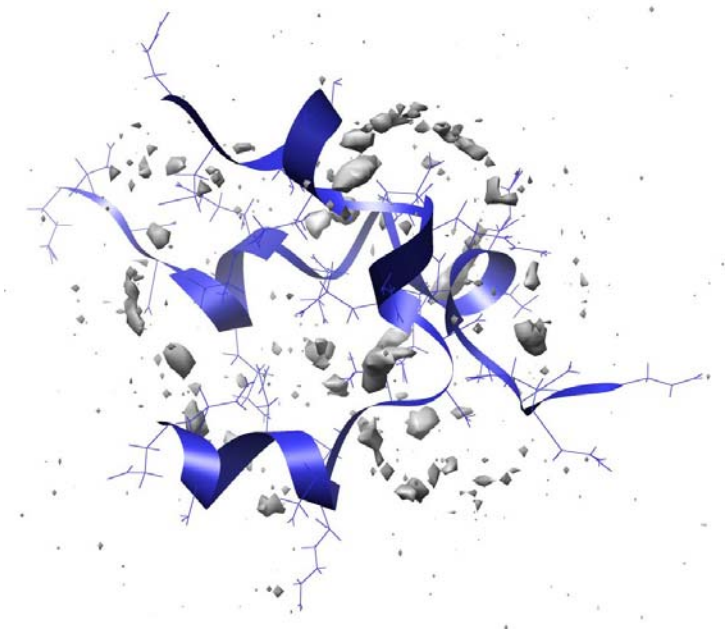
Top Down View



IAAL-E3/O3

Monomer A	EIAALEKEIAALEKEIAALEK
Monomer B	OIAALOEOIAALOEOIAALOE

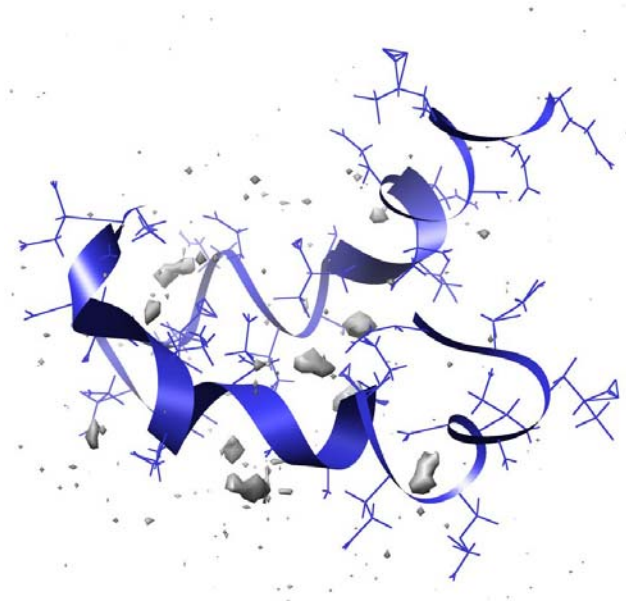
Side View



IAA(hFLeu)-E3/O3

Monomer A	EIAAĹEKEIAAĹEKEIAAĹEK
Monomer B	OIAAĹOEOIAAĹOEOIAAĹOE

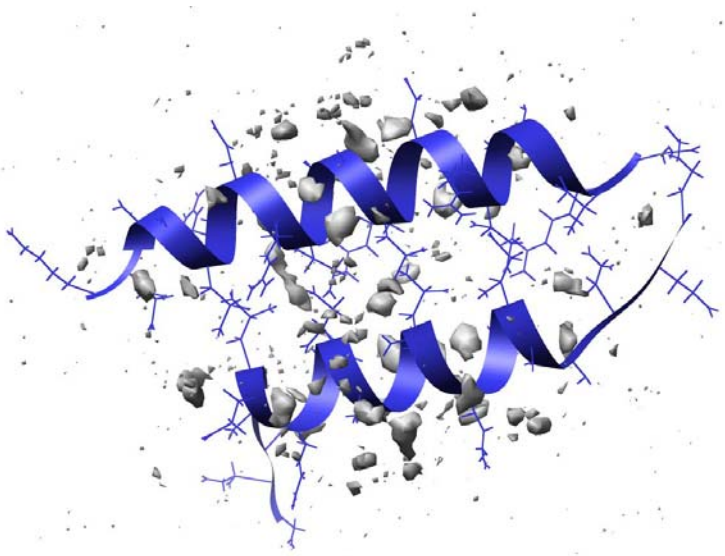
Side View



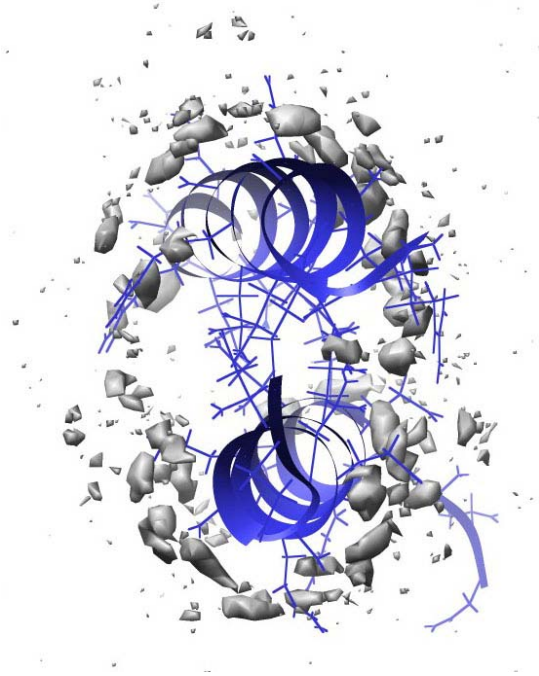
IAAL-E3/R3

Monomer A	EIAALEKEIAALEKEIAALEK
Monomer B	RIAALRERIAALRERIAALRE

Side View



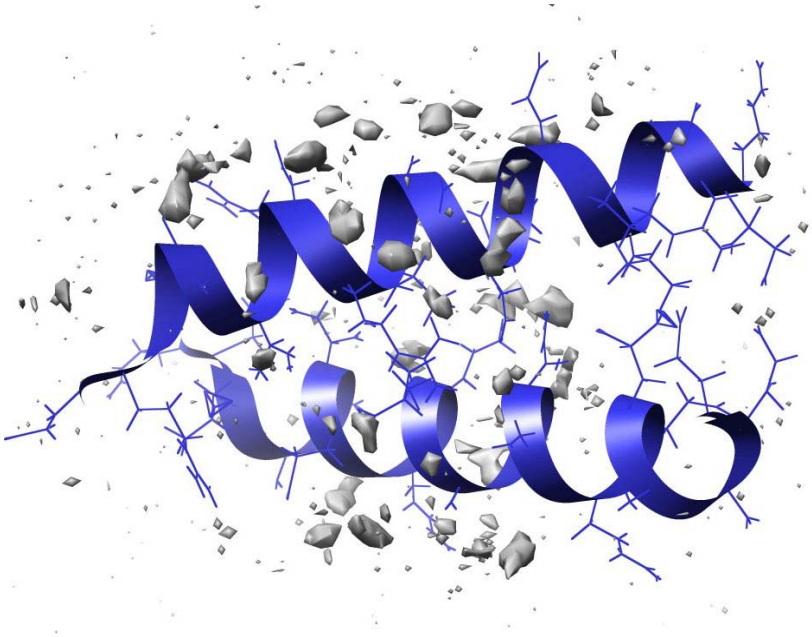
Top Down View



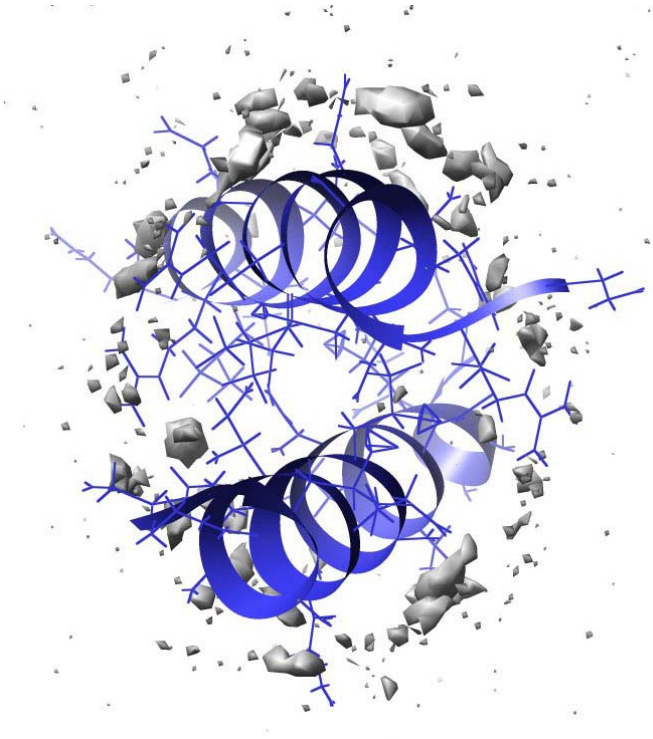
IAA(hFLeu)-E3/R3

Monomer A	EIAAĹEKEIAAĹEKEIAAĹEK
Monomer B	RIAAĹRERIAAĹRERIAAĹRE

Side View



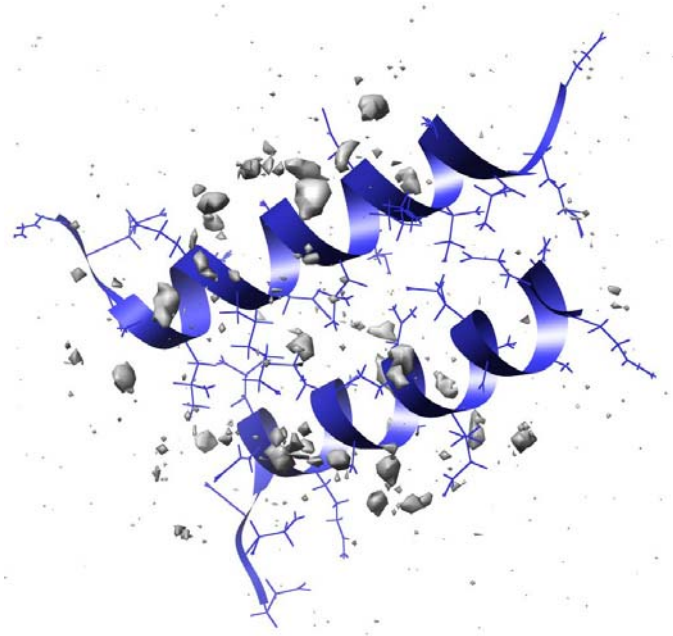
Top Down View



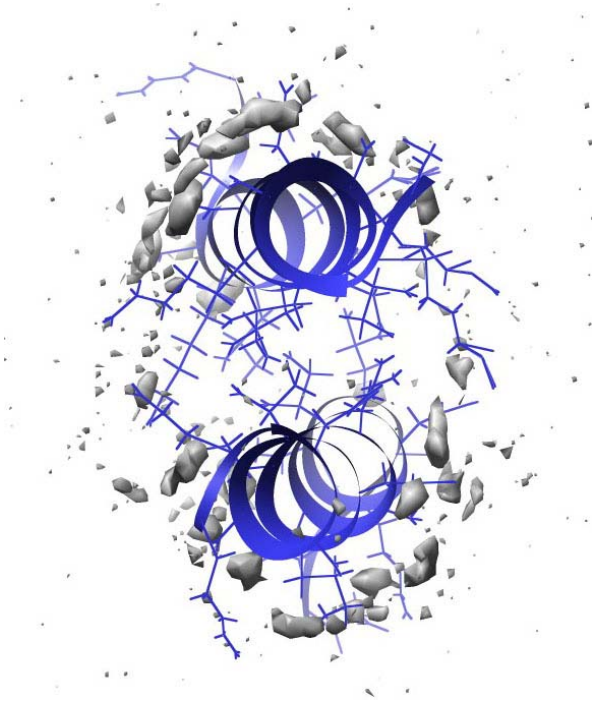
IAAL-E3/Ĥ3

Monomer A	EIAALEKEIAALEKEIAALEK
Monomer B	ĤIAALĤEĤIAALĤEĤIAALĤE

Side View



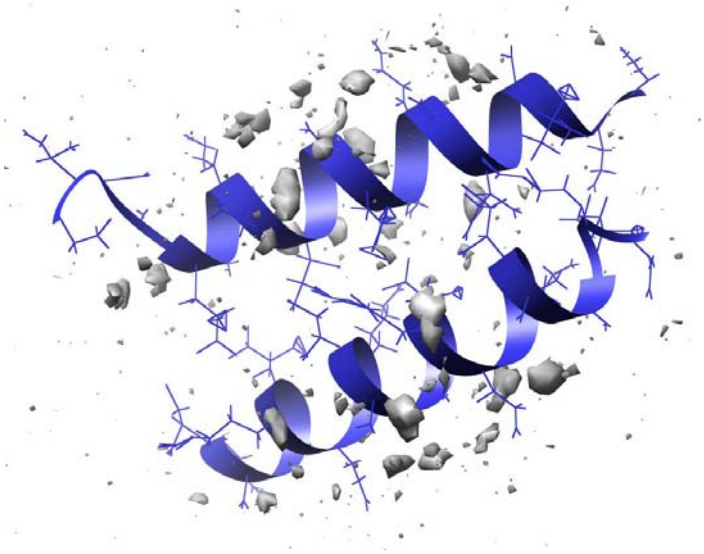
Top Down View



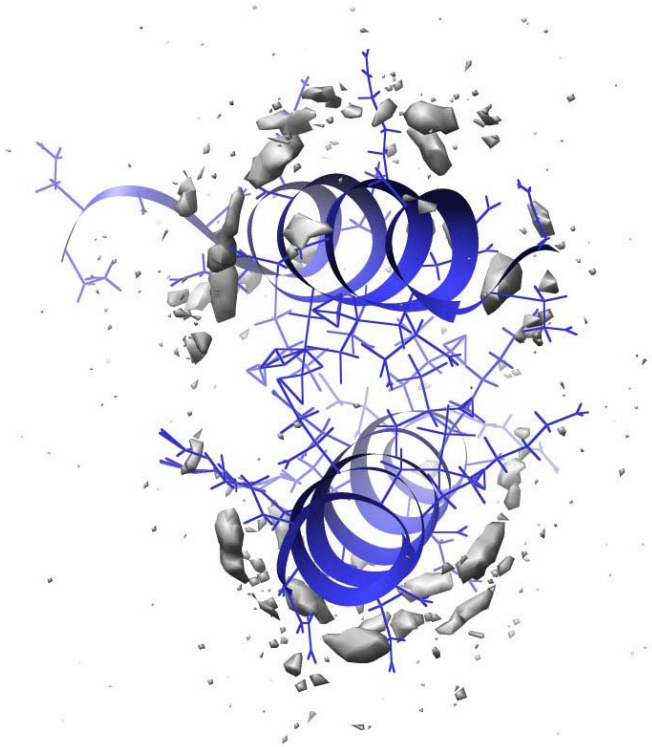
IAA(hFLeu)-E3/Ĥ3

Monomer A	EIAAĹEKEIAAĹEKEIAAĹEK
Monomer B	ĤIAAĹĤEĤIAAĹĤEĤIAAĹĤE

Side View



Top Down View



O = ornithine
Ĺ = 5,5,5,5’,5’,5’-hexafluoroleucine
Ĥ = (S)-2,7-diaminoheptanoic acid

Table S1 – Detailed calculation of relative free energy of hydration as measured by thermodynamic integration of leucine to 5,5,5,5',5',5'-hexafluoroleucine in gas and aqueous environments.

Thermodynamic Integration - Leucine to HFL Transition

Water

	λ	∂V/∂λ	weight	A	Stdev-∂V/∂λ	Stdev-A
1	0.00922	1075.069	0.02359	25.36088	6.897373	0.162709
2	0.04794	476.9183	0.05347	25.50082	5.168323	0.27635
3	0.11505	0.338758	0.08004	0.027114	3.050577	0.244168
4	0.20634	-156.939	0.10158	-15.9419	1.232714	0.125219
5	0.31608	-119.702	0.11675	-13.9752	0.724443	0.084579
6	0.43738	-53.1574	0.12457	-6.62182	0.192584	0.02399
7	0.56262	-15.8672	0.12457	-1.97658	0.061205	0.007624
8	0.68392	-3.16454	0.11675	-0.36946	0.01208	0.00141
9	0.79366	-0.37622	0.10158	-0.03822	0.00122	0.000124
10	0.88495	-0.02025	0.08004	-0.00162	8.76E-05	7.01E-06
11	0.95206	-0.0003	0.05347	-1.6E-05	0	0
12	0.99078	0	0.02359	0	0	0
				11.96403		***0.431197***

Gas Phase

	λ	∂V/∂λ	weight	A	Stdev-∂V/∂λ	Stdev-A
1	0.00922	1046.252	0.02359	24.68108	2.081009	0.049091
2	0.04794	466.5309	0.05347	24.94541	1.661893	0.088861
3	0.11505	-2.23157	0.08004	-0.17861	1.778744	0.142371
4	0.20634	-155.164	0.10158	-15.7616	1.477358	0.15007
5	0.31608	-118.813	0.11675	-13.8714	0.75396	0.088025
6	0.43738	-52.6247	0.12457	-6.55546	0.34311	0.042741
7	0.56262	-15.7366	0.12457	-1.96031	0.100182	0.01248
8	0.68392	-3.13473	0.11675	-0.36598	0.020711	0.002418
9	0.79366	-0.37241	0.10158	-0.03783	0.002498	0.000254
10	0.88495	-0.02009	0.08004	-0.00161	0.000129	1.03E-05
11	0.95206	-0.00028	0.05347	-1.5E-05	3.77E-05	2.01E-06
12	0.99078	0	0.02359	0	0	0
				10.8937		***0.250666***

Relative Free Energy of Hydration ***1.070333*** ***Stdev*** ***0.350852***

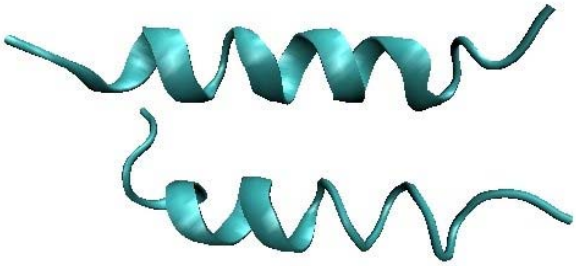
Table S2 – Leucine rotamers in the d position of published coiled coil dimers. Shown below are the measured leucine rotamers for both coiled coil monomers of the leucine sidechain atoms. All structures were taken from the RCSB Protein Data Bank. Please refer to tables 3 and 4 of the main article for rotamer definitions.

Residue	Dihedral 1	Dihedral2	Dihedral 1	Dihedral 2	Rotamer 1	Rotamer 2
1U0I (Model 1)						
5	-51.19	163.48	-51.19	163.48	gauche -	trans
12	-47.47	117.96	-47.47	117.96	gauche -	gauche +
19	113.66	86.09	113.66	86.09	gauche+	gauche +
5	-78.41	-63.49	-78.41	-63.49	gauche -	gauche -
12	-42.54	-175.79	-42.54	184.21	gauche -	trans
19	-48.06	131.59	-48.06	131.59	gauche -	trans
1U0I (Model 2)						
5	151	75	151	75	trans	gauche +
12	170.95	-81.97	170.95	-81.97	trans	gauche -
19	-170.71	-77.27	189.29	-77.27	trans	gauche -
5	104.53	75.14	104.53	75.14	gauche +	gauche +
12	112	87.09	112	87.09	gauche +	gauche +
19	139.21	71.95	139.21	71.95	trans	gauche +
1A93						
13	-83.86	163.46	-83.86	163.46	gauche -	trans
20	-89.86	153.88	-89.86	153.88	gauche -	trans
27	-69.95	163.01	-69.95	163.01	gauche -	trans
20	-72.9	154.15	-72.9	154.15	gauche -	trans
27	-79.48	161.09	-79.48	161.09	gauche -	trans
1P9I						
4	-135.56	-24.2	224.44	-24.2	trans	gauche-
11	-77.3	167.76	-77.3	167.76	gauche -	trans
18	-71.26	166.7	-71.26	166.7	gauche -	trans
25	-67	171.45	-67	171.45	gauche -	trans
1R48 (Multiple Models)						
20	-44.93	-168.48	-44.93	191.52	gauche -	trans
27	-90.8	141.59	-90.8	141.59	gauche -	trans
20	-46.01	-163.46	-46.01	196.54	gauche -	trans
27	-89.44	132.63	-89.44	132.63	gauche -	trans
2A93 (Multiple Models)						
13	-92.82	168.09	-92.82	168.09	gauche -	trans
20	-90.47	151.34	-90.47	151.34	gauche -	trans
27	-72.31	157.53	-72.31	157.53	gauche -	trans
34	-89.41	172.26	-89.41	172.26	gauche -	trans
20	-75.6	162.8	-75.6	162.8	gauche -	trans
27	-88.34	158.62	-88.34	158.62	gauche -	trans
34	-88.98	163.09	-88.98	163.09	gauche -	trans

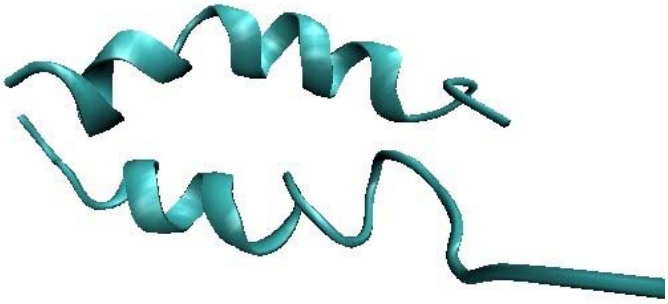
2ZTA						
	5	-74.57	157.4	-74.57	157.4	gauche - trans
	12	-73.06	162.89	-73.06	162.89	gauche - trans
	19	-66.58	160.65	-66.58	160.65	gauche - trans
	26	-59.8	137.2	-59.8	137.2	gauche - trans
	5	-70.22	169.44	-70.22	169.44	gauche - trans
	12	-74.38	174.09	-74.38	174.09	gauche - trans
	19	-71.07	171.55	-71.07	171.55	gauche - trans
	26	-75.21	-177.21	-75.21	182.79	gauche - trans
1KDD						
	5	-63.26	-179.22	-63.26	180.78	gauche - trans
	12	-70.54	174.59	-70.54	174.59	gauche - trans
	19	-74.22	-172.2	-74.22	187.8	gauche - trans
	26	-69.78	171.76	-69.78	171.76	gauche - trans
	5	-70.35	-176.41	-70.35	183.59	gauche - trans
	12	-65.56	178.96	-65.56	178.96	gauche - trans
	19	-77.68	166.46	-77.68	166.46	gauche - trans
	26	-71.3	175.8	-71.3	175.8	gauche - trans
1KD8						
	5	-84.39	-65.33	-84.39	-65.33	gauche - gauche -
	19	-69.85	-169.22	-69.85	190.78	gauche - trans
	26	-73.81	167.72	-73.81	167.72	gauche - trans
	5	-68.76	168.4	-68.76	168.4	gauche - trans
	12	-73.7	176.62	-73.7	176.62	gauche - trans
	19	-72.25	162.18	-72.25	162.18	gauche - trans
	26	-67.27	-179.37	-67.27	180.63	gauche - trans
1KD9						
	5	-60.52	-154.52	-60.52	205.48	gauche - trans
	12	-62.84	171.24	-62.84	171.24	gauche - trans
	19	-70.15	158.01	-70.15	158.01	gauche - trans
	26	-71.64	165.62	-71.64	165.62	gauche - trans
	5	-75.65	-174.53	-75.65	185.47	gauche - trans
	12	-67.07	170.37	-67.07	170.37	gauche - trans
	19	-77.11	172.6	-77.11	172.6	gauche - trans
	26	-65.89	168.68	-65.89	168.68	gauche - trans

Figure S6 – Degradation of IAAL-E3/O3 secondary structure. Unfolding of IAAL-E3/O3 shown by secondary structure snapshots. Final structure shows significant loss of helical character and deformation of the protein-protein binding domain while association is maintained.

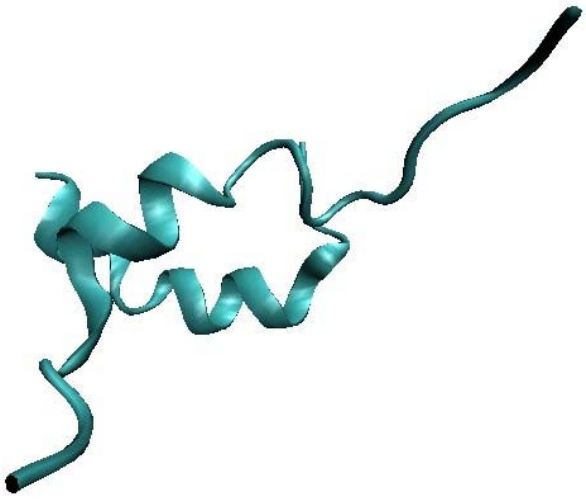
4.47 ns



10.0 ns



14 ns



15 ns

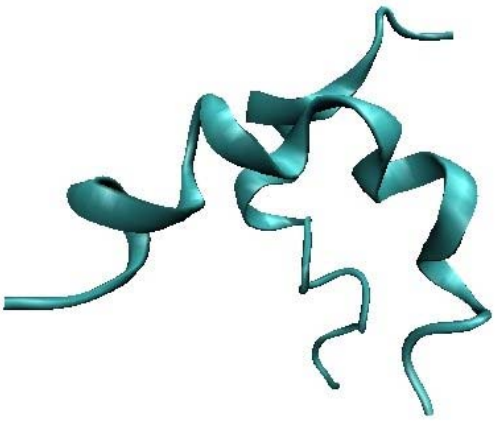


Table S3 – Rotamer contributions to the overall hydrophobic sidechain conformational entropy. Rotamer contributions calculated from pooled fluorinated and nonfluorinated coiled coils using equation 1 of the article.

Isoleucine Rotamers		**Contribution (%)**	**Entropy (cal/mol K)**	**Entropy (kcal/mol)**
Nonfluorinated				
gauche+, trans	gauche-, trans	29.74	0.31	0.09
gauche-, trans	gauche-, trans	2.61	0.08	0.02
gauche-, gauche-	gauche+, trans	0.98	0.04	0.01
trans, trans	trans, trans	47.54	0.31	0.09
gauche-, trans	trans, trans	9.96	0.20	0.06
trans, gauche+	trans, trans	5.83	0.14	0.04
gauche-, trans	trans, gauche+	2.10	0.07	0.02
trans, gauche+	trans, gauche+	0.71	0.03	0.01
trans, trans	gauche-, gauche-	0.53	0.02	0.01
Total				**0.36**
Fluorinated				
gauche-, trans	gauche+, trans	25.13	0.30	0.09
gauche-, gauche-	gauche+, trans	8.20	0.18	0.05
gauche-, trans	trans, trans	38.48	0.32	0.09
trans, gauche+	gauche-, trans	15.29	0.25	0.07
gauche-, gauche-	trans, trans	0.70	0.03	0.01
trans, trans	trans, trans	2.05	0.07	0.02
gauche-, gauche+	trans, trans	0.60	0.03	0.01
trans, gauche+	gauche-, gauche-	6.41	0.15	0.05
trans, trans	gauche-, trans	3.09	0.09	0.03
Total				**0.42**
Leucine Rotamers		**Contribution (%)**	**Entropy (cal/mol K)**	**Entropy (kcal/mol)**
Nonfluorinated				
gauche-, trans	gauche-, trans	31.90	0.31	0.09
trans, gauche+	trans, gauche+	32.80	0.32	0.09
trans, gauche+	gauche-, trans	26.27	0.30	0.09
trans, gauche+	gauche-, gauche+	5.27	0.13	0.04
trans, gauche+	gauche-, gauche-	1.40	0.05	0.02
gauche-, trans	gauche-, trans	0.97	0.04	0.01
Total				**0.34**
Fluorinated				
gauche-, trans	gauche-, trans	96.77	0.03	0.01
trans, gauche+	gauche-, trans	3.09	0.09	0.03
Total				**0.04**

Figure S7 – Leucine χ_1 and χ_2 angles measured in the IAAL-E3/R3 as shown in a time course. Measurements reported below were taken from residue 12 of the central heptad of both monomers. Time course shown as simulation time in ps (x axis) against torsion angle in degrees (y axis)

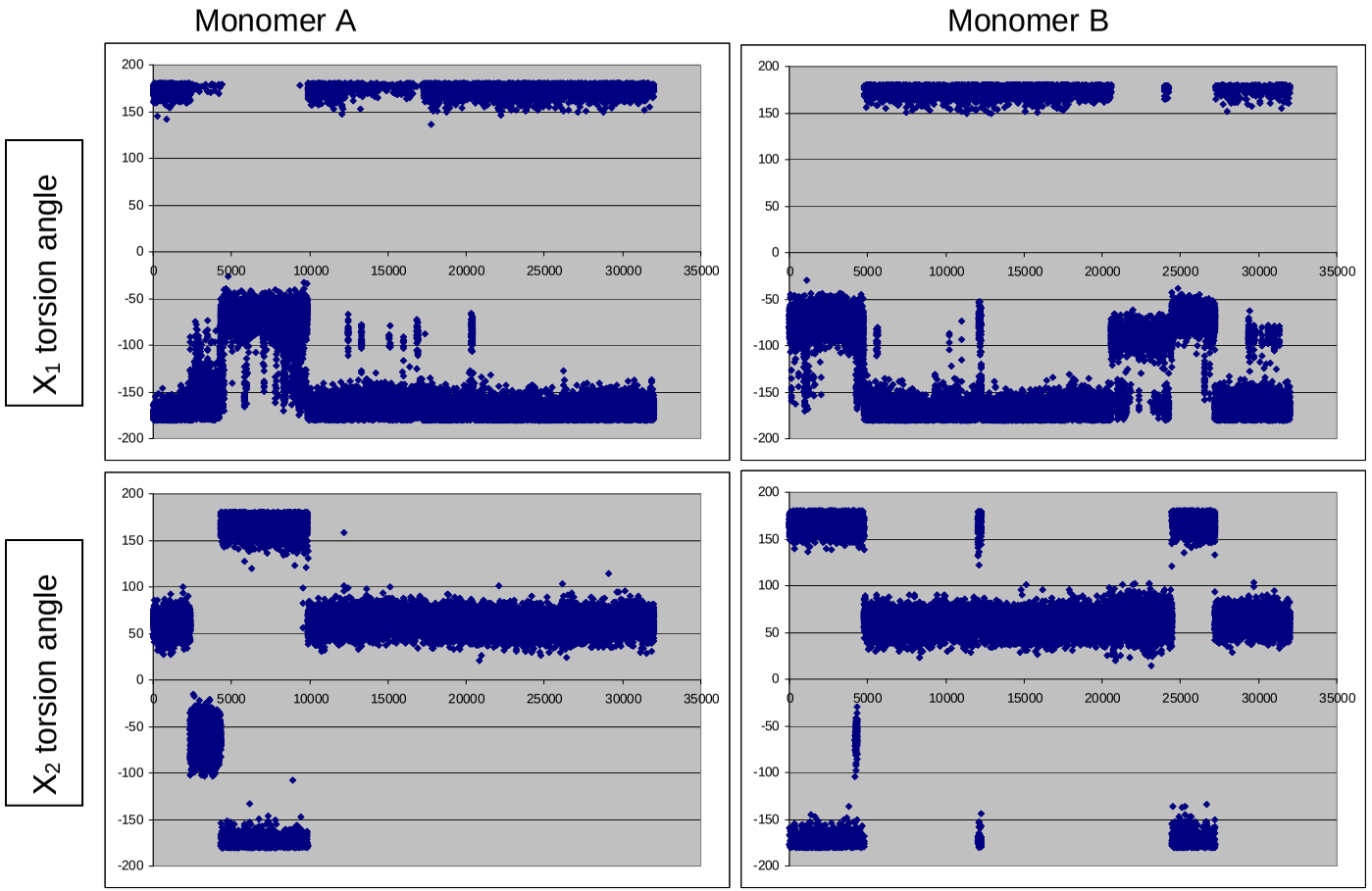


Figure S8 – Isoleucine χ_1 and χ_2 angles measured in the IAAL-E3/R3 as shown in a time course. Measurements reported below were taken from residue 9 of the central heptad of both monomers. Time course shown as simulation time in ps (x axis) against torsion angle in degrees (y axis).

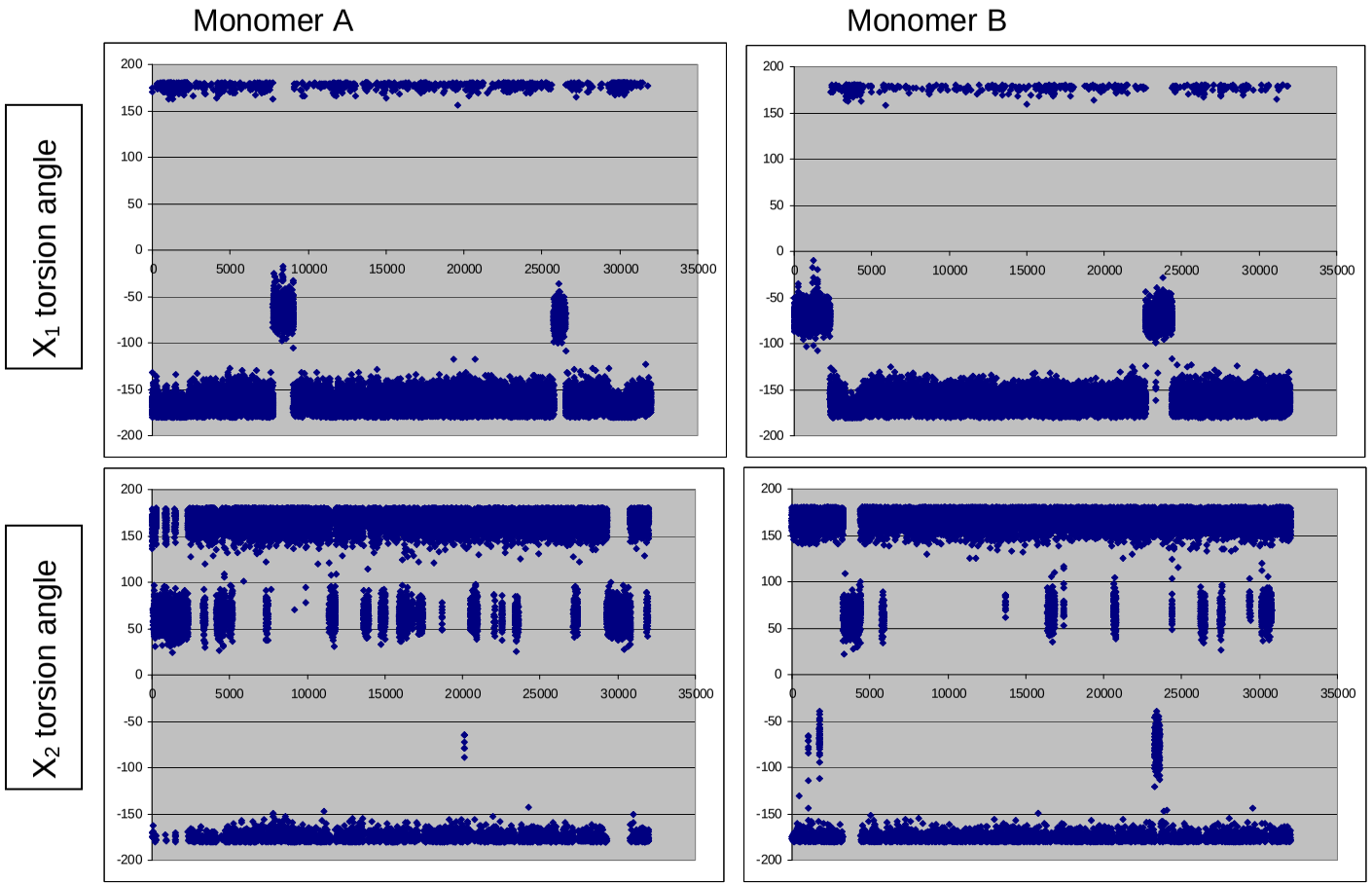


Figure S9 – Salt bridge distance time course for fluorinated and non-fluorinated coiled coil dimer pairs. Salt bridge distance in angstroms (y axis) is plotted against simulation time measure in ps (x axis). In panel A, the time progression is shown for IAAL-E3/K3 (in black) and IAA(hFLeu)-E3/K3 in red. In panel B, the time progression is shown for IAAL-E3/R3 (in black) and IAA(hFLeu)-E3/R3 in red. In panel C, the time progression is shown for IAAL-E3/Ĥ3 (in black) and IAA(hFLeu)-E3/Ĥ3 in red.

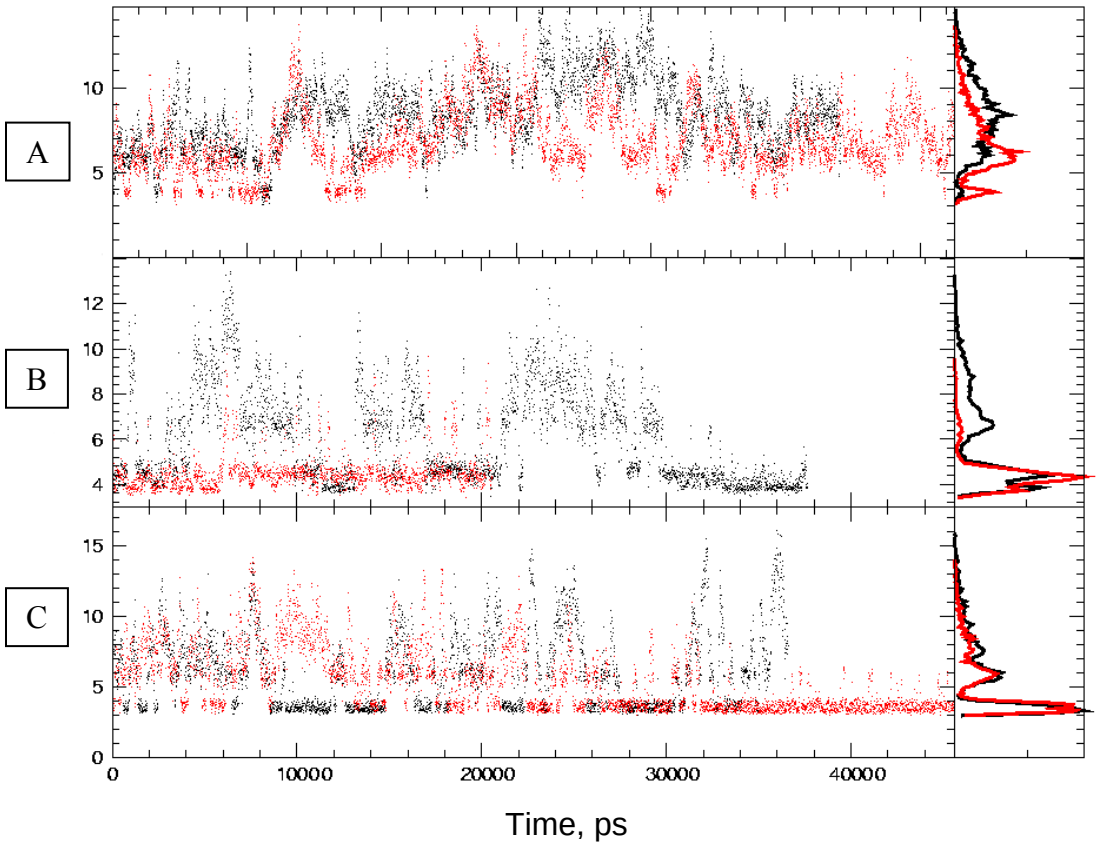


Table S4 – Pairwise free energy decomposition for interactions composing the Arg13, Glu29, and hFLeu33 stabilizing triad. Units in kcal/mol.

Coiled Coil Dimer	**Snapshot**	**Interaction**	**VDW**	**EEL**	
IAAL-E3/R3	1	13->33	-0.612	-0.696	
	2	13->33	-0.342	-0.510	
	3	13->33	-0.444	-0.994	
	4	13->33	-0.728	-0.992	
	5	13->33	-0.605	-0.555	
	6	13->33	-0.421	-1.164	
	7	13->33	-0.362	-0.656	
	8	13->33	-0.422	-1.050	
	9	13->33	-0.898	-0.879	
	10	13->33	-0.601	-0.781	
			-0.544	***-0.828***	
	1	29->33	-0.608	-0.245	
	2	29->33	-0.416	-1.109	
	3	29->33	-1.460	0.634	
	4	29->33	-0.619	-0.348	
	5	29->33	-0.671	-0.483	
	6	29->33	-0.647	-0.219	
	7	29->33	-1.184	-1.076	
	8	29->33	-0.739	-0.267	
	9	29->33	-1.666	-0.571	
	10	29->33	-1.577	0.111	
			-0.959	***-0.357***	
IAA(hFLeu)-E3/R3	1	13->33	-0.684	0.733	
	2	13->33	-0.394	0.698	
	3	13->33	-0.162	0.473	
	4	13->33	-0.639	0.988	
	5	13->33	-0.544	0.567	
	6	13->33	-0.367	0.328	
	7	13->33	-0.910	0.438	
	8	13->33	-0.476	0.528	
	9	13->33	-0.512	0.612	
	10	13->33	-0.947	0.868	
			-0.564	***0.623***	
	1	29->33	-2.019	-3.234	
	2	29->33	-1.799	-2.875	
	3	29->33	-1.038	-3.688	
	4	29->33	-1.734	-3.044	
	5	29->33	-0.701	-3.828	
	6	29->33	-1.052	-3.071	
	7	29->33	-1.116	-2.489	
	8	29->33	-1.702	-2.796	
	9	29->33	-1.591	-2.853	
	10	29->33	-1.504	-3.280	
			-1.426	***-3.116***	
Fluorination Differences		13->33	-0.020	1.451	
		29->33	-0.467	-2.759	
		Combined	-0.487	-1.308	***-1.794***

Sources cited in supplemental materials

1. Wang, J., P. Cieplak, and P.A. Kollman, *How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?* J. Comput. Chem., 2000. **21**(12): p. 1049-1074.

2. Simmerling, C., B. Strockbine, and A.E. Roitberg, *All-atom structure prediction and folding simulations of a stable protein.* J. Am. Chem. Soc., 2002. **124**(38): p. 11258-9.

3. Hornak, V., et al., *Comparison of multiple Amber force fields and development of improved protein backbone parameters.* Proteins, 2006. **65**(3): p. 712-25.

4. Duan, Y., et al., *A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations.* J. Comput. Chem., 2003. **24**(16): p. 1999-2012.

5. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J. Comput. Chem., 2004. **25**(13): p. 1605-12.

CHAPTER 3

COMPUTATIONAL GUIDED DESIGN OF AN APOPTOSIS INDUCING BCR MUTANT FOR THE TREATMENT OF CHRONIC MYELOGENOUS LEUKEMIA

Note: This chapter describes a collaborative study between the Cheatham and Lim research groups at the University of Utah. Computational approaches and molecular dynamic simulations were completed by Scott Pendley from the Cheatham group. Gene cloning, translocation assays, microscopy, phosporylation and activity assays, and experimental approaches were completed by the Lim laboratory. The choice of specific mutations and protein engineering approaches used in the design of the Bcr mutant were a combined effort. The design of a disulfide bond, incorporation of alanine in the backbone of the coiled-coil dimer, the C38A mutation, and the separate consideration of the Asp mutations to the hydrophobic core were ideas proposed by Scott Pendley.

Abstract

Chronic myelogenous leukemia (CML) is a condition resulting from the reciprocal translocation of chromosome 22 onto chromosome 9 to form the bcr-abl oncogene. The protein translation of this gene, Bcr-Abl, is responsible for a massive clonal expansion of hematopoietic progenitor cells and the resulting cancer. The current treatment using Gleevec® is effective; however, resistance can develop and it is only moderately effective in the treatment of the disease in later stages. We present in this article our efforts to use computational and experimental approaches to develop a mutant Bcr protein capable of binding with the oncoprotein and inducing apoptosis in CML cells. Molecular dynamics simulations were used to predict and evaluate key mutations to improve coiled-coil binding at the protein interface and specifity for heterodimer binding with the Bcr-Abl oncoprotein. Improvements to protein binding focused

primarily on mutations to improve ionic interactions across the salt bridges. Rational design approaches to improve specificity were tested to determine energetic and structural consequences of introducing charged amino acids into the hydrophobic domain. Experimental testing of the final mutant heterodimer, CC-CCmut1, suggests a significant improvement to binding over the wild-type oncoprotein dimer as well as the ability to decrease cell proliferation and induce apoptosis in expressing cells.

Introduction

Chronic myelogenous leukemia (CML) is a cancer that results in the massive clonal expansion of hematopoietic progenitor cells which eventually lose their ability to differentiate (1-3). A characteristic of CML is the Philadelphia (Ph) chromosome which is formed by reciprocal chromosomal translocation of chromosome 22 onto chromosome 9. This chromosomal translocation results in a fused gene composed of the oligomerization domain of the bcr gene from chromosome 22 with the abl gene, which encodes a tyrosine kinase, on chromosome 9 (1, 4, 5). The fused bcr-abl gene is located on the resulting, shorter chromosome 22. Translation of the fused genes results in the oncoprotein Bcr-Abl, a homotetramer tyrosine kinase, which is found in nearly all patients with CML. The N-terminal region of Bcr-Abl is comprised of an oligomerization domain which forms a homotetramer, antiparallel coiled-coil (6, 7). Bcr-Abl association occurs when two monomers bind and exchange N-terminal helices and form an antiparallel homodimer with C-terminal helices (6). Two dimers then associate; stack together to form the homotetramer (see Figure 3.1) (6). Oligomerization of Bcr-Abl results in the constitutive activation of the oncoprotein through auto-phosphorylation of regulatory tyrosines (8, 9). Bcr-Abl is known to activate several downstream signal transduction pathways involved in cell proliferation and apoptosis (10). Bcr-Abl activation of PI3 kinase (through either phosphorylation of Crkl or Gab2) prevents apoptosis by inhibiting caspase activation, decreasing FoxO3 and GSK3β activity, and increasing the phosphorylation of Bad and the cytosolic concentration of anti-apoptotic Bcl-2 family members (11-17). Bcr-Abl activation of Stat5 and Ras contributes to increased nuclear expression of anti-apoptotic Bcl-2 family members MCL-1 and BCL-x_L as well as increased expression of cyclin D, an important molecule in cell cycle regulation (18-23). Bcr-

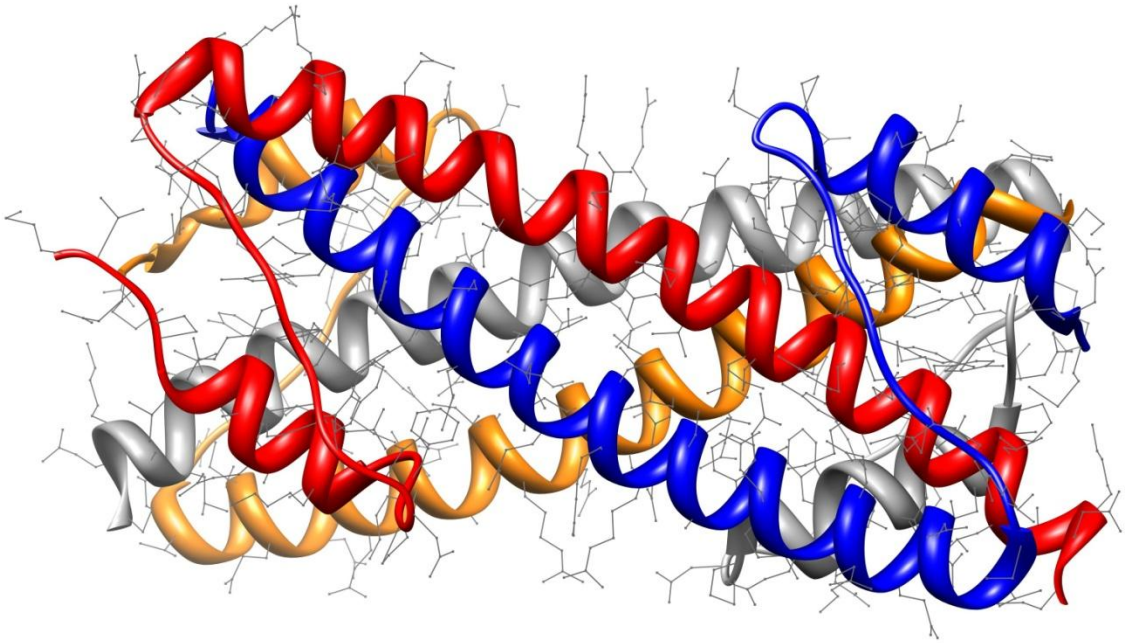


Figure 3.1 - A molecular graphic representation of the anti-parallel, Bcr homotetramer coiled-coil. Secondary structure of the molecule is shown using a ribbon structure while heavy atoms and bonds of the main chain and side-chains are represented using a wire model. Individual chains (monomers) are colored differently to emphasize the protein-protein interactions in the formation of the homotetramer.

Abl transformed cells were found to proliferate even in low serum conditions due to activation of CDK2, an molecule involved in the G_1/S cell-cycle progression, and a decreased nuclear concentration of cyclin dependent kinase inhibitor $p27^{kip1}$ (24). This information taken together indicates that Bcr-Abl activated signaling pathways that lead to cell survival and cell cycle progression contribute to the leukogenic potential of the oncoprotein (11). Further, *in vitro* and animal studies have found that the expression of Bcr-Abl is responsible for the development of CML in patients (25-28).

Treatment of CML usually involves the introduction of Gleevec® (imatinib mesylate), a small molecule tyrosine kinase inhibitor (29, 30). Gleevec® binds in the ATP binding site of Bcr when the kinase is enzymatically inactive (31). Resistance to Gleevec® can develop due to point mutations in the Bcr-Abl protein that develop with the progression of the cancer(31) which either hinder drug binding to the active site or affect the activation loop switch. As a result, Gleevec® is only moderately effective in later stages of the disease (29). Tasigna® (nilotinib), another tyrosine kinase inhibitor, was developed to surmount developed resistance to Gleevec® in patients diagnosed with CML (32). Use of Tasigna® can result in QT prolongation (irregular heartbeat that can lead to fainting, loss of consciousness, seizures, and sudden death) (32) and neither tyrosine kinase inhibitor is curative. Alternative treatments that could replace or be used in conjunction with these kinase inhibitors would significantly improve the treatment of CML and patient outcomes. Recent drug development research has focused on two main targets: the oligomerization domain and the cellular localization of the oncoprotein. The oligomerization domain has been found to be essential for the phenotypic disease transformation (8). Approaches using an exogenous peptide that can bind at the protein-protein domain can suppress the transformed phenotype and increase Gleevec® sensitivity (33). An alternative treatment discovered by Vigneri and Wang found that nuclear importation of cytoplasmic Bcr-Abl resulted in the induction of apoptosis in CML positive cell lines (34).

We are interested in building on prior experimental research (7, 8, 34) to develop a mutant Bcr coiled-coil to facilitate treatment of CML by sequestering Bcr-Abl and induction of apoptosis. The design of the Bcr mutant protein will focus on optimal Bcr-Abl protein binding with

minimal self-association. To be of pharmaceutical interest the mutant protein must be deliverable at nano to picomolar concentration and show competitive or improved binding compared to the wild-type Bcr-Abl binding partner (35) and heterodimer specificity. To accomplish this task we proposed several modifications to the coiled-coil binding domain of Bcr and used computational modeling with molecular dynamics approaches to predict favorable attributes, specificity, and energetics. A final proposed mutant was chosen and cloned into a plasmid vector. Experimental approaches using a nuclear translocation assay of the expressed protein show that the proposed heterodimer does improve binding and specificity over the wild-type coiled-coil dimer.

Methods

Starting Geometries

The initial model used in these computational studies was derived from the A and B chains of the published X-ray crystal structure (refined to 2.2 Å resolution) of the N-terminal oligomerization domain of Bcr-Abl (PDB ID: 1K1F) (6). Selenomethionine groups from that structure were mutated to methionine amino acid residues and position 38 was back-mutated to cysteine to maintain consistency with the original amino acid sequence. Mutations to the amino acid sequence were introduced into the crystal structure model using the Deepview (Swiss PDB Viewer) program (36) and the LEAP module from AMBER 9 (37).

Molecular Dynamics Simulations

All simulations of the proteins were completed using the ff03 force field (38) from the AMBER modeling suite (39). All proteins in this study were solvated by surrounding the compound with at least a 10 Å water layer in all directions within a truncated octahedron using explicit solvent. This amounts to approximately 12,000 TIP3P (40) waters. Explicit Na^+ and Cl^- salt ions using the Aqvist parameters set (41) were added to neutralize the system. Ionization of amino acids reflected the default solvent-exposed pKa states at physiological pH, i.e., charged Arg, Lys, Glu, Gln, Asn, and Asp residues.

Energy minimization was performed for 500 steps first in the system with restrained substrate atoms (50 kcal mol^{-1} $Å^{-2}$) and then in an unrestrained system. Initial minimization was

followed by heating to 300K at constant volume over a period of 10 ps using harmonic restraints of 2 kcal mol^{-1} $Å^{-2}$ on the protein atoms. Subsequent unrestrained equilibration at 300K followed for 500 ps. Bond lengths involving hydrogen atoms were constrained with SHAKE (42, 43) (for simulations involved TIP3P waters) with a geometric tolerance for the constraint of 0.00001 Å during the coordinate resetting. Periodic boundary conditions were applied using the particle mesh Ewald method (PME) with a less than 1 Å charge grid and cubic B-spline interpolation (44).

All of the production molecular dynamics (MD) simulations were performed with a 2 fs time step and a direct space nonbonded cutoff of 10 Å with the pair list of atomic interactions built out to 11 Å and heuristic update of the pair list triggered when any atom moved more than 0.5 Å since the previous update. During production runs, the center of mass translational motion of the entire system was removed after the initial velocity assignments and subsequently every 5000 MD steps. Constant temperature was maintained with weak coupling to a heat bath with a 2 ps time constant (45). Pressure (1 atm) was maintained using isotropic position scaling with Berendsen weak coupling algorithm with a 1.0 ps pressure relaxation time (45). A 40-90 ns production molecular dynamics simulation for each peptide was completed.

Circular Dichroism Calculations

Quantification and comparison of structural helical content was measured by calculating mean residue ellipticities representing the CD spectra of five individual 500 ps average structures spanning the final 5 ns of simulation of each coiled coil dimer using the DichroCalc program (46). A Gaussian curve type was assumed with a bandwidth at half maximum of 12.5 nm and two backbone transitions. For these calculations, the Hirst et al. semiempirical parameter set (47) was used due to its accuracy with helical proteins (46). The calculated values reported show the average mean residue ellipticities at 222 nm and the standard deviation among the structures.

Percent Helicity Calculations

Structural helical content was also calculated based on secondary structure as determined by peptide backbone Ψ and Φ torsions. The secondary structure of an ensemble of molecular dynamics snapshots from the final 10 ns of production simulation were evaluation

using the DSSP method (48) as implemented in UCSF Chimera (49). Percent helicity was calculated by determining the number of residues assigned as an α-helix divided by the total number of residues in the protein.

Intrahelical Hydrogen Bonds

Hydrogen bonds forming between the carbonyl oxygen and the amine hydrogen of residues separated by four residues are characteristic of the α-helix secondary structure. These interactions were calculated for each residue during the final 10 ns of production simulation using a distance calculation between the carbonyl oxygen of residue 'i' and the amine nitrogen of residue 'i+4'. All distances less than 3.5 Å were considered indicative of an existing hydrogen bond. The number of distance measurements less than 3.5 Å were counted every ps and divided by the total number of potential interactions (the number of residue minus 4). The averaged percent value of hydrogen bonds was reported as a further indication of secondary structure.

MM-PBSA

The MM-PBSA energy calculation methodology was applied to estimate the binding free energies of Bcr mutants as implemented in AMBER and as described by Gohlke and Case (50). Separate trajectories for the dimer and individual monomers were modeled. Solvent and ion molecules were stripped from the trajectories and the free energy of binding was calculated by subtracting absolute free energies of the monomers from the dimer.

For the free energy of binding calculations, 250 snapshots were taken spanning the final 5 ns of stable simulation. Normal mode analysis calculations to determine translation, rotational, and vibrational entropies were completed using a subset of 50 snapshots spanning the range of the original 250 snapshot set. Polar contributions to the solvation free energy were calculated using the Poisson-Boltzmann (PB) equation as implemented in Delphi II (51). For these calculations two grid points per Å were used, the solute filled 80% of the grid box, and 5000 finite difference iterations were performed to ensure convergence of the results. Atomic parse radii consistent with prior Amber Delphi parameterization were used (50, 52).

For the nonpolar contribution to the free energy of solvation, the molecular surface area was calculated using the molsurf program implemented in AMBER (39) with a 1.4 Å probe radius(50). The surface tension proportionality constant was set to 0.00542 kcal mol^{-1} $Å^{-2}$, and the free energy of nonpolar solvation for a point solute *b* was set to 0.92 kcal mol^{-1}. This is consistent with the use of Delphi as the PB solver (50).

Entropy contributions were calculated using normal mode analysis. Minimization of each snapshot in the gas phase using the conjugate gradient method with a distance dependent dielectric screened by a dielectric constant of 4 until the RMS of the elements of the gradient vector were less than 10^{-6} kcal mol^{-1} $Å^{-1}$.

Thermodynamic Integration Free Energy Calculations

Calculations of the relative free energy of binding with respect to the wild-type dimer ($\Delta\Delta G_{binding}$) were completed using thermodynamic integration for the CCmut1 dimers. On the basis of a thermodynamic cycle (see Figure 3.2a), the relative free energy of binding can be calculated by "mutating" the original protein ($\lambda = 0$) to incorporate designed amino acid point mutations ($\lambda = 1$) in both the dimer and monomeric states. Incorporation of the five amino acid mutations considered using this approach was accomplished stepwise (see Figure 3.2b). Two steps were required to incorporate all five mutations to form the heterodimer mutant (CC-CCmut1) and an additional two steps to perturb the transition dimer into the homodimer (CCmut1-CCmut1). Similarly, two steps were used to incorporate the five mutations in the unbound monomer.

Transitions from the wild-type dimer to the CCmut1 dimers were accomplished using three stages of thermodynamic integration at each step of the perturbation: the removal of atomic charges for changing atoms, a soft core potential stage for the conversion of atom types, and a stage to add atomic charges to the mutated atoms. All thermodynamic integrations simulations were performed using a particle mesh Ewald (PME) treatment of electrostatics, as previously described, with explicit solvent and neutralizing ions. Equilibrated structures were allowed to relax during 6 ns of molecular dynamics simulation followed by thermodynamic integration sampling for a minimum of 6 ns at each sampling point. Convergence of thermodynamic

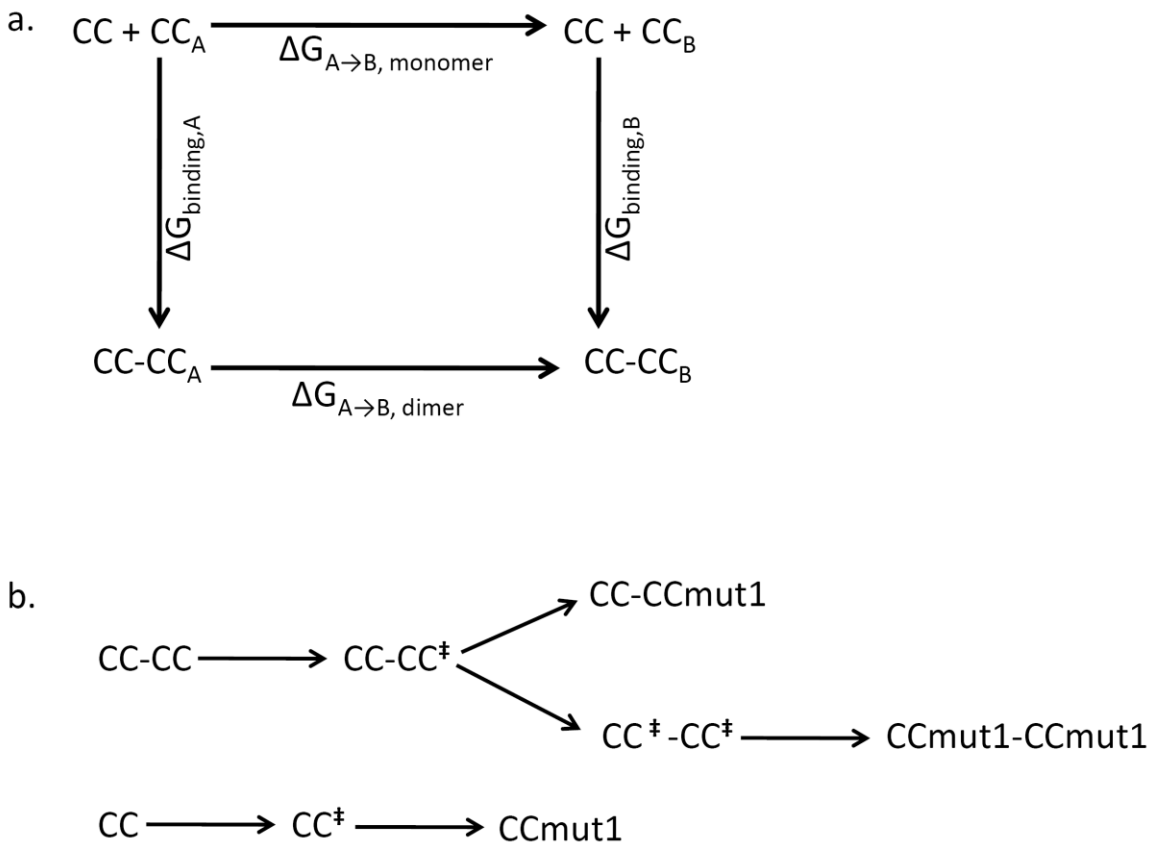


Figure 3.2 - Thermodynamic integration approaches to calculate the relative free energy of binding for coiled-coil dimers were based a thermodynamic cycle (a). Using this cycle the relative free energy of binding ($\Delta\Delta G_{binding}$) can be found using the free energy differences found by incorporating point mutations in the dimer and unbound monomers; $\Delta\Delta G_{binding} = \Delta G_{binding,B} - \Delta G_{binding,A} = \Delta G_{A\rightarrow B,dimer} - \Delta G_{A\rightarrow B,monomer}$. Five point mutations were designed in each monomer using the scheme presented in (b). Note that $CC^{‡}$ represents a transition coiled-coil that incorporates the C38A, S41R, E48R, and Q60E mutations.

integration sampling was visualized by a plateau in the $\partial V/\partial \lambda$ time course. The first step transitions which incorporated four of the five mutations required significant sampling time exceeding 12 ns to affirm that convergence had occurred. For thermodynamic integration calculations involving soft core potentials, 19 linear sampling points of λ between 0.05 and 0.95 were used. Calculations involving only the removal or addition of atomic charges used 10 sampling points of λ between 0.5 and 0.95. Trapezoidal numerical integration was used to sum the free energy of transitions. Comparing this approach against integrating a polynomial fit of the average weighted energetic value against λ resulted in a difference of less than 0.1 kcal/mol value.

Results and Discussion

The design of the mutant Bcr fusion protein focused on mutations that would improve dimer binding while maintaining heterodimer specificity. Computational models evaluated several approaches to increase stability including rational design mutations to stabilize the heterodimer, incorporation of alanine residues at the **f** positions to increase helicity(53), design of a disulfide bond, and the incorporation of mutant-homodimer destabilizing rational design mutations. As an initial quick and somewhat simple estimate of the free energy of binding, helicity was used as a correlate of the relative free energy of binding. In coiled-coils, association is highly correlated with secondary structure formation (54) and the use of helicity is often used as a means of calculating the free energy of binding (54, 55). In the coiled-coil domain of the Bcr protein, some of the secondary structure is preorganized prior to binding (49.8 % α-helical); however, helicity does increase upon oligomerization (66.5% α-helical) (54, 56). Three metrics were used as a measure of secondary structure in the designed mutations. These include a calculation of the circular dichroism (Θ_{222}), percent helicity as calculated from secondary structure using Ψ and Φ torsions, and the percentage of α-helical specific hydrogen bonds that are formed in the dimer (see Table 3.1).

The initial rational design approach focused on stabilizing the dimer by improving salt bridge interactions (see Figure 3.3). Three mutations were designed to the wild-type monomer and comprise a serine to arginine mutation at position 41 (S41R), a glutamine to glutamate

Table 3.1 - Overview of comparative helicity measurements for various mutations of the Bcr coiled-coil domain. Helicity of simulated coiled-coil dimers using the calculated circular dichroism of the peptide, the percent of the residue which is defined as α-helical according to Ψ and Φ backbone dihedral torsions (using the DSSP method), and the percent of α-helical specific hydrogen bonds formed between i and i+4 residues.

Mutations	**Circular Dichroism**		**Secondary Structure**		**Hydrogen Bonds**	
	$[\Theta]_{222}$	$[\Theta]_{222}$ Stdev	Helicity (%)	Helicity Stdev (%)	I,I+4 hb (%)	I,I+4 Stdev (%)
Homodimer Wildtype	**-20943.36**	857.82	**66.55**	1.70	**56.55**	2.71
Heterodimer S41R,Q60E,E48R	**-21580.20**	671.73	**70.01**	2.00	**56.90**	2.75
Homodimer S41R,Q60E,E48R	**-20961.25**	678.30	**66.49**	2.86	**59.38**	2.68
Heterodimer Q33A,Q47A,F54A,T61A	**-20884.41**	476.53	**66.28**	2.49	**60.57**	2.53
Homodimer Q33A,Q47A,F54A,T61A	**-20786.79**	760.39	**65.07**	2.24	**56.39**	3.22
Heterodimer E52C	**-20521.39**	795.64	**67.77**	2.16	**55.90**	2.67
Heterodimer S41R,Q60E,E48R, L45D	**-20330.04**	1329.00	**66.82**	1.87	**52.92**	2.98
Homodimer S41R,Q60E,E48R, L45D	**-19397.75**	679.03	**66.21**	1.95	**56.48**	2.41
Heterodimer S41R,Q60E,E48R, L45D, V49D	**-19695.85**	764.71	**67.84**	1.99	**56.90**	2.65
Homodimer S41R,Q60E,E48R, L45D, V49D	**-20352.10**	671.48	**69.10**	2.53	**59.22**	2.80
Heterodimer S41R,Q60E,E48R,C38A	**-21493.33**	89.72	**69.95**	2.39	**61.50**	2.48
Homodimer S41R,Q60E,E48R,C38A	**-21462.52**	264.90	**61.47**	2.68	**61.42**	2.52
Heterodimer S41R,Q60E,E48R, L45D,C38A	**-21186.16**	695.23	**67.71**	2.39	**59.06**	2.84
Homodimer S41R,Q60E,E48R, L45D,C38A	**-17963.68**	507.00	**59.29**	2.56	**50.71**	2.43
Heterodimer S41R,Q60E,E48R, L45D, V49D,C38A	**-19651.33**	455.70	**64.18**	2.52	**54.02**	2.52
Homodimer S41R,Q60E,E48R, L45D, V49D,C38A	**-16770.20**	311.90	**64.18**	1.77	**50.44**	3.27

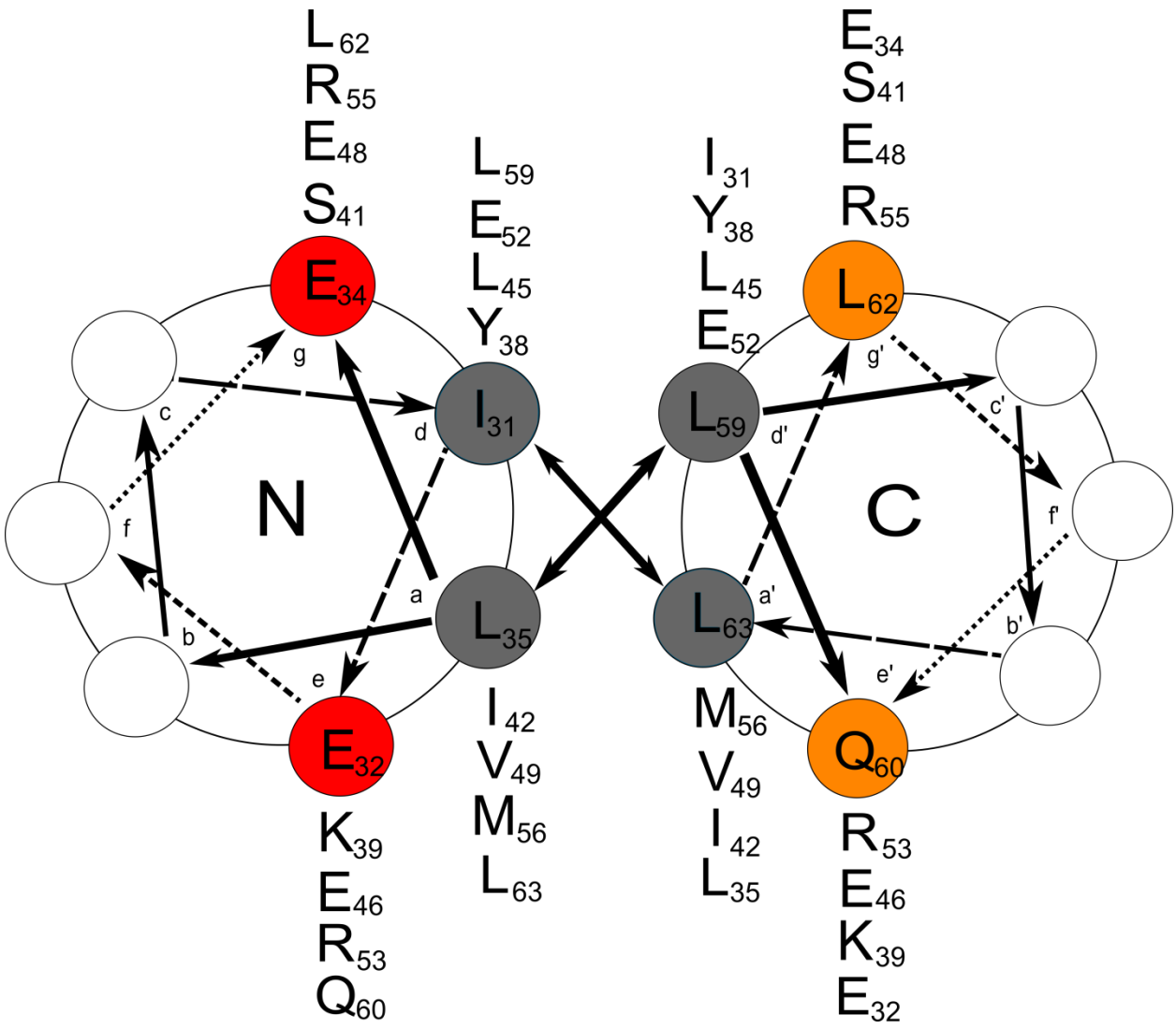


Figure 3.3 - A heptad helical wheel representation showing interacting amino acids in the hydrophobic domain and residues contributing to inter-helical salt bridges. The primary structure of the coiled-coil is denoted **abcdefg** where residues **e** and **g** are typically charged and resides **a** and **d** comprise the hydrophobic domain. Residues in position **g** form a salt bridge with the **g** position of the opposing coil (and **e** with **e**). The crossed arrows at the center denote hydrophobic interactions.

mutation at position 60 (Q60E), and a glutamate to arginine mutation at position 48 (E48R). These mutations would favor the heterodimer over both the homodimer wild-type and homodimer mutant by increasing the favorable electrostatic interactions (E48-R41, S41-R48, and K39-E60) in the heterodimer and causing charge repulsion (R41-R48) interactions in the homodimer mutant. Analysis of the secondary structure metrics suggests that these mutations increase helicity, indicative of a higher binding affinity. The heterodimer shows improved helicity over both the homodimer wild-type and the homodimer mutant. Unfortunately, the homodimer mutant was found to have a higher helicity than the homodimer wild-type which may decrease the specificity for the heterodimer form.

Other approaches to increase the α-helicity included the incorporation of alanine residues in the peptide backbone of the C-terminal coiled-coil region and the design of a disulfide bond to improve binding kinetics. The amino acid residue alanine has the highest helical propensity of all amino acids (53). Alanine mutations were designed at residues Gln33, Gln47, Phe54, and Thr61. While the helicity metrics (α-helical specific hydrogen bonds) suggest an increase of secondary structure, comparison of helicity as measured by secondary structure between the homodimer and heterodimer suggest that this design actually decreases specificity of the heterodimer over the homodimer and may disrupt local intra-helical hydrogen bonds (see Table 3.1 and Figure 3.4). A disulfide bond was designed by mutating Glu52 to a cysteine residue. Cys38 exists as unbound free thiol in the native protein so we hypothesized that the close approach of this residue to a cysteine at position 52 might allow formation of a disulfide bond that could further stabilize the structure. Visualization of the heterodimer (see Figure 3.5) and analysis of the structure helicity suggests that the geometry of the disulfide is not ideal and introduces structural disturbances. Further, MM-PBSA analysis of this dimer showed poor van der Waals contacts and unusual vibrational entropies suggesting that the engineered disulfide bond decreases the stability of the coiled-coil dimer.

Following these early models, additional mutations were proposed to increase the specificity for the heterodimer. While the earlier rational design models focused on mutations that would stabilize interactions (especially ionic interactions) between the dimers, these new

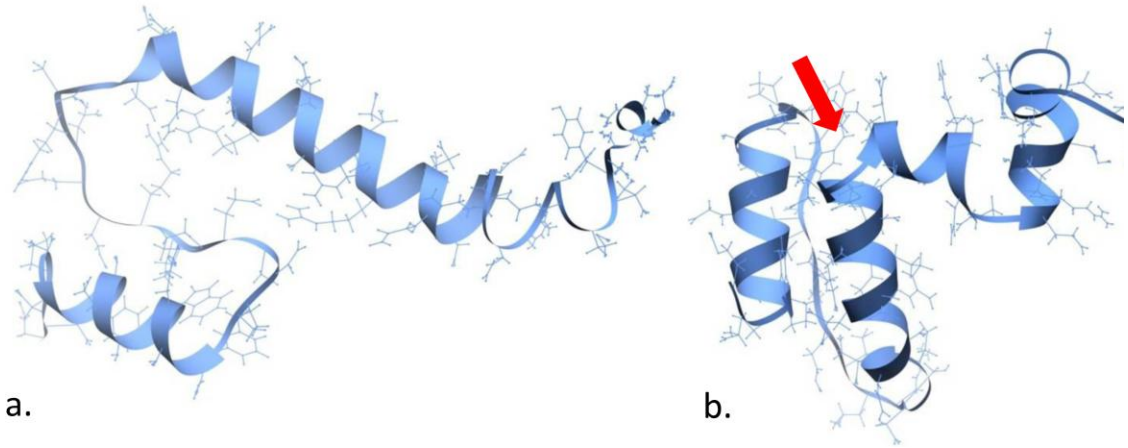


Figure 3.4 - Incorporation of alanine in position **f** of the coiled-coil heptad disrupts intra-helical hydrogen bonds that stabilize the protein in monomeric and dimer form. Shown is a molecular graphic representation of the secondary structure of the Bcr coiled-coil monomer (a) and mutant coiled-coil with alanine modifications (b). A red arrow has been added to indicate the position of the structural kink.

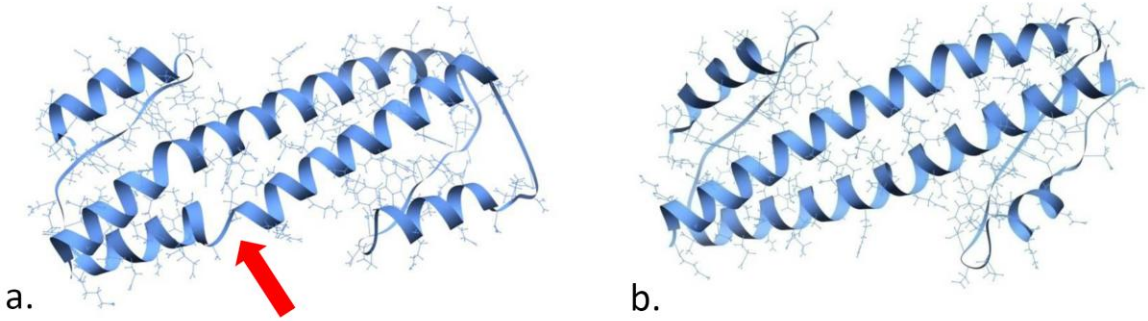


Figure 3-5 - The design of a disulfide bond to improve folding kinetics at position 52 creates a structural disruption due to the geometry of the bound pairs. A molecular graphic representation of the secondary structure is shown with (a) and without (b) the engineered disulfide bond. A red arrow has been added to indicate the position of the disulfide bond.

mutations would introduce charged amino acid residues into the hydrophobic domain that would destabilize the homodimer mutant and hopefully to a lesser extent destabilize the heterodimer. Two mutations were proposed: a leucine to aspartate mutation at position 45 (L45D) and a valine to aspartate mutation at position 49 (V49D). Because these were destabilizing mutations, they were modeled as a single mutation (position 45) and a double mutation (positions 45 and 49). As we had expected, the incorporation of these destabilizing mutations did decrease the overall helicity of the dimers. The inclusion of the second destabilizing mutation altered the structure enough to actually favor the homodimer mutant over the heterodimer as seen in the overall helicity of the coiled-coil dimers. A final rational design consideration was made to mutate Cys38 to alanine to minimize free thiol interactions. Inclusion of this final mutation to our earlier rational design models shows that the C38A mutation had very little effect in the dimers with no destabilizing mutations; however, helicity and specificity increased in models which incorporated rationally designed destabilizing mutations. These increases in helicity indirectly suggest increased stability and to validate this assertion, more detailed free energetic approaches were applied to estimate the free energy of binding.

MM-PBSA and thermodynamic integration free energy of binding calculations were completed for the wild-type Bcr dimer and the mutant dimers that incorporated the rational design choices as defined in Table 3.2. For comparison two Bcr mutants were chosen, CCmut1 and CCmut2. The sequence of the proteins differ in the incorporation of both destabilizing hydrophobic mutations in CCmut2 (CCmut1 only contains the L45D mutation) and the lack of the C38A mutation in CCmut2. MM-PBSA and thermodynamic integration results are shown in Table 3.3. Variation from the mean values was reported using standard deviation and standard error(50). In computational approaches to calculating the free energy of binding a large deviation from the mean is common and can be attributed to two causes. First, many energy approaches calculate the free energy of binding as the sum of individual energetics components which frequently exceed ten of thousand of kcal/mol energy units to arrive at a final value near 0 kcal/mol. Small errors in values of such a magnitude difference can be negated in the mean but are emphasized in the deviation. Second, the final value reported is the mean free energy of

Table 3.2 - Coiled-coil names and defining mutations for dimers used in MM-PBSA and thermodynamic integration energy calculation methodologies and experimental translocation assays.

Coiled Coil Dimer	**Mutations**
CC-CC	Homodimer Wildtype
CC-CCmut1	Heterodimer C38A,S41R,L45D,E48R,Q60E
CCmut1-CCmut1	Homodimer C38A,S41R,L45D,E48R,Q60E
CC-CCmut2	Heterodimer S41R, L45D,E48R,V49D,Q60E
CCmut2-CCmut2	Homodimer S41R, L45D,E48R,V49D,Q60E

Table 3.3 - MM-PBSA and thermodynamic relative free energy of binding results (in kcal/mol). MM-PBSA results were found by subtracting the absolute free energies of the unbound monomers from the calculated free energy of coiled-coil dimer using separate MD trajectories. Thermodynamic integration calculations followed the scheme described in Figure 2 using intermediate coiled-coil dimers to build a consistent transition from the wild type coiled-coil dimer to the CCmut1dimers. Results from both calculations were reported relative to the wild-type coiled-coil dimer (CC-CC) with respective standard deviation and standard error.

MM-PBSA Calculation Results			
Coiled Coil Dimer	$\Delta\Delta G_{binding}$ (kcal/mol)	Stdev (kcal/mol)	Stderr (kcal/mol)
CC-CC	0.00	49.93	3.16
CC-CCmut1	-1.28	49.28	3.12
CCmut1-CCmut1	11.41	48.10	3.04
CC-CCmut2	11.68	48.14	3.04
CCmut2-CCmut2	25.53	49.07	3.10
Thermodynamic Integration Results			
Coiled Coil Dimer	$\Delta\Delta G_{binding}$ (kcal/mol)	Stdev (kcal/mol)	Stderr (kcal/mol)
CC-CC	0.00	0.00	NA
CC-CCmut1	-1.14	12.04	0.54
CCmut1-CCmut1	3.25	17.10	0.76

binding and represent the average energetic value of an ensemble of molecular configurations. Because of the large range of molecular configurations sampled at room temperature, a similarly large range of energetic values exist. The reported value, if sampled over a large enough ensemble of configurations, represent the average value for the free energy of binding similar to the average position of a vibrating guitar string. Overall, the measurements for the free energy of binding follow trends and ranking seen in the initial helicity measurement. In all cases, CC-CCmut1 is shown to be slightly more stable than both the wild-type dimer and its complementary homodimer, CCmut1-CCmut1. Some variation is seen in the free energy of the CCmut2 homodimer and heterodimer. While the wild-type dimer (CC-CC) and CC-CCmut1 dimer fold tightly with burial of hydrophobic residues, the CCmut1-CCmut1 dimer and CCmut2 dimers have wider centers due to charge repulsion which expose hydrophobic residues to the solvent (see Figure 3.6). In the MM-PBSA measurements, this translates to a decrease in van der Waals energies, a slight increase in entropy due to the exposed hydrophobic residues (less ordered), and a change in the ordering and solvent exposure of charged residues. The MM-PBSA free energy of binding calculations ranks the CCmut2 heterodimer as the more favored of the pair which differs from the helicity calculations; however, both calculations show their stability significantly below the wild-type and CC-CCmut1 dimers. The variation seen between thermodynamic integration and MM-PBSA results may be attributed to the differences in the surface area of the dimers. The use of a surface area term to approximate the nonpolar contribution for the free energy of solvation in MM-PBSA calculations has recently been called into question (57-60). In coiled-coils, large charge densities exist and surface areas can change significantly due to secondary structure formation upon binding which may propagate errors in this energy term (see Chapter 5 for additional discussion of MM-PBSA energetics and coiled-coils).

Having established a competitive, highly specific heterodimer Bcr-mutant coiled-coil by simple and high-level computational approaches, experimental approaches were applied to determine the feasibility of using this heterodimer mutant in the pharmaceutical treatment of CML. The most favorable Bcr mutant, CCmut1, was designed with five mutations in the Bcr coiled-coil

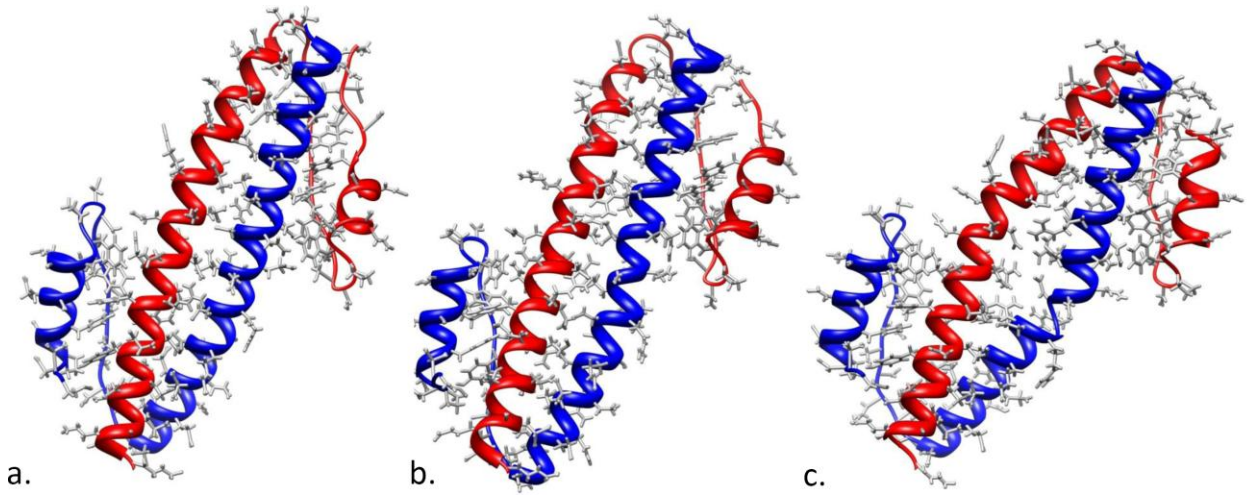


Figure 3.6 - A molecular graphics representation of the coiled-coil dimers: CC-CC(a), CC-CCmut1(b), and CCmut2-CCmut2(c). While good secondary structure formation and hydrophobic core packing is evident in the former two dimers, opposing charges and destabilizing charges in the hydrophobic core result in a widening protein-protein interface of the central heptad and solvent exposure of hydrophobic residues in CCmut2-CCmut2.

domain: C38A, S41R, L45D, E48R, and Q60E. Experimental studies in the Lim laboratory using this mutant in transfected cells confirmed improved binding to the oncoprotein over the heterodimer and found that this protein construct decreased autophosphorylation of the Bcr-Abl oncoprotein, downstream phophorylation activity and cell proliferation in expressing cells and induced caspase activity and apoptosis (61). A summary of their experimental approaches and results can be found in Table 3.4.

Conclusions

A combination of computational and experimental approaches were used to designed a heterodimer specific mutant Bcr-Abl binding partner for use as a therapeutic in the treatment of Chronic Myelogenous Leukemia (CML). Initial computational approaches monitored increases in α-helicity as a correlate for improved free energy of binding of mutant pairs. Mutations focused on approaches to improve salt bridge interactions, stability by forming of a disulfide bond, helicity in the backbone by incorporating alanine mutations to position **f** of the coiled-coil heptad, and designing destabilizing mutation to improve specificity. We found that a rational design approach that improved salt bridge interactions also significantly increased the helicity (and presumably the free energy of binding) of both the homodimer and heterodimer pairs. When the three designed point mutations were incorporated, helicity calculations of the homodimer and heterodimer mutants exceeded the wild-type dimer which may have decrease specificity for the heterodimer form. Approaches to incorporate a disulfide bond and alanine residues in the coiled-coil backbone caused structural distortions and decreased the specificity of heterodimer association as suggested by the short MD simulations and helicity calculations. Destabilizing point mutations were pursued to improve the heterodimer specificity of the prior rational design. Aspartate mutations were incorporated into the hydrophobic core and evaluated as a single and double mutation. Both mutations decreased the secondary structure (and stability) of the heterodimer while only the single point mutation maintained heterodimer specificity. A final planned mutation converted an exposed thiol group (Cys38) to alanine and in conjunction with the four prior mutations stabilized the heterodimer over both the homodimer and the wild-type dimer. Specificity as measured by helicity suggests that the homodimer (CCmut1-CCmut1) is

Table 3.4 – Experimental studies performed by the Lim laboratory show improved binding of the designed heterodimer protein construct. Binding of the protein construct decreased auto-phosphorylation and activity in the oncoprotein and induced caspase activity and apoptosis.

Experiment	Results
Nuclear Translocation Assay	Binding between coiled-coils was measured as a function of the ability of the designed mutant to translocate the protein dimer to the cell nucleus. The only interaction found to be statistically different (better) than the negative control is the binding of designed mutant with the wildtype coiled-coil domain (CC:CCmut1).
Mammalian Two-Hybrid Assay	The Two Hybrid Assay calculates binding between protein domains as a function of activation of downstream reporter genes. The greatest binding was found between the mutant and the wildtype coiled-coil domain (CC:CCmut1). Homo-oligomerization of the designed coiled-coil mutant domains (CCmut1:CCmut1) was statistically indistiquishable from the negative control.
Western Blotting with an Antibody Specific for Phosphorylated Bcr-Abl	Activation of the Bcr-Abl oncoprotein occurs through autophosporylation. Disruption in the oligomerization or phosphorylation of the Bcr-Abl proteins results in a decrease in activity. The wildtype coiled-coil domain has a minimal effect on the level of Bcr-Abl phosphorylation while the mutant coiled-coil domain reduced the phosphorylation level by 35%.
Western Blotting with an Antibody Specific for Phosphorylated CrkL	Bcr-Abl activity was monitored by measuring downstream phosphorylation of CrkL. The wildtype coiled-coil domain had minimal effect on the phosphorylation of CrkL while the mutant coiled-coil domain reduced the phosphoration of CrkL.
Western Blotting with an Antibody Specific for Phosphorylated STAT5	Downstream phosphorylation of STAT5 was monitored to detect changes to Bcr-Abl activity. The wildtype coiled-coil domain had minimal effect on the phosphorylation of STAT5 while the mutant coiled-coil domain reduced the phosphoration of STAT5.
Cell Proliferation Assay with Trypan Blue Exclusion	A decrease in cell proliferation was seen using both the wildtype (CC) and designed coiled-coil construct (CCmut1). The wildtype coiled-coil showed a slight decrease in proliferation while the designed construct showed was most effective as decreasing the number of proliferating cells.
Colony Forming Assay	The mutant coiled-coil domain was found to have the greatest effect at reducing the cell profileration among all proteins. This effect was similar to that seen using imatinib (Gleevec).
Fluorimetric Assay of Caspase-3/7 Activity	Caspase activity was used to monitor induction of apoptosis. The mutant coiled-coil domain showed the greatest activity and was only protein construct able to induce the activation of caspase at a statistically significant level.
DNA Segmentation	A metric to measure late stage apoptosis. Cells transfected with the CCmut1 protein construct showed segmented nuclei, consist with late stage apoptosis. Wildtype (CC) transfected and control cells showed healthy (round) nuclei.

significantly less stable than the wild-type (CC-CC) and the heterodimer (CC-CCmut1). More quantitative and accurate energetic approaches using MM-PBSA and thermodynamic integration also suggest improved stability of the heterodimer compared to the wild-type and specificity of the heterodimer over the homodimer. Experimental approaches using translocation and two-hybrid assays verified the improved binding of the heterodimer that we found using computational approaches. Further, experimental assays found that expression of the designed protein decreased the activity of the oncoprotein as measured in phosphorylation studies. This inactivitation of Bcr-Abl decreased cell proliferation and induced caspase and apoptosis activity in transformed cells.

References

1. Druker B, Talpaz M, Resta D, et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. New England Journal of Medicine 2001; 344: 1031-1037.

2. Faderl S, Talpaz M, Estrov Z, O'Brien S, Kurzrock R, Kantarjian H. The biology of chronic myeloid leukemia. New England Journal of Medicine 1999; 341: 164-172.

3. Sawyers C. Chronic myeloid leukemia. New England Journal of Medicine 1999; 340: 1330-1340.

4. Nowell P, Hungerford D. A minute chromosome in human chronic granulocytic leukemia. Science 1960; 132: 1-5.

5. Rowley J. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature 1973; 243: 290-296.

6. Zhao X, Ghaffari S, Lodish H, Malashkevich V, Kim P. Structure of the Bcr-Abl oncoprotein oligomerization domain. Nature Structural & Molecular Biology 2002; 9: 117-120.

7. Taylor C, Keating A. Orientation and oligomerization specificity of the Bcr coiled-coil oligomerization domain. Biochemistry 2005; 44: 16246.

8. McWhirter J, Galasso D, Wang J. A coiled-coil oligomerization domain of Bcr is essential for the transforming function of Bcr-Abl oncoproteins. Molecular and Cellular Biology 1993; 13: 7587-7595.

9. Pendergast A, Gishizky M, Havlik M, Witte O. SH1 domain autophosphorylation of P210 BCR/ABL is required for transformation but not growth factor independence. Molecular and Cellular Biology 1993; 13: 1728-1736.

10. Warmuth M, Danhauser-Riedl S, Hallek M. Molecular pathogenesis of chronic myeloid leukemia: implications for new therapeutic strategies. Annals of Hematology 1999; 78: 49-64.

11. Hazlehurst L, Bewry N, Nair R, Pinilla-Ibarz J. Signaling networks associated With Bcr-Abl–dependent transformation. Cancer Control 2009; 16: 100–107.

12. Kharas M, Deane J, Wong S, et al. Phosphoinositide 3-kinase signaling is essential for ABL oncogene-mediated transformation of B-lineage cells. Blood 2004; 103: 4268-4275.

13. Skorski T, Bellacosa A, Nieborowska-Skorska M, et al. Transformation of hematopoietic cells by BCR/ABL requires activation of a PI-3k/Akt-dependent pathway. The EMBO Journal 1997; 16: 6151-6161.

14. Sattler M, Salgia R, Okuda K, et al. The proto-oncogene product p120CBL and the adaptor proteins CRKL and c-CRK link c-ABL, p190BCR/ABL and p210BCR/ABL to the phosphatidylinositol-3'kinase pathway. Oncogene 1996; 12: 839-846.

15. Brunet A, Kanai F, Stehn J, et al. 14-3-3 transits to the nucleus and participates in dynamic nucleocytoplasmic transport. The Journal of Cell Biology 2002; 156: 817.

16. Obsilova V, Vecer J, Herman P, et al. 14-3-3 Protein interacts with nuclear localization sequence of forkhead transcription factor FoxO4. Biochemistry 2005; 44: 11608-11617.

17. Pap M, Cooper G. Role of glycogen synthase kinase-3 in the phosphatidylinositol 3-kinase/Akt cell survival pathway. Journal of Biological Chemistry 1998; 273: 19929-19932.

18. Gesbert F, Griffin J. Bcr/Abl activates transcription of the Bcl-X gene through STAT5. Blood 2000; 96: 2269-2276.

19. Huang M, Dorsey J, Epling-Burnette P, et al. Inhibition of Bcr-Abl kinase activity by PD180970 blocks constitutive activation of Stat5 and growth of CML cells. Oncogene 2002; 21: 8804-8816.

20. Magne S, Caron S, Charon M, Rouyez M, Dusanter-Fourt I. STAT5 and Oct-1 form a stable complex that modulates cyclin D1 expression. Molecular and Cellular Biology 2003; 23: 8934-8945.

21. Sillaber C, Gesbert F, Frank D, Sattler M, Griffin J. STAT5 activation contributes to growth and viability in Bcr/Abl-transformed cells. Blood 2000; 95: 2118-2125.

22. de Groot R, Raaijmakers J, Lammers J, Koenderman L. STAT5-dependent CyclinD1 and Bcl-xL expression in Bcr-Abl-transformed cells. Molecular Cell Biology Research Communications 2000; 3: 299-305.

23. Nieborowska-Skorska M, Wasik M, Slupianek A, et al. Signal transducer and activator of transcription (STAT5) activation by BCR/ABL is dependent on intact Src homology (SH) 3 and SH2 domains of BCR/ABL and is required for leukemogenesis. The Journal of Experimental Medicine 1999; 189: 1229-1242.

24. Cortez D, Reuther G, Pendergast A. The Bcr-Abl tyrosine kinase activates mitogenic signaling pathways and stimulates G1-to-S phase transition in hematopoietic cells. Oncogene 1997; 15: 2333-2342.

25. Daley G, Van Etten R, Baltimore D. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. Science 1990; 247: 824-830.

26. Kelliher M, McLaughlin J, Witte O, Rosenberg N. Induction of a chronic myelogenous leukemia-like syndrome in mice with v-abl and BCR/ABL. Proceedings of the National Academy of Sciences 1990; 87: 6649-6653.

27. Heisterkamp N, Jenster G, Ten Hoeve J, Zovich D, Pattengale P, Groffen J. Acute leukaemia in bcr/abl transgenic mice. Nature 1990; 344: 251 - 253.

28. Lugo T, Pendergast A, Muller A, Witte O. Tyrosine kinase activity and transformation potency of bcr-abl oncogene products. Science 1990; 247: 1079-1082.

29. Nardi V, Azam M, Daley G. Mechanisms and implications of imatinib resistance mutations in BCR-ABL. Current Opinion in Hematology 2004; 11: 35-43.

30. Sawyers C. Disabling Abl—Perspectives on Abl kinase regulation and cancer therapeutics. Cancer Cell 2002; 1: 13-15.

31. Daley G. Gleevec resistance: lessons for target-directed drug development. Cell Cycle 2003; 2: 190-191.

32. Hazarika M, Jiang X, Liu Q, et al. Tasigna for chronic and accelerated phase Philadelphia chromosome–positive chronic myelogenous leukemia resistant to or intolerant of imatinib. Clinical Cancer Research 2008; 14: 5325-5331.

33. Beissert T, Puccetti E, Bianchini A, et al. Targeting of the N-terminal coiled coil oligomerization interface of BCR interferes with the transformation potential of BCR-ABL and increases sensitivity to STI571. Blood 2003; 102: 2985-2993.

34. Vigneri P, Wang J. Induction of apoptosis in chronic myelogenous leukemia cells through nuclear entrapment of BCR–ABL tyrosine kinase. Nature Medicine 2001; 7: 228-234.

35. Shire S, Shahrokh Z, Liu J. Challenges in the development of high protein concentration formulations. Current Trends in Monoclonal Antibody Development and Manufacturing: Springer New York; 2010. p. 131-147.

36. Guex N, Peitsch M. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 1997; 18: 2714-2723.

37. Case D, Darden T, Cheatham III T, et al. AMBER 9. University of California, San Francisco; 2006.

38. Duan Y, Wu C, Chowdhury S, et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. Journal of Computational Chemistry 2003; 24: 1999-2012.

39. Case DA, Cheatham TE, 3rd, Darden T, et al. The Amber biomolecular simulation programs. Journal of Computational Chemistry 2005; 26: 1668-1688.

40. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparisons of simple potential functions for simulating liquid water. Journal of Chemical Physics 1983; 79: 926-935.

41. Aqvist J. Ion-water interaction potentials derived from free energy perturbation simulations. Journal of Physical Chemistry 1990; 94: 8021-8024.

42. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. Journal of Computational Physics 1977; 23: 327-341.

43. Barth E, Kuczera K, Leimkuhler B, Skeel RD. Algorithms for constrained molecular dynamics. Journal of Computational Chemistry 1995; 16: 1192 - 1209.

44. Essmann U, Perera L, Berkowitz M, Darden T, Lee H, Pedersen L. A smooth particle mesh Ewald method. Journal of Chemical Physics 1995; 103: 8577-8593.

45. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. Journal of Computational Physics 1984; 81: 3684-3690.

46. Bulheller B, Hirst J. DichroCalc--circular and linear dichroism online. Bioinformatics 2009; 25: 539-540.

47. Cox R, Hirst W. Calculation of the circular dichroism of bihelical ribonucleic acids as a function of nucleotide composition. Biochemical Journal 1976; 155: 292-295.

48. Rost B, Sander C. Prediction of protein secondary structure at better than 70% accuracy. Journal of Molecular Biology 1993; 232: 584-584.

49. Pettersen E, Goddard T, Huang C, et al. UCSF Chimera—a visualization system for exploratory research and analysis. Journal of Computational Chemistry 2004; 25: 1605-1612.

50. Gohlke H, Case DA. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. J Comput Chem 2004; 25: 238-250.

51. Rocchia W, Alexov E, Honig B. Extending the applicability of the nonlinear Poisson-Boltzmann equation: multiple dielectric constants and multivalent ions. Journal of Physical Chemistry B 2001; 105: 6507-6514.

52. Massova I, Kollman P. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. Perspectives in Drug Discovery and Design 2000; 18: 113-135.

53. Zitzewitz J, Ibarra-Molero B, Fishel D, Terry K, Robert Matthews C. Preformed secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system. Journal of Molecular Biology 2000; 296: 1105-1116.

54. Jelesarov I, Bosshard H. Thermodynamic characterization of the coupled folding and association of heterodimeric coiled coils (leucine zippers). Journal of Molecular Biology 1996; 263: 344-358.

55. Hodges R, Saund A, Chong P, St-Pierre S, Reid R. Synthetic model for two-stranded alpha-helical coiled-coils. Design, synthesis, and characterization of an 86-residue analog of tropomyosin. Journal of Biological Chemistry 1981; 256: 1214-1224.

56. Mason J, Arndt K. Coiled coil domains: stability, specificity, and biological implications. ChemBioChem 2004; 5: 170-176.

57. Pearlman D. Evaluating the molecular mechanics poisson-boltzmann surface area free energy method using a congeneric series of ligands to p 38 MAP kinase. Journal of Medicinal Chemistry 2005; 48: 7796-7807.

58. Gouda H, Kuntz I, Case D, Kollman P. Free energy calculations for theophylline binding to an RNA aptamer: comparison of MM-PBSA and thermodynamic integration methods. Peptide Science; 68: 16-34.

59. Levy R, Zhang L, Gallicchio E, Felts A. On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute-solvent interaction energy. Journal of the American Chemical Society 2003; 125: 9523-9530.

60. Chen J, III C. Implicit modeling of nonpolar solvation for simulating protein folding and conformational transitions. Physical Chemistry Chemical Physics 2008; 10: 471-481.

61. Dixon AS, Pendley SS, Bruno BJ, et al. Disruption of Bcr-Abl coiled-coil oligomerization by design. Submitted to the Journal of Biological Chemistry.

CHAPTER 4

ASSESSING MODEL STRUCTURES OF SMALL, PARALLEL IAAL-E3/K3 DIMERIC COILED-COILS WITH THE AMBER FORCE FIELDS

Abstract

The use of all-atom molecular dynamics (MD) with simple additive and pair-wise molecular mechanic potential energy representations to model bio-molecules, especially proteins, has shown great promise and many successes in the last few decades. Much of that success has been due to improved force field parameters including better fits for peptide backbone dihedrals. We confirm that although many modern force fields correctly model large coiled-coils, attempts to model smaller heterodimeric, parallel coiled-coils can result in significant structural distortions and loss of secondary structure. As is well appreciated in the literature, failures with the earlier AMBER force fields in modeling the IAAL-E3/K3 coiled-coil and other proteins result from inaccurate parameterization of peptide backbone dihedral terms. In this report, we further suggest that biases in the side chain dihedral parameters, such as with isoleucine—even with balanced force fields including AMBER ff99SB and ff03—can lead to disruption of small coiled-coil structure.

Introduction

The use of simulation methods to model and understand atomistic and molecular behavior, kinetics, and energetics often requires the use and development of representative force fields at the molecular mechanics level of resolution (1-6). With application and additional testing, modifications to these force fields become necessary to correct for known biases and errors in the development of these models. In the AMBER suite of program (7), the Cornell et al. force field (8) (denoted as ff94) formed the basis of the second generation of additive nonpolarizable force field for proteins, nucleic acids and organic compounds in condensed phase. This all atom

force field was largely inspired by the optimized potential for liquid simulations (OPLS) (9) and utilized RESP charges (10, 11) calculated at the Hartree-Fock 6-31G* level to compensate for the lack of polarizability. Characteristic features included fixed partial charges on atom centers and explicit use of hydrogen atoms. Parameterization schemes for bonds, angles, and dihedrals were developed with the aim that they would be easily transferable across molecules. This is relatively straightforward for bonds and angles since the equilibrium bond lengths, angles, and vibrational frequencies are relatively decoupled from their surroundings. This is less true with the dihedrals and therefore the dihedral terms are typically parameterized last as these provide the free parameters or slop to augment the other molecular mechanical terms to reproduce relative energetics due to rotation around bonds. The protein Φ and Ψ dihedral specific energy rotational barriers in the AMBER force fields were originally determined using QM simulations of several glycine and alanine conformations. Extensive testing of this force field revealed a limitation in the fitting of the peptide backbone dihedral parameters leading to a preference for α-helicity in secondary structure (12). Revisions to these dihedral terms were proposed in subsequent modifications of ff94, first ff96 (13) and later ff99 (14).

Parameterization of protein backbone dihedral terms in ff96 used equivalent Φ and Ψ terms adjusted to reproduce the energy differences in extended and constrained alanine tetrapeptides. Following use and testing of ff96, a preference for β-sheet conformations in secondary structure was found (15-18). Dihedral parameterization continued in ff99 by expanding the training set to include 11 representative structures of alanine tetrapeptides as well as alanine dipeptides. Similar to the original ff94 parameterization, an over-stabilization of α-helical character was shown in ff99 (12, 19). New parameters sets seeking to correct these biases have been proposed by many authors including Garcia (12), Pande (20, 21), Simmerling (ff99-mod1 and ff99SB) (19, 22, 23), Wang and Luo (ff99-mod2) (7), and Duan (ff03) (24). Assessment of AMBER force fields by direct comparison of different model systems has shown significant differences between simulations using implicit or explicit solvent (25). Attempts to model α-helical and β-sheet peptides favored ff96 and ff03 (using the implicit solvent model of Onufriev, Bashford, and Case) over ff94, ff99, and ff99SB (25, 26). Similarly, attempts to fold the

β-hairpin from the C-terminal region of Protein G using a Poisson Boltzmann solvent model showed a strong preference for ff03 over ff94, ff99-mod1, and ff99-mod2 following the zip-in folding pathway (27). In explicit solvent, folding simulations of helical proteins, β-sheets, β-turns, and native and decoy structures have shown ff03 and ff99SB to be superior to other non-polarizable AMBER force fields in correctly balancing secondary structure bias as determined by protein structure, and also with agreement between experimental and calculated order parameters (23, 24, 28-31). Recent work has also suggested that ff03 may favor right handed α-helical structures in simulations of trialanine in explicit water, however additional validation of this observation is warranted (23).

While the focus of this paper is to assess the ability of modern AMBER force fields to model coiled-coil structure and dynamics, other force field models have shown parallel improvements in modeling protein structure and dynamics following peptide backbone torsion refinement including: the GROMOS 53A6 force field (32-36), the most recent version of the OPLS-AA force field (37-41), and the CHARMM22/CMAP force field (42-46). The latest GROMOS force field, 53A6, has shown promise in correctly modeling the structure of hen egg-white lysozyme and the β^3-dodecapeptide (32, 33). Metadynamics derived free energy landscapes created using 53A6 of the α-helix portion of the G-protein (34) reproduced several experimental observations, and 53A6 was successful in predicting behavior and structure of two β-peptides in explicit methanol solvent (35). However, GROMOS 53A6 has been shown to overestimate the radius of gyration and underestimated the PPII content of Ace-(diaminobutyric acid)$_2$-(Ala)$_7$-(ornithine)$_2$-NH_2, and umbrella sampling of alanine and glycine dipeptides showed that backbone torsional angles varied significantly from quantum mechanical and crystal structure models (35, 36, 47). Additional parameterization to correct the backbone torsional angles has been recently published by Liu and coworkers (36).

In 2001, Jorgensen et al. reparameterized the OPLS-AA force field by refitting key Fourier torsional coefficients using ab initio quantum mechanical data of alanine di- and tetra-peptides and charged dipeptides (37). Using these parameters, Jorgensen calculated the energetics and interactions of HEPT and 20 nevirapine non-nucleoside inhibitors of HIV-1 reverse

transcriptase (38). Potential descriptors of binding affinity were determined and found to correlate with experimental activities with an r^2 value of 0.94 (38). Friesner and Rapp were able to show that these new parameters significantly improved side chain prediction on a database of 36 proteins (39). Jorgensen and Felts also showed that the new parameters in OPLS-AA could effectively be used in a scoring function (41) and to predict the free energy of hydration for neutral organic molecules (40).

MacKerell and coworkers used a grid based energy correction to the Ψ/Φ two dimensional surface with empirical adjustments to the alanine and glycine based surface to account for systematic differences observed between molecular dynamics and experimentally observed Ψ/Φ distributions in their design of the CHARMM/CMAP force field (42). These corrections were shown to improve conformational properties in protein crystals and alanine dipeptides (43). In hen egg-white lysozyme simulations, application of CMAP corrections eliminated substantial deviations from the backbone root mean-square fluctuations and N-H NMR order parameters, although some discrepancies remained (44). Continuum representations with CHARMM/CMAP of the B1 immunoglobulin-binding domain of streptococcal protein G and bovine pancreatic trypsin inhibitor showed significant improvement of CHARM22 at correctly modeling structural and dynamic properties (45) and application of the CHARM22/CMAP corrections to apoliprotein A–I models fixed the π-helix bias seen in earlier CHARMM 22 force field simulations (46).

In our work with coiled-coils we have attempted to model protein dynamics and behavior using a variety of the AMBER force fields. Similar to Brünger's work with the ff94 force field (48), we found the ff99 force field was parameterized sufficiently well to correctly model large coiled-coils including the p1 domain of the GCN4 leucine zipper (49) and Keating's synthetic heterodimer d12La16I/d12La16L (50). However, attempts to model small, heterodimeric parallel coiled-coils (3 heptad units or 21 residues per monomer) based on Hodges' highly-stable IAAL-E3K3 coiled-coil (pdb id: 1U0I) (51) resulted in denatured structures in nanosecond MD simulations while using the ff99 force field parameters. We report here our efforts to further model this system using various all-atom AMBER force fields including ff94, ff96, ff99, ff99-mod1,

ff99-mod2, ff99SB, and ff03. We have also included a united atom model based on the Duan ff03 force field (entitled ff03ua) (52), as well as the recently published polarizable force field ff02pol.r1 (53). Of the force fields tested, we found that the ff99SB and ff03 force fields most closely reproduced the structure of the IAAL-E3/K3 coiled-coil, however, energetic biases in isoleucine or potentially other side-chain torsions may lead to significant structural distortions, separations, and loss of dimer orientation in molecular dynamics simulations.

Methods

Starting Geometries

Initial structures were derived from the first NMR structure model of the IAAL-E3/K3 coiled-coil dimer (PDB ID: 1U0I, see Figure 4.1) (51). The studied sequence is represented as a helical wheel in Figure 4.2. NMR models 2 through 5 were also modeled using the ff99SB and ff03 force fields to investigate the starting conformation dependence on protein binding stability. Investigational studies using the GCN4 leucine zipper coiled-coil dimer (PDB ID: 2ZTA) and the Keating d12La16I/ d12La16L mutation of the GCN4 protein (PDB ID: 1KDD) were also performed using the ff99 force field. Terminal amino acid residues were not blocked and were assumed to exist in a charged state. Note that charged amino acids (Asp, Glu, Lys) were assumed to exist in ionic states consistent with direct exposure to solvent at pH 7.0. Coiled-coils were solvated by surrounding the protein with at least a 12 Å water layer in all directions within a truncated octahedron using explicit solvent. This amounts to approximately 4,500-6,000 waters. For all non-polarizable force fields TIP3P waters (54) were used, while POL3 waters were used in simulations using the ff02.polr1 force field. No explicit salt ions were initially added to the model because the system was neutral. Follow up studies exploring the effects of explicit salt concentrations in this force field assessment found that the addition of salt ions (150 mM Na^+ Cl^- explicit salt concentrations using the Joung/Cheatham parameters (55)) and different starting orientations accelerated the instability of the simulated coiled-coils.

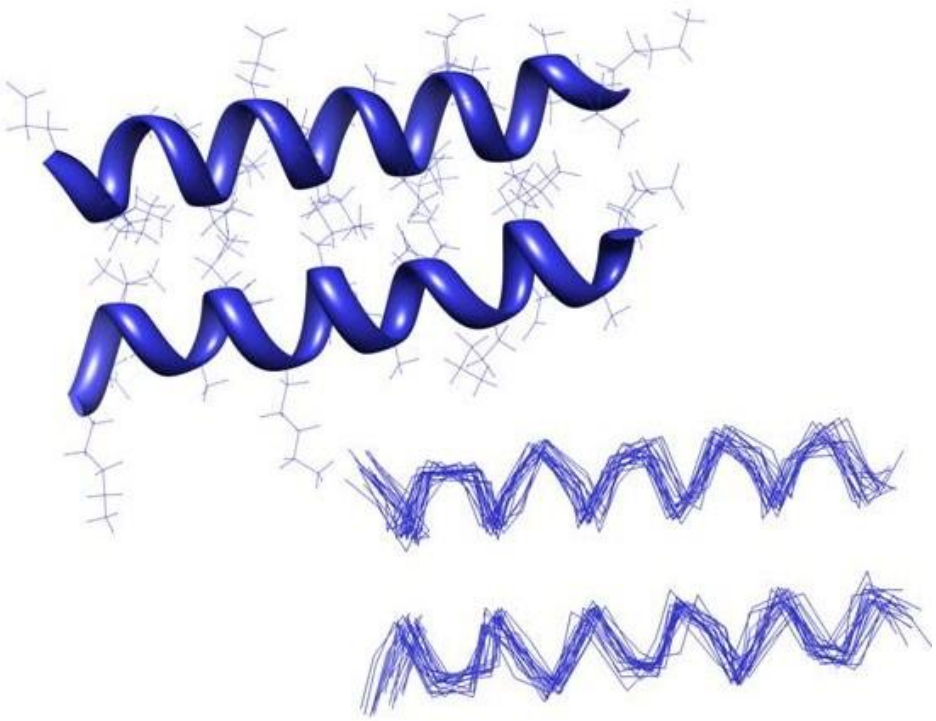


Figure 4.1 - Representative models of the IAAL-E3/K3 NMR structure. Twenty NMR models comprise the published structure (pdb id: 1U0I). The initial NMR model is shown as an all-atom, ball and stick representation with a ribbon representation of the secondary structure (top). Hydrophobic residues comprising the binding domain are shown in blue and grey. Ionic residues that form inter-helical salt bridges are shown in red and orange. A wire-frame model of the backbone atoms of all twenty NMR models shows similarity and freedom of movement in backbone atom positions (bottom).

IAAL-E3/K3

Position in heptad:	gabcdefgabcdefgabcdef
Monomer A	**E**IAAL**E**K**E**IAAL**E**K**E**IAAL**E**K
Monomer B	**K**IAAL**K**E**K**IAAL**K**E**K**IAAL**K**E

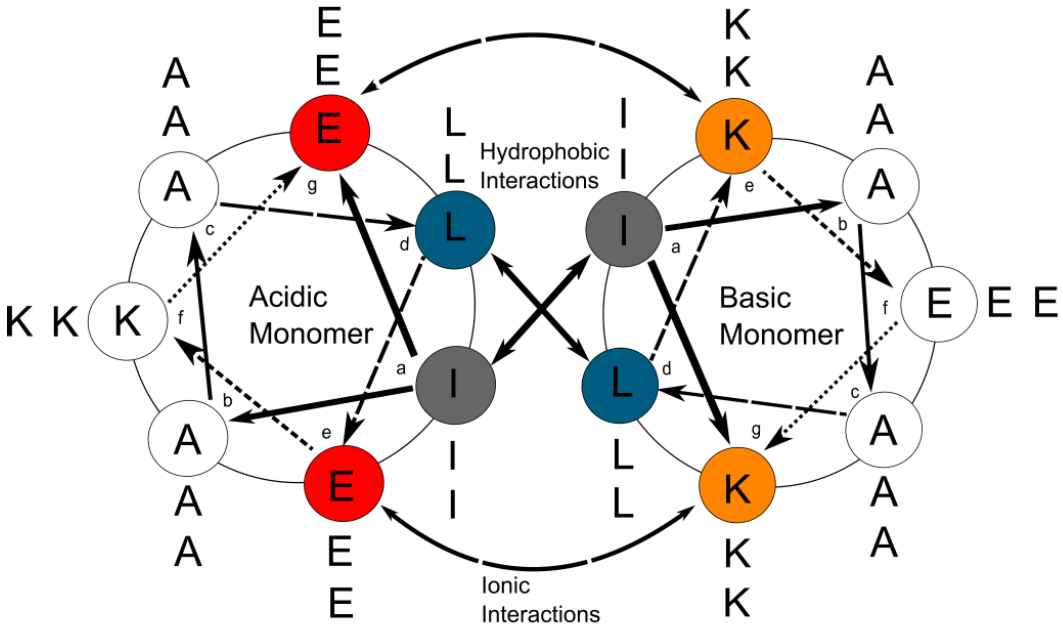


Figure 4.2 - The sequence (top) and a helical wheel representation (bottom) of IAAL-E3/K3 are shown. Hydrophobic residues comprising the binding domain are shown in blue and grey. Ionic residues that form inter-helical salt bridges are shown in red and orange. As seen from the representation and published studies, IAAL-E3K3 is completely alpha-helical with a low dissociation constant (70 nm) and a highly stable free energy of folding and association of 9.8 kcal mol^{-1}. Helical wheel representation adapted from Sykes et al(51).

Molecular Dynamics Simulations

All simulations of the IAAL-E3/K3 coiled-coil were completed using AMBER sander and pmemd modules with several different force fields including ff94, ff96, ff02.polr1 (a polarizable force field with refit Φ Ψ torsion parameters), ff99, ff99SB, ff03, ff03ua (united atom force field modification of the ff03 force field), Simmerling's original Φ/Ψ re-parameterization of the ff99 force field (denoted ff99-mod1), and Wang and Luo's re-parameterization of the ff99 force field (denoted ff99-mod2) (7).

Energy minimization was performed for 500 steps first in the system with restrained protein atoms (50 kcal mol^{-1} $Å^{-2}$) and then in an unrestrained system. Initial minimization was followed by heating to 300K at constant volume over a period of 10 ps using harmonic restraints of 2 kcal mol^{-1} $Å^{-2}$ on the protein atoms. Subsequent unrestrained simulation at 300K followed for 500 ps. Bond lengths involving hydrogen atoms were constrained with SHAKE (56, 57) (for simulations involved TIP3P waters) with a geometric tolerance for the constraint of 0.00001 Å during the coordinate resetting. Periodic boundary conditions were applied using the particle mesh Ewald method (PME) with a less than 1 Å charge grid and cubic B-spline interpolation (58). Follow up studies with ff99, ff99-mod2, and ff02pol.r1 force field models using restrained equilibration for 500 ps preceding production simulation found that the onset of secondary structure loss was slowed but not stopped using restraints and long equilibration times.

All of the production molecular dynamics (MD) simulations were performed with a 2 fs time step and a direct space nonbonded cutoff of 10 Å with the pair list of atomic interactions built out to 11 Å and heuristic update of the pair list triggered when any atom moved more than 0.5 Å since the previous update. During production runs, the center of mass translational motion of the entire system was removed after the initial velocity assignments and subsequently every 5000 MD steps. Constant temperature was maintained using the weak coupling algorithm and heat bath coupling with a 2 ps time constant (59). Pressure (1 atm) was maintained using isotropic position scaling with Berendsen weak coupling algorithm with a 1.0 ps pressure relaxation time (59). For the polarizable force field models, the time step was decreased to 1 fs and the Car-Parrinello scheme for assigning a fictitious mass to dipoles and integrating times steps was

followed allowing a maximum of 10 iterations per time step (7, 60-62). Temperature control on the dipoles was set to 1.0 ps units and 1-4 charge-dipole and dipole-dipole interactions were not scaled (53).

NMR Restraints

In some studies, intramolecular NOE-based distance and backbone torsion restraints were applied to the IAAL-E3/K3 model (using the ff99SB and ff03 force fields). The Ψ and Φ dihedral angles were restrained using a 2 kcal mol^{-1} rad^{-2} rotation barrier. Distance restraints were limited to main-chain interactions and used a 20 kcal mol^{-1} restraint. No intermolecular restraints were applied across the protein-protein interface.

Analysis

2D RMSD. Convergence was visualized by overlaying a 2-dimensional root mean square deviation plots (2D-RMSD) with total RMSD against the initial frame. 2D-RMSD plots were calculated using the AmberTools development version of the ptraj module. Approximately 1000 snapshots were chosen which were spaced equi-distance along the completed molecular dynamics trajectory. Typically, convergence is seen where comparative RMSD goes to a minimum; this is visualized as a large blue field occupying the upper right hand corner of the plot (and a plateau region in the 1D RMSD plot). Alternatively, convergence can be seen with smaller adjacent blue fields along the diagonal suggest that the protein may be sampling different minima or substrates.

Scanning frame RMSD. Regions of large structural deviation and flexibility in the simulations were visualized by calculating the root mean square deviation (RMSD) of a small region of the protein against the same region of average NMR structure and then shifting the window in which the RMSD was calculated. Due to the repeating heptad nature of the coiled-coil amino acid sequence, we chose to calculate all-atom RMSD based on seven amino acid segments starting initially with residues 1-7 then recalculated at residues 2-8, 3-9, 4-10 and so forth. Examining the differences as a function of simulation time allows the visualization of regions where structural distortions initiate and propagate. For the graphs presented in this

study, individual monomers were shown separated with the primary sequence (N terminal to C terminal) plotted on the x-axis, the time course of the y-axis (moving backwards into the plane with the progression of time), and RMSD variation on the z-axis.

NOE analysis. The AMBER 9.0 ptraj module was used to calculate distance measurements to compare the averaged unrestrained modeled structure against NOE-derived distance restraints over the initial 100 ns of MD simulation. Distances were measured every ps and average distances were compared against distance restraints used to construct the NMR model.

Ramachandran plots. Secondary structure of individual residues was determined based on the backbone Ψ and Φ dihedral angles and visualized using a Ramachandran plot (63). Plots appearing in this paper were created using the RAMPAGE Ramachandran plot assessment (64) of the coiled-coil structure based on molecule snapshots taken during the final 5 ns of MD simulation. The Ramachandran plot combines four overlapping graphs illustrating favorable regions for general amino acids and proline (blue), glycine favored (brown), glycine allowed (pink), and general amino acids and proline allowed (light green).

Rotamer analysis. Side-chain rotamers of leucine and isoleucine were calculated consistent with the definitions of Dunbrack and Cohen (65) and our earlier work (66). Rotamer pairs were assigned at each ps and the percent occupancy was calculated at 20 ns bins spanning 100 ns following an initial 20 ns MD equilibration.

Molecular graphics. Molecular graphics snapshots of metastable states (determined by plateaus in 1D-RMSD and blue fields in 2D-RMSD) and distance specific time steps were created using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (67). Structures were oriented N-terminal to C-terminal with the acidic monomer (monomer A) above and the basic monomer (monomer B) beneath. A ribbon representation of the secondary structure was superimposed over a ball and stick representation of atom positions and bonds.

Results and Discussion

All of the initial molecular dynamic simulations in this study were initiated with the first model from the set of submitted IAAL-E3/K3 NMR structures (pdb: 1UOI). Despite starting from the refined NMR model and using simulation times often exceeding 100 ns, 2D-RMSD plots showed that none of the force field derived models, perhaps with the exception of the very stable ff94 runs, was able to converge to a point of full stabilization or equilibration (see Figure 4.3). Examination of the snapshots taken at metastable states (as judged by 2D-RMSD) show very large structural deviations in the ff99, ff99-mod2, and ff02pol.r1 force fields in the interior heptads leading to deformation of the binding domain (see Figure 4.4). Similarly, large initial structural deformations occurred in the interior of ff99-mod1 force field model and continued to deform the C-terminus. These structure variances in the sampled substates, along with consideration of the recent literature (12, 15, 23), allows us to discount these force fields as viable force fields for use in modeling small coiled-coils and other proteins.

Several different AMBER force fields were tested despite well-known secondary structure biases. AMBER force fields ff94 and ff99 have been shown to favor the formation of α-helical structures (12, 15). Similarly, the AMBER ff96 force field is known to over stabilize β-sheets (15-18). Despite these tendencies, ff99 showed a reduced ability to maintain the coiled-coil structure compared to the β-sheet favoring force field ff96. Ramachandran plots (Figure 4.5) of the α-helical biased force fields show that while interior domains of the ff99 model are contained in the α-helical defined region, they have spread left of the classical domain towards the π-helix region consistent with the structural distortion seen in the snapshots in Figure 4.4. Ramachandran plots of ff96 show considerable occupancy of the β-sheet region which is less obvious in the secondary structure representations of the structural snapshots. The scanning RMSD (see Figure 4.6) does suggest the ff96 model is beginning to denature as a large peak in RMSD is seen in the C-terminal basic monomer just prior to the end of simulation. This is confirmed by observation of peaks seen in the acidic monomer C-terminus showing the largest deviation among all the tested force fields from the averaged NMR structure. Of the helical favoring force fields, ff94 showed very little freedom of movement in the terminal residues compared to ff03 and ff99SB. Since

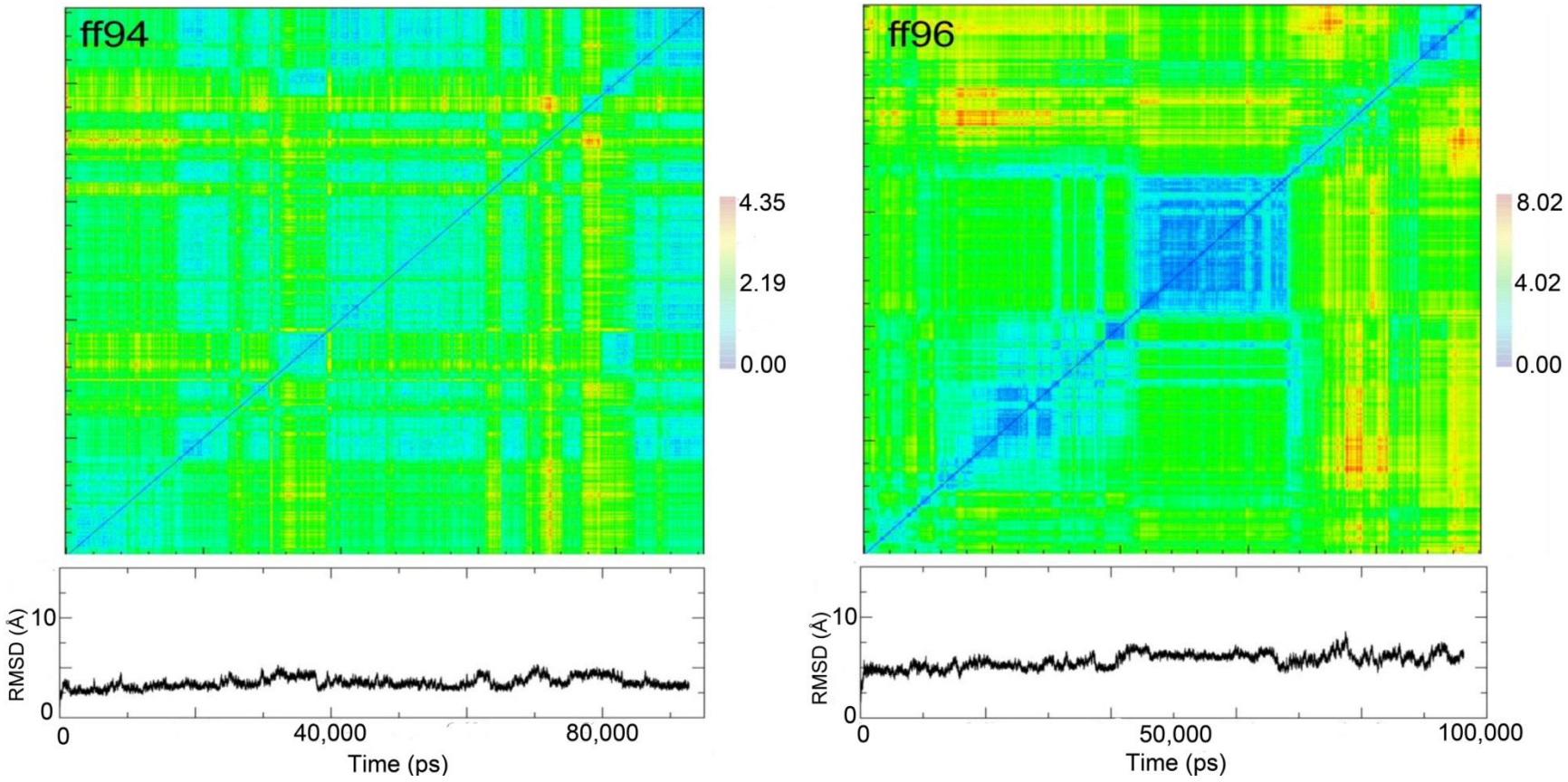


Figure 4.3 - Combined 1D and 2D root mean square deviation (RMSD) plots. 2D RMSD plots were calculated using 1000 individual snapshots taken at spaced interval to represent the entire production molecular dynamics trajectory. Comparison of both 1D and 2D RMSD plots above suggest that none of the force field simulation were able to achieve equilibration, though simulations using the ff99-mod1 and ff02pol.r1 force fields come closest to that measure. Note that scales in each RMSD plot are different and that the ff02pol.r1 force field RMSD calculations refer to the POL3 water model only.

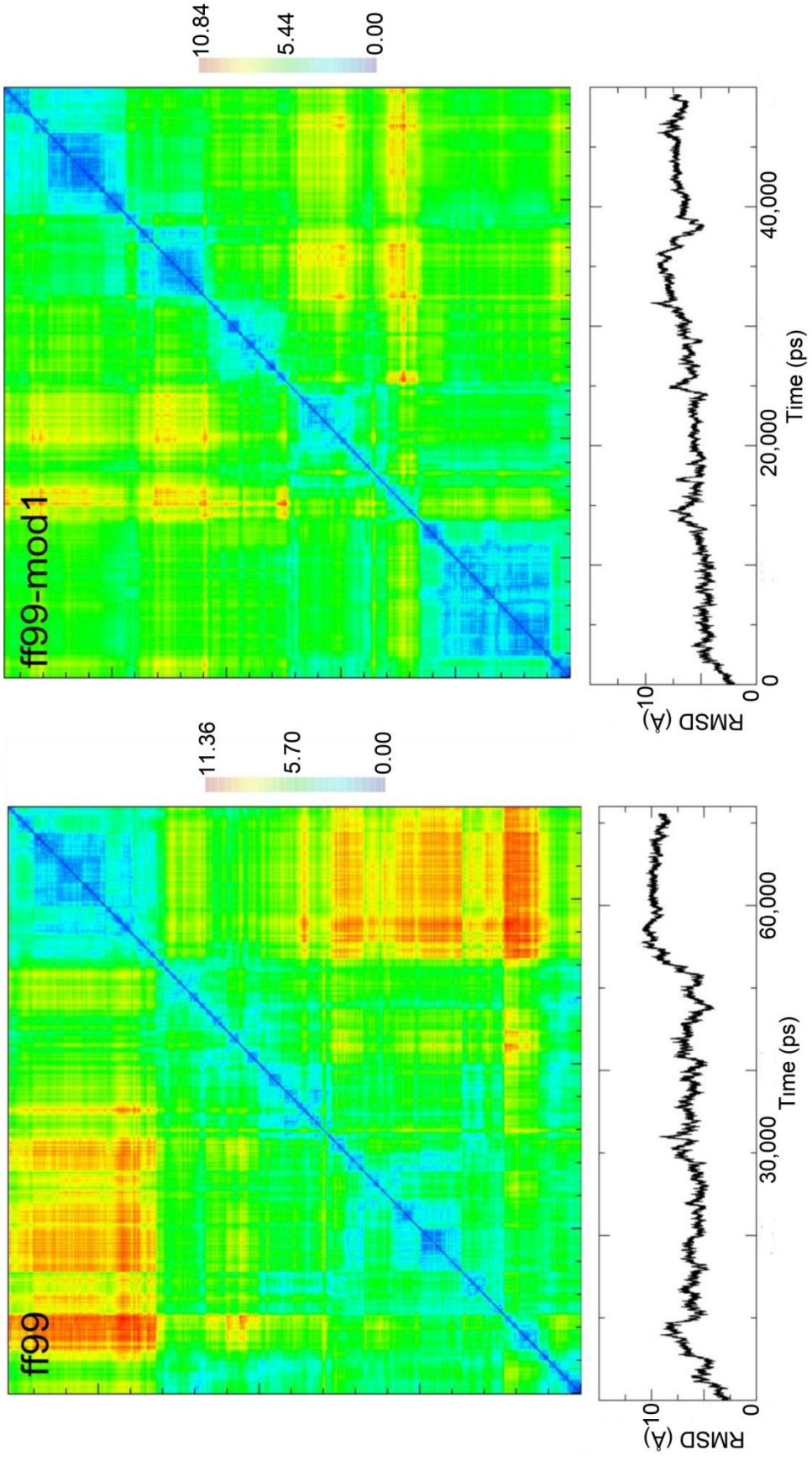


Figure 4.3 Continued

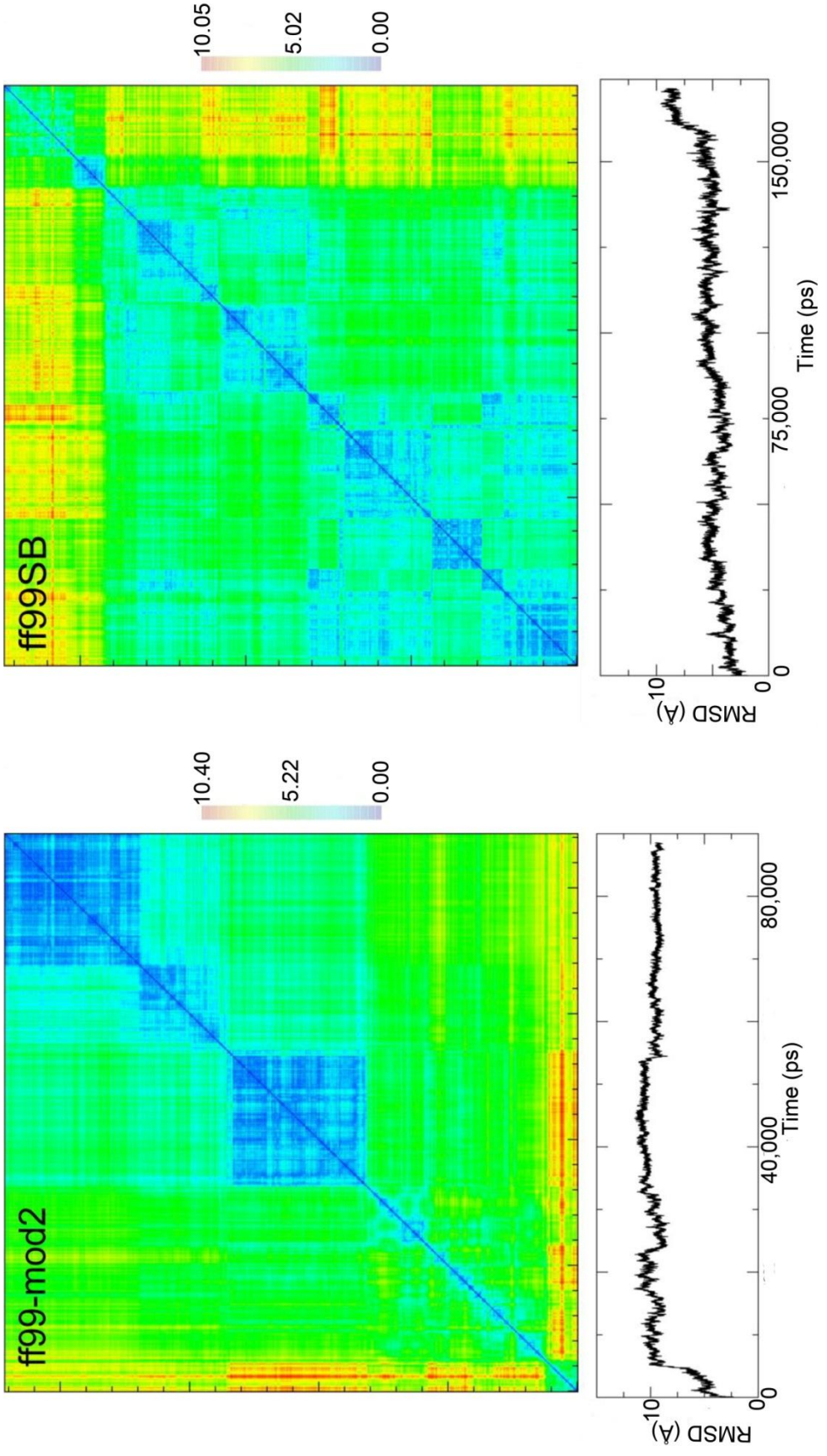
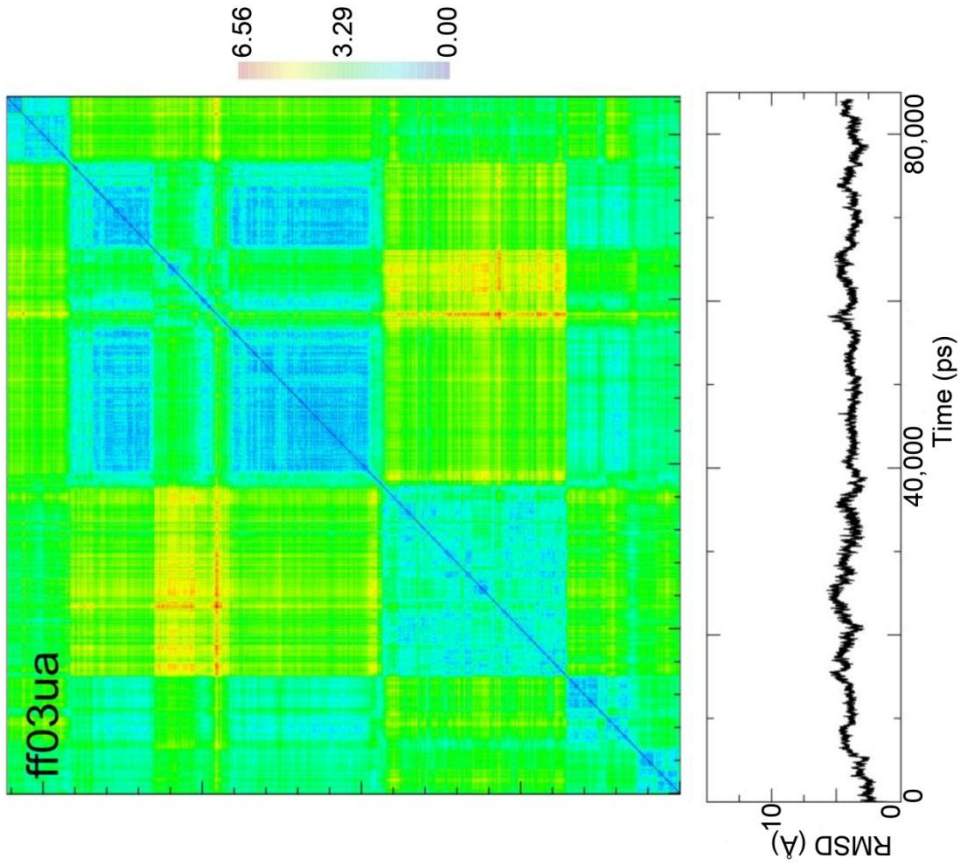


Figure 4.3 Continued



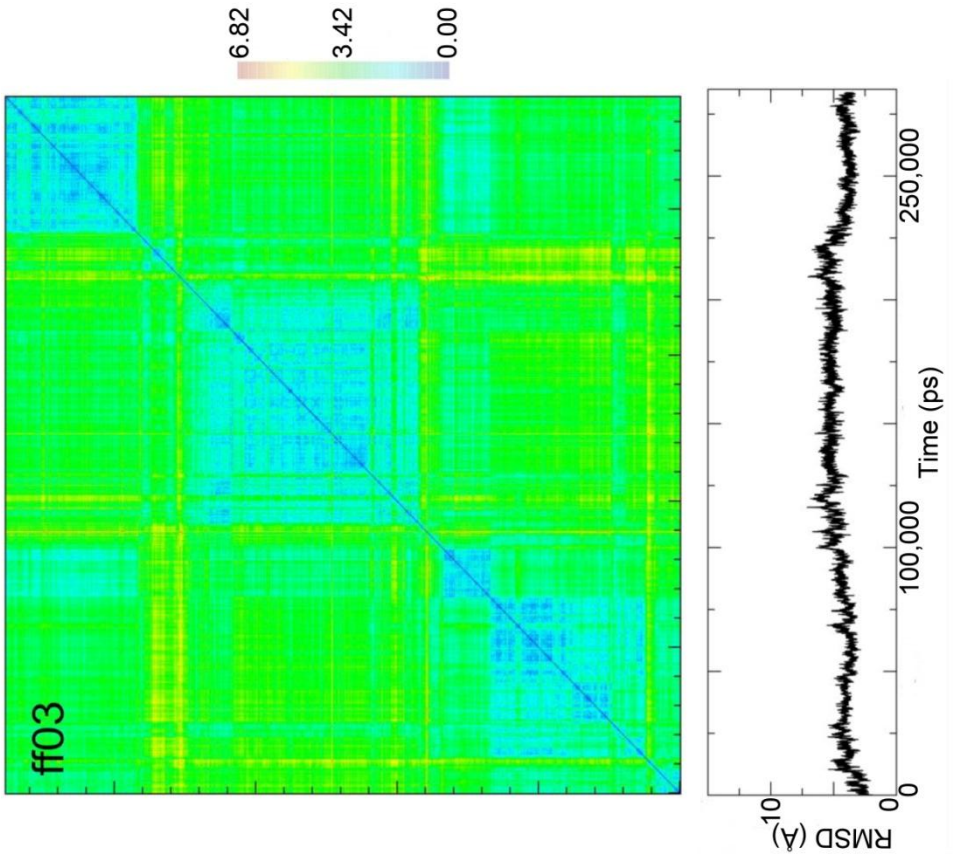


Figure 4.3 Continued

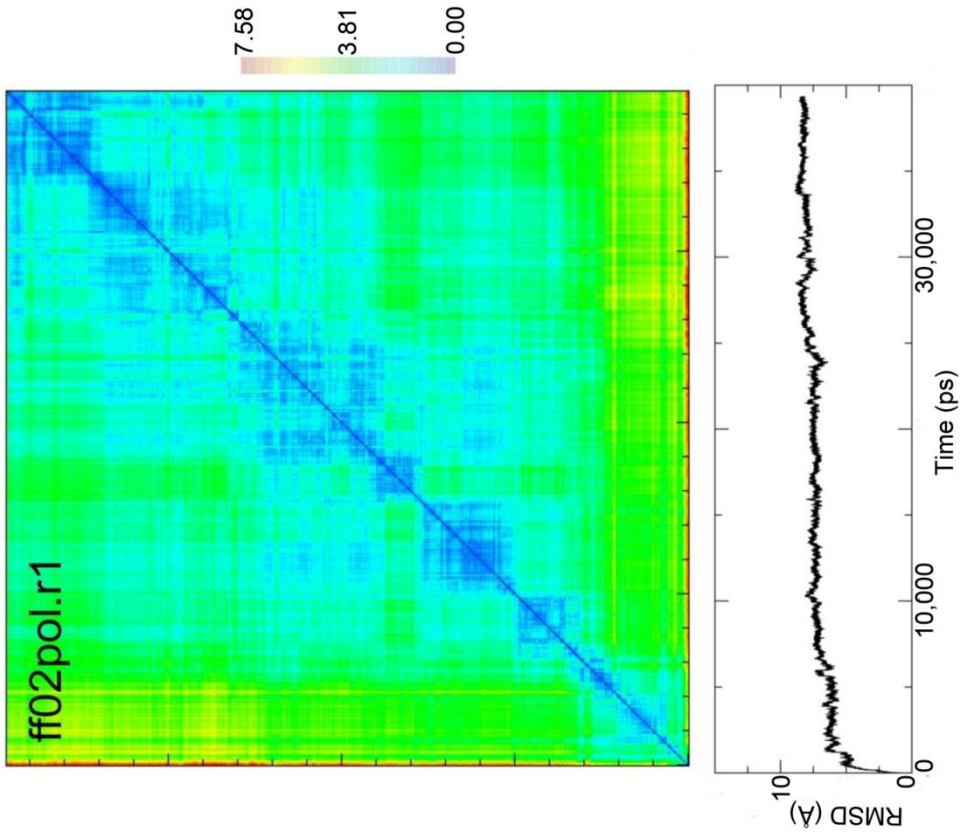


Figure 4.3 Continued

Figure 4.4 - Snapshots generated along the molecular dynamics time-course. Time periods for representative snapshots were chosen to best represent moments of dynamic structural change and metastable states as determined by 1D and 2D RMSD calculations. Figurative models were created showing all atom positions with ball and stick representations and secondary structure using a ribbon representation.

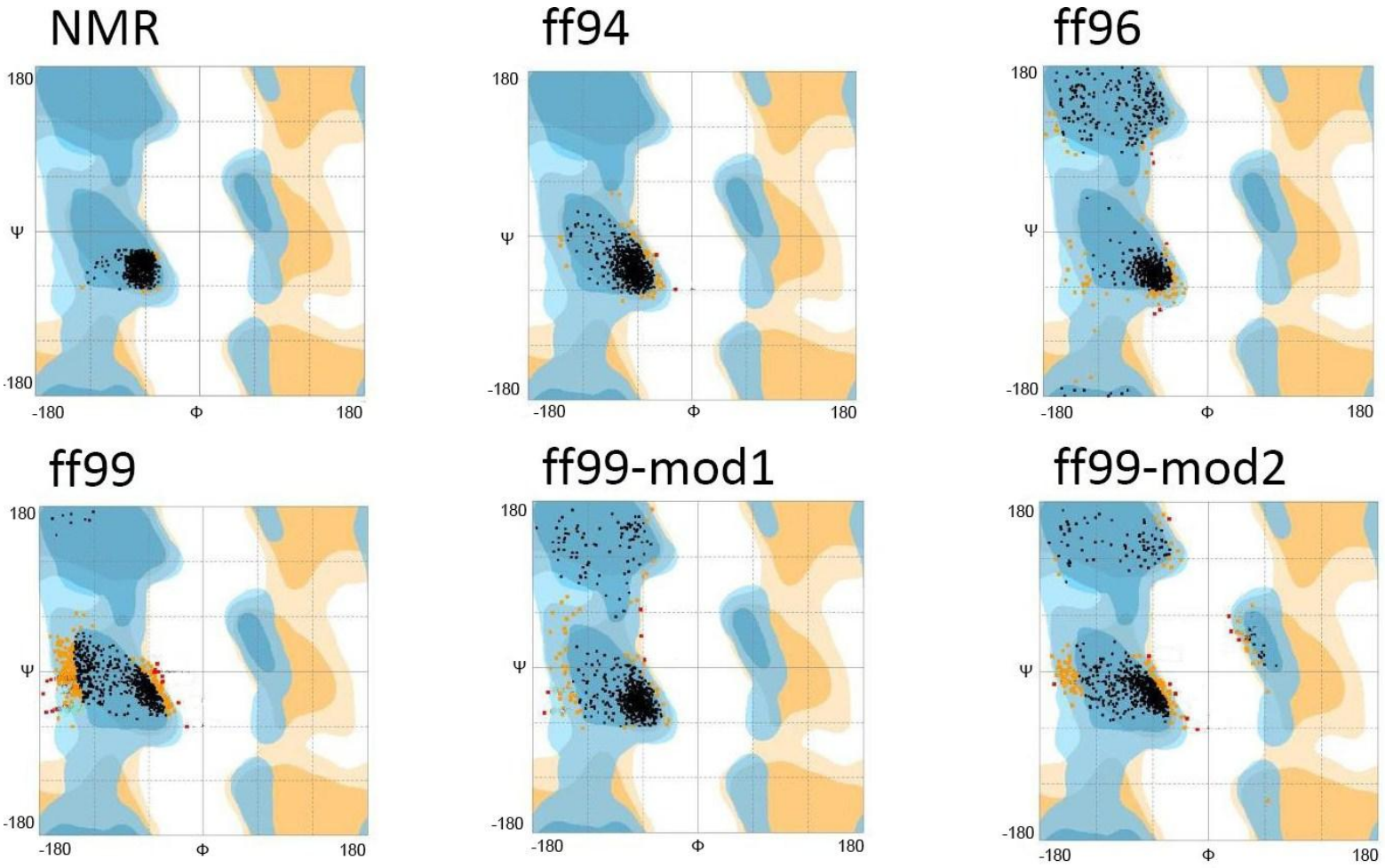


Figure 4.5 - Ramachandran plots created from twenty MD snapshots taken during the final 5ns of production simulation using the RAMAGE Ramachandran Plot Assessment. Light blue regions show areas where amino acid occupancy is allowed while dark blue regions show regions where amino acid occupancy is favored according to a sampling of the Protein Data Bank. (Orange and green regions show favored and allowed occupancy for glycine and proline residues, respectively.)

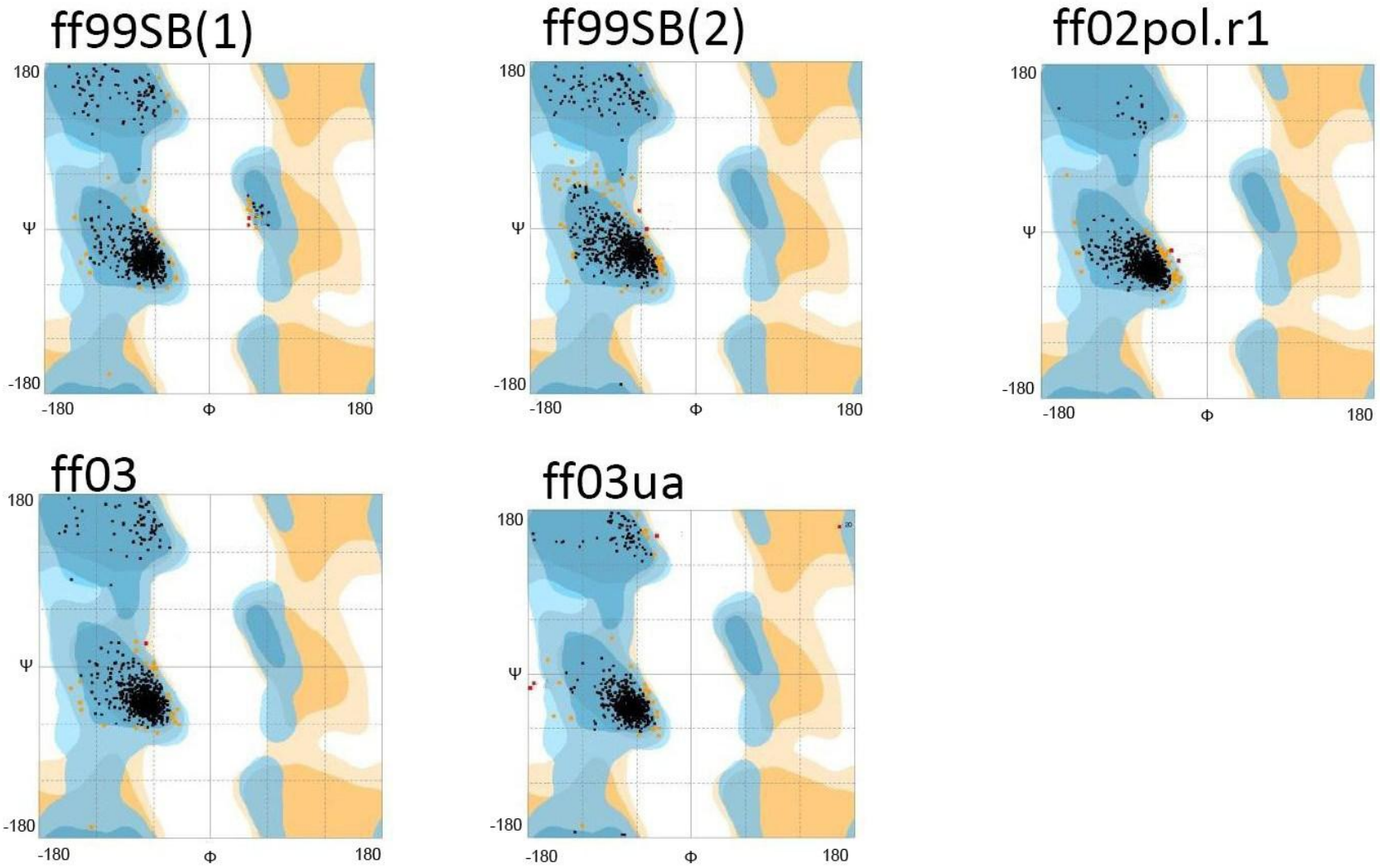


Figure 4.5 Continued

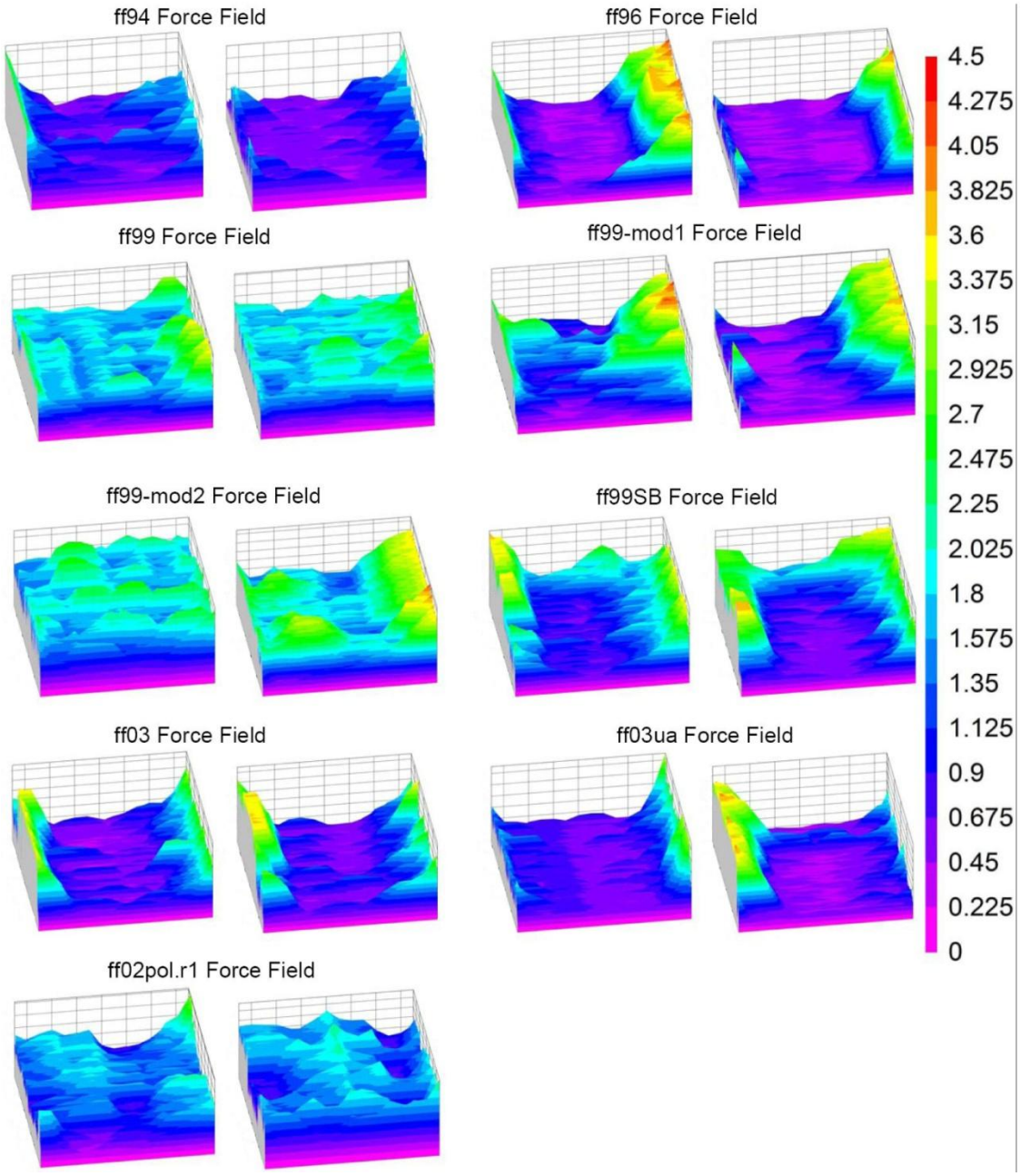


Figure 4.6 - Scanning RMSD of force field models showing regions at variance from the averaged NMR structure. Shown are both the monomer A and monomer B regions. Scanning RMSD measurements were calculated using snapshots taken every 100 ps. RMSD was calculated based on the backbone of seven amino acids starting initially with residues 1-7, then recalculated at residues 2-8, 3-9, 4-10, etc. Comparing these regions in a time course allows the visualization of structural distortions and entropic movements in the peptide structure. Time progression is shown along the X axis. Positioning along the primary sequence of the coiled coil is shown on the Y axis (N-terminal to C-terminal, left to right). RMSD of 7 residue regions (in angstroms) shown along the Z axis (enlarged color coded scale shown on right).

coiled-coils are known to be a flexible protein motif, especially in the terminal regions (68), we expect that this decreased movement is due to artificial stabilization of α-helical structures. In the models built using the ff99-mod2 and ff02pol.r1 force fields, an intertwined horseshoe structure is seen in the final snapshots. In a prior study we reported on the formation of this structural instability as a result of shortened salt bridge length (66). We speculated that this intertwined structure allows minimal exposure of hydrophobic residues to the aqueous environment. As secondary structure is correlated with association, an actual increase in helical character is seen in the final snapshots as this structure is formed and hydrophobic residues are stabilized. In the ff02pol.r1 model this is seen in significant α-helical character in the Ramachandran plots but poor coiled-coil tertiary interactions in the molecular snapshots.

Of the remaining nonpolarizable force fields, we see reasonable agreement between ff99SB, ff03, and united atom model ff03ua. These three force fields show RMSDs very similar to the NMR averaged structure in their interior regions while showing high flexibility in the terminal, unstructured residues (primarily in the N-terminal residues for ff99SB and ff03). The united atom model, ff03ua, differs from both all-atom models with respect to the flexibility of the terminal atoms. In the scanning RMSD measurements (see Figure 4.6), the N-terminus of the acidic monomer and C-terminus of the basic monomer are surprisingly rigid compared to the all-atom version of this force field. Examination of the snapshots suggests the denaturation in the terminal residues for these three force fields is dynamic. Uncoiling and loss of secondary structure can be followed by reformation of α-helical regions in the terminal four residues of both the N and C termini. It is likely this dynamic process, associated with uncoiling and recoiling of the terminal residues, prevents convergence on the time scales observed in the 2D-RMSD. This large movement resulting from the formation and loss of secondary structure is at odds with the NMR structure which assigns NOE-derived distance restraints through the terminal residues (51) suggesting that the occurrence or the extent of terminal unraveling is artificial. Ramachandran plots (Figure 4.5) of these three force fields show that loose and frayed termini occupy the β-sheet region of the Ψ/Φ dihedral space. As the simulation of the IAAL-E3/K3 coiled-coil continues with the ff99SB force field we see at 160 ns a sharp increase in the 1D-RMSD plot

showing a dramatic loss of secondary structure. Analysis of the structure shows that with time, structural destabilization extends from the terminal regions into the interior domain (see Figure 4.6). Additional simulations of IAAL-E3/K3 NMR models 2-5 using the ff99SB and ff03 force fields were pursued to determine the dependence of structure stability on the starting trajectory. It was found that independent simulations did not all converge; some models were more stable on the MD simulation time scales whereas other move significantly away from the NMR model geometries (see Figure 4.7). Simulations of models 2 and 3 using the ff03 force field were found to have lost structural orientation and have switched from a parallel to an anti-parallel orientation which is in conflict with experimental observation (51).

Comparing MD structural internal average distances to the NOE-derived distance restraints applied to the NMR models shows that significant deviations occur in the terminal four residues. To overcome this, NOE-based dihedral and distance restraints on the backbone were applied to the initial NMR model using ff99SB and ff03 force fields. While the restraints stabilized the structure and decreased RMSD variance to thermal fluctuations in the ff99SB model, the ff03 model quickly denatured, separated, and changed orientation from a parallel to an anti-parallel dimer (see Figure 4.8). Applying only the backbone torsional restraints to the ff99SB force field model resulted in a rapid onset of deformation which suggests that the structure denaturation must have an internal energy contribution apart from the backbone torsions. Under-parameterized or incorrectly parameterized terms would naturally move away from the initial NMR starting structure and lead to structural deformation. This should occur less in the models applying NOE-derived distance restraints. In the unrestrained models, a decrease in the dihedral energy was observed over time (see Figure 4.9) while the bond and angle energies remained constant or experienced a slight increase. Since the dihedral term had not equilibrated in 100 ns and the dihedral energies were still decreasing, this suggests that the dihedral energy is contributing to the instability of the force field model.

In proteins, the dihedral terms have contributions from both the main chain and side-chain torsions. Since addition of main-chain dihedral restraints to the Ψ and Φ angles did not by themselves prevent denaturation, the side-chain dihedrals became the focus as the source of

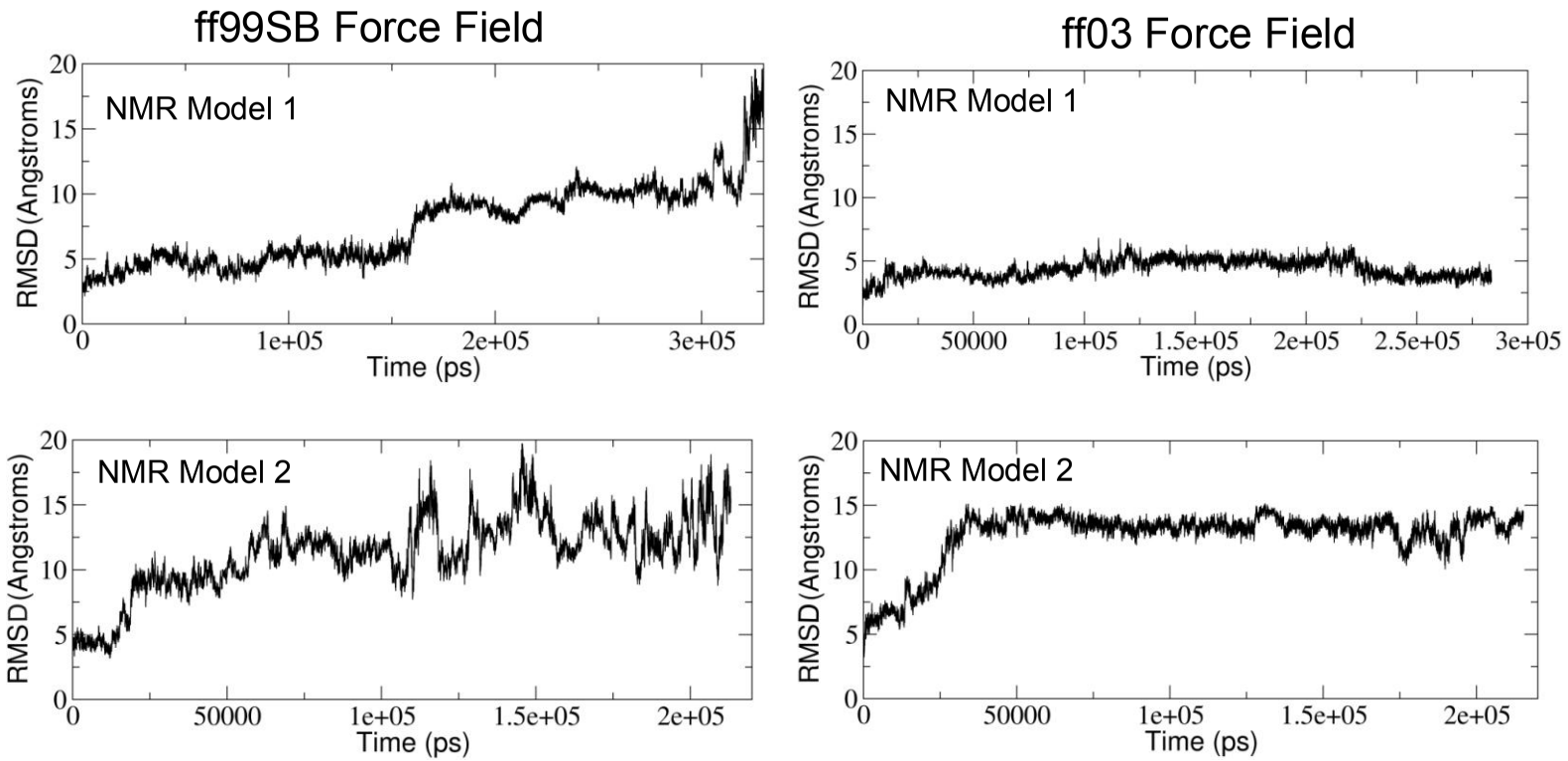


Figure 4.7 - One dimensional RMSD chart of ff99SB and ff03 force field models of the IAAL-E3/K3 coiled-coil from starting trajectories created using the initial 5 NMR models of the same protein. Significant differences in stability show a strong dependence on the initial start model.

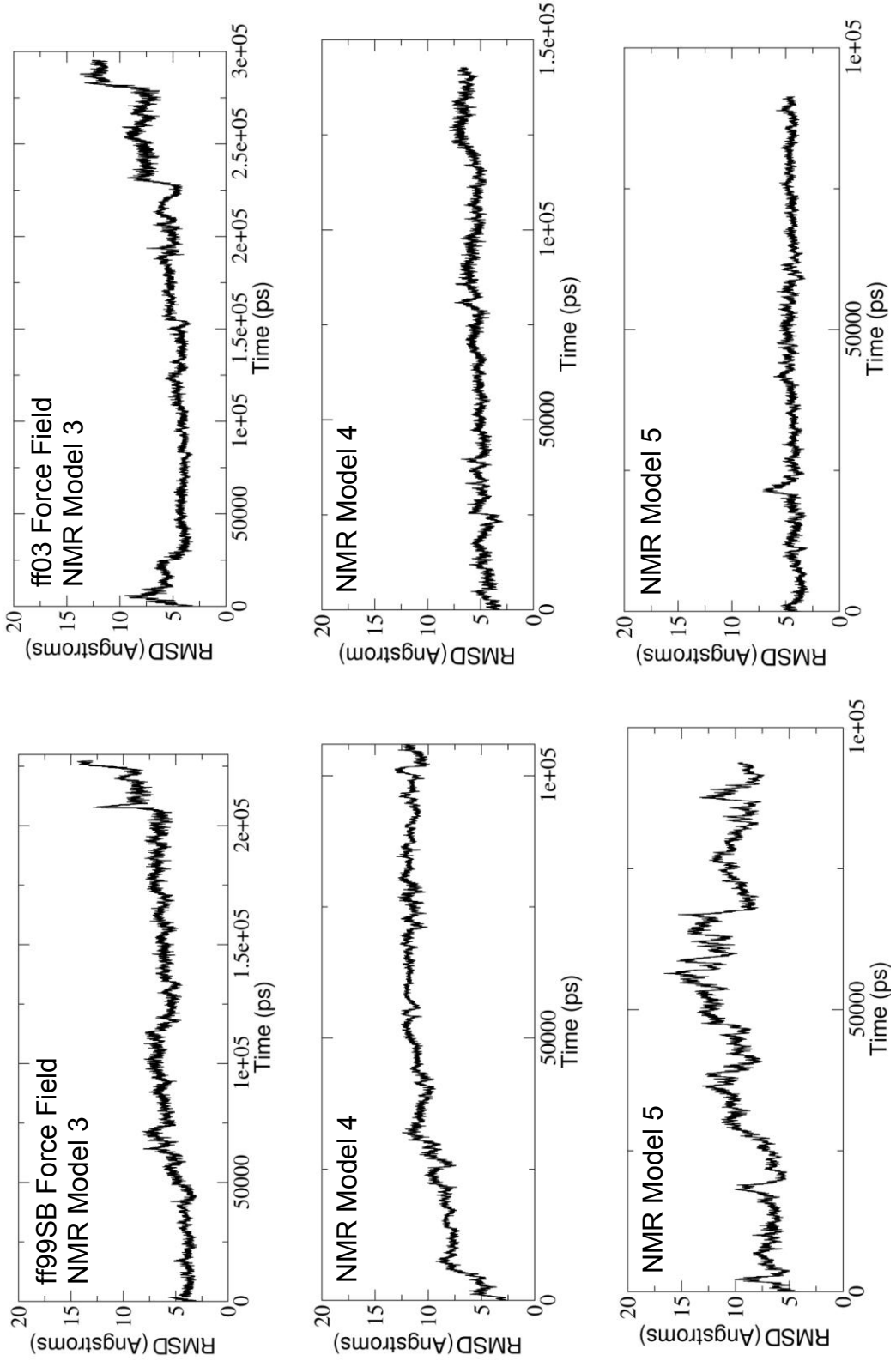


Figure 4.7 Continued

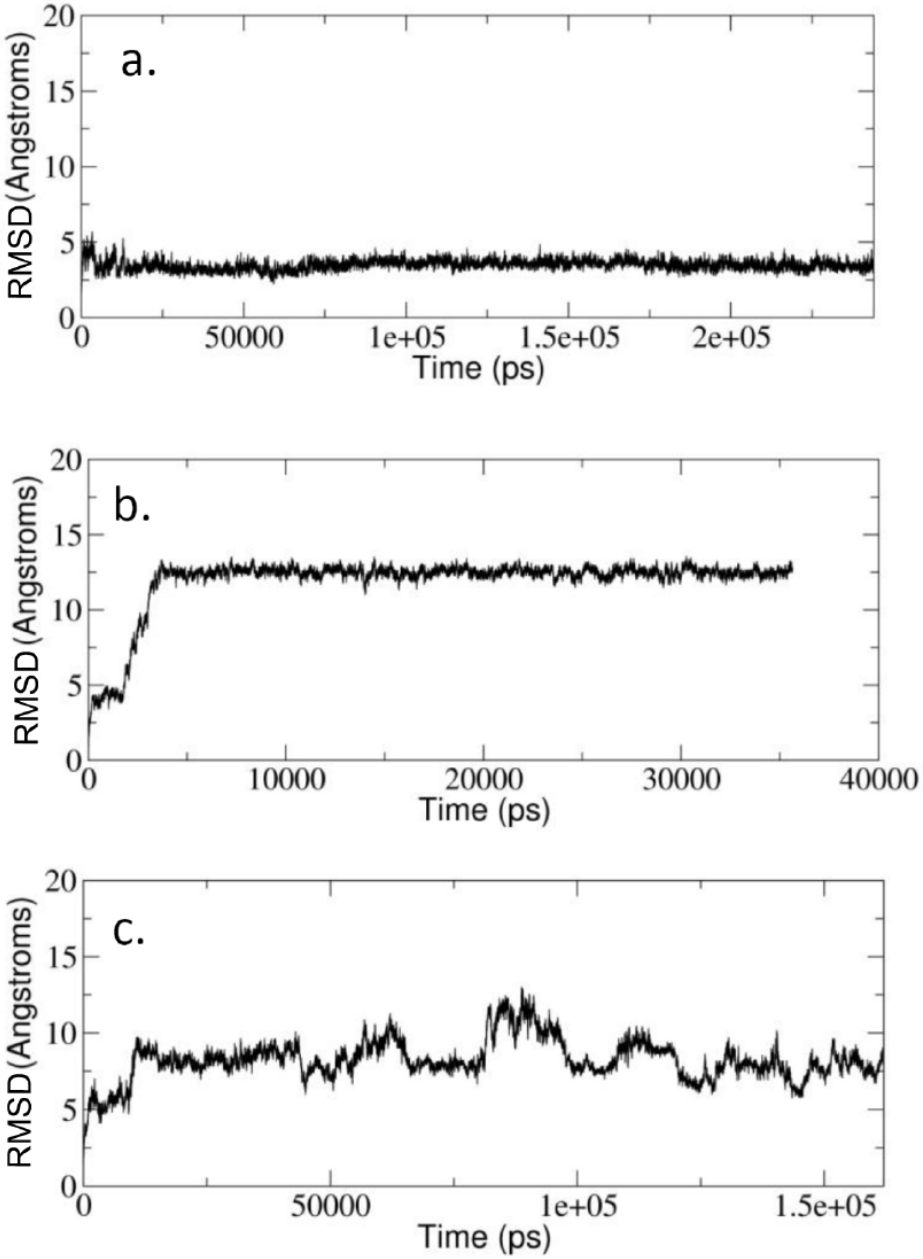


Figure 4.8 - One dimensional RMSD chart showing application of NOE-derived distance and main-chain dihedral restraints to the initial NMR structure as modeled by ff99SB (a) and ff03 (b). Reversal of dimer orientation is seen shortly after the initiation of MD in the ff03 model. When NOE-derived main-chain dihedral restraints were applied without distance restraints to the ff99SB model (c), a rapid structural deformation is observed.

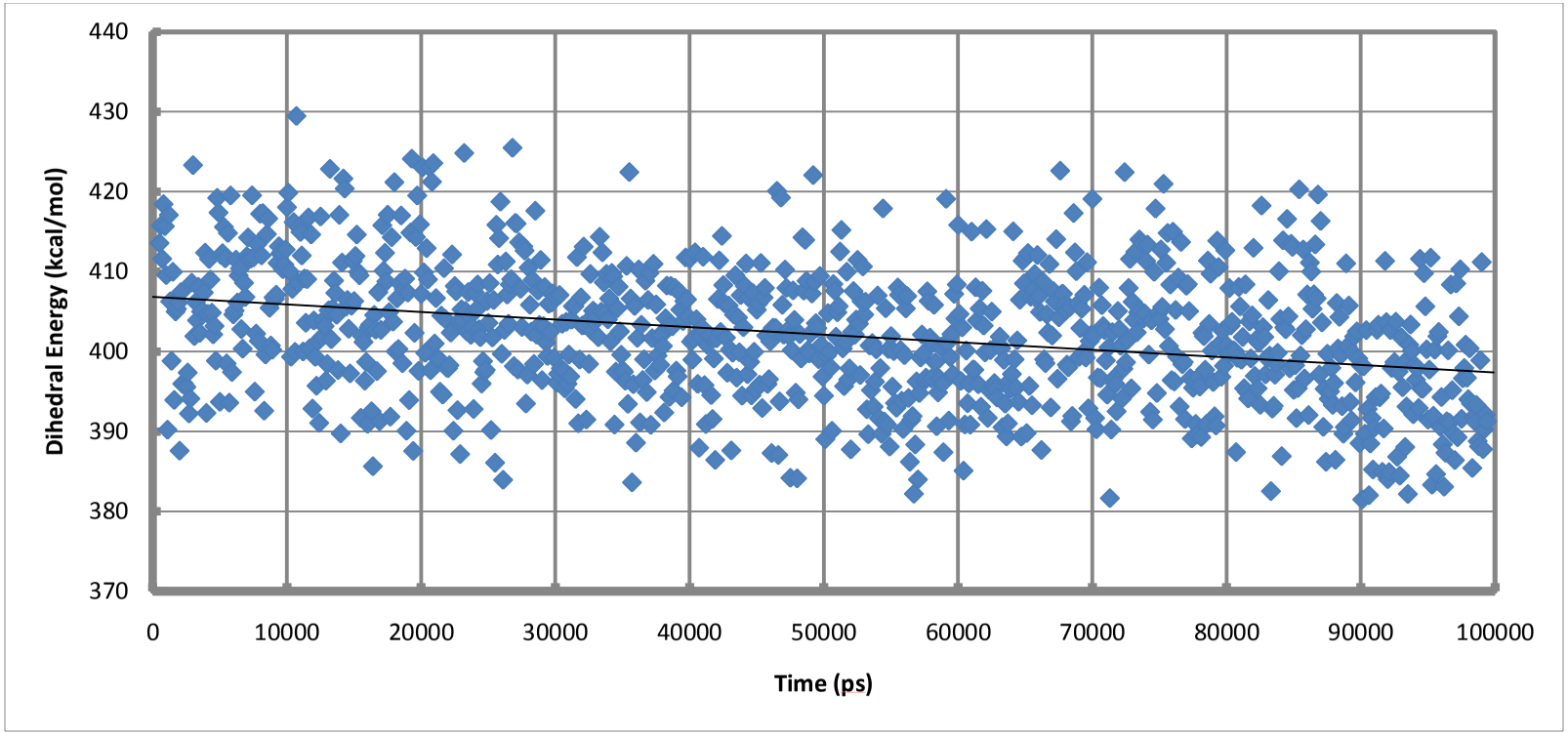


Figure 4.9 - A dihedral energy time course showing the sum of dihedral energy of the IAAL-E3/K3 coiled-coil as modeled using the ff99SB force field (without NOE restraints) during the first 100 ns of production MD simulation.

instability in ff99SB model. In the coiled-coil motif, ordered side-chains that contribute to the tertiary structure are primarily found in the hydrophobic core at positions **a** and **d**. In IAAL-E3/K3 these positions are occupied by leucine and isoleucine residues. Rotamer analysis of internal leucine and isoleucine side-chains using well-established definitions (65) found significant differences in the isoleucine rotamer sampling between the unrestrained and models where NOE-derived restraints were applied. In the acidic monomer (monomer A) during restrained sampling, isoleucine side chains showed high occupancy in the *gauche*+ (r1), *trans* (r2) rotamer pair with minor contributions in the gauche+, gauche+ rotamer pair. In the unrestrained models these rotamers were essentially unsampled in interior, unfrayed positions during the 100 ns monitored. This suggests that the isoleucine side chain dihedrals could be a significant contributor to the structural deformation seen when modeling IAAL-E3/K3. Note that since this model system only contains five amino acid types (Ala, Ile, Leu, Glu, and Lys), this model does not exclude the possibility that additional side chain biases may be evident with other amino acids.

Analysis of the structural models between restrained and unrestrained coiled-coil trajectories showed that application of intramolecular NOE restraints to the backbone of the ff99SB dimer resulted in a decrease in interhelical distance which is especially significant in the terminal regions of the coiled-coil dimer (see Table 4.2). The choice of isoleucine rotamers in the hydrophobic interior is a reflection of the energetic interplay between internal energetics (especially dihedral torsions), the hydrophobic effect, and the formation of van der Waals contacts between proximal nonpolar sidechains (packing). The decrease in interhelical distance, leads to a change of isoleucine rotamer choices (see Table 4.1) in order to maximize van der Waals contacts and minimize internal energy. Interestingly, the change in rotamers also leads to an increase in intrahelical distance measured as the distance between Cβ atoms of opposing isoleucine residues (see Figure 4.10) and maintains similar van der Waals energy between adjacent leucine residues (see Table 4.3). In the ff99SB model without the NOE restraints, isoleucine rotamer choices result in improved van der Waals energetics between opposing isoleucine residues, a widening of interhelical distance, and a consequential decrease in favorable van der Waals contacts outside of the hydrophobic core (see Table 4.3). While the

Table 4.1 - Isoleucine rotamers of IAAL-E3/K3 as modeled using the ff99SB force field. Side-chain dihedral angles were assigned to rotamer preferences every ps and occupancy was determined by averaging over a 20 ns bin. By comparing 20 ns average bin over a 100 ns sampling region, rotamer transitions were visualized. Rotameric differences between unrestrained models and models where NOE-derived restrained were applied showed significant differences with regard to the *gauche+, trans* rotamer pairs and trans, trans rotamer pairs.

Isoleucine Rotamers (Unrestrained)

	MonA - ILE2									MonA - ILE9									MonA - ILE16								
	gauche+,gauche+	gauche+,trans	gauche+,gauche-	trans,gauche+	trans,trans	trans,gauche-	gauche-,gauche+	gauche-,trans	gauche-,gauche-	gauche+,gauche+	gauche+,trans	gauche+,gauche-	trans,gauche+	trans,trans	trans,gauche-	gauche-,gauche+	gauche-,trans	gauche-,gauche-	gauche+,gauche+	gauche+,trans	gauche+,gauche-	trans,gauche+	trans,trans	trans,gauche-	gauche-,gauche+	gauche-,trans	gauche-,gauche-
20-40 ns	13	58	2	11	10	0.1	0	0	6	0	0	0	38	62	0	0	0	0	0	0	0	13	85	0.1	0.1	0.5	0.7
40-60 ns	8.6	91	0.7	0	0	0	0	0	0	0	0	0	7.1	93	0	0	0	0	0	0	0	18	66	0.1	0.3	13	3.1
60-80 ns	16	84	0	0	0	0	0	0	0	0	0	0	37	50	1.5	0.1	0	11	0	0	0	30	41	0.2	0.3	27	1.1
80-100 ns	6.3	93	0.7	0	0	0	0	0	0	0	0	0	23	77	0	0	0	0	0	0	0	31	55	0.1	1.2	12	0
100-120 ns	4.6	91	0.5	0	0	0	0	3	0	0	0	0	9.8	90	0	0	0	0	0	0	0	5.1	44	0.4	0.6	37	13

	MonB - ILE2									MonB - ILE9									MonB - ILE16								
20-40 ns	7.3	47	3	20	15	0.6	0	8	0	0	0	0	14	30	1	0.1	50	4.9	0	0	0	3.4	56	0.6	0.8	37	2.5
40-60 ns	12	61	1.7	5.5	13	0	0	6	0	0	0	0	3.6	86	2	0	8.5	0	0	0	0	12	59	0	1.8	27	0
60-80 ns	14	33	0.3	35	17	0	0	0	0	0.2	1.6	0	0	0.2	0	0.6	94	3.9	0	0	0	11	21	0.2	0	65	2.3
80-100 ns	1.3	0.3	0.1	57	26	1.7	3	3	7	0	0	0	7.6	52	0.1	0	37	3.3	0	0	0	5.7	40	0	0.2	50	4.1
100-120 ns	0.3	13	0	26	51	1.9	2	3	3	0	0	0	7.9	85	0.3	0	6.1	0.6	0	0	0	25	61	0	0.1	14	0

Table 4.1 Continued

Isoleucine Rotamers (NOE-based restraints)

	MonA - ILE2									MonA - ILE9									MonA - ILE16								
	gauche+,gauche+	gauche+,trans	gauche+,gauche-	trans,gauche+	trans,trans	trans,gauche-	gauche-,gauche+	gauche-,trans	gauche-,gauche-	gauche+,gauche+	gauche+,trans	gauche+,gauche-	trans,gauche+	trans,trans	trans,gauche-	gauche-,gauche+	gauche-,trans	gauche-,gauche-	gauche+,gauche+	gauche+,trans	gauche+,gauche-	trans,gauche+	trans,trans	trans,gauche-	gauche-,gauche+	gauche-,trans	gauche-,gauche-
20-40 ns	16	84	0.2	0	0	0	0	0	0	5	95	0	0	0	0	0	0	0	5.3	74	0	0	0	0	0.1	21	0.2
40-60 ns	22	76	1.8	0	0	0	0	0	0	6.9	93	0	0	0	0	0	0	0	1.9	17	0.1	1.5	29	0	0	51	0
60-80 ns	4.7	29	0.5	21	23	0	0	20	2	7	93	0	0	0	0	0	0	0	6.6	93	0	0	0	0	0	0	0
80-100 ns	0	0	0	3.9	13	0	0	3	79	2.2	40	0	1	7.3	0	0	50	0	4.4	96	0	0	0	0	0	0	0
100-120 ns	8.9	68	0.2	0	0	0	0	14	9	0	0	0	1	25	0.6	0	74	0	1.6	22	0	0	2.1	0.3	0	70	3.7

	MonB - ILE2									MonB - ILE9									MonB - ILE16								
20-40 ns	0	0	0	54	0	46	0	0	0	0	0	0	47	25	0.9	0	21	5.9	0	0	0	18	34	0.5	0.1	46	1.7
40-60 ns	0	0	0	52	0	48	0	0	0	0.1	0.8	0	30	41	1.1	0	26	0.2	0	0	0	24	53	0.3	0.1	21	2.7
60-80 ns	0	0	0	52	0	48	0	0	0	0	0	0	30	16	2.5	0.1	43	8.1	0	0	0	15	40	0.3	0.8	38	5.2
80-100 ns	0	0	0	71	0	29	0	0	0	0	0	0	13	52	1.7	0	26	6.6	0	0	0	0.6	3.3	0.7	0.3	66	29
100-120 ns	0	0	0	77	1.5	21	0	0	0	0	0	0	47	39	1.5	0	11	0.5	0	0	0	21	75	0.1	0	0.1	4.6

Table 4.2 - Isoleucine rotamers of IAAL-E3/K3 as modeled using the ff03 force field. Side-chain dihedral angles were assigned to rotamer preferences every ps and occupancy was determined by averaging over a 20 ns bin.

Isoleucine Rotamers (Unrestrained)

	MonA - ILE2									MonA - ILE9									MonA - ILE16								
	gauche+,gauche+	gauche+,trans	gauche+,gauche-	trans,gauche+	trans,trans	trans,gauche-	gauche-,gauche+	gauche-,trans	gauche-,gauche-	gauche+,gauche+	gauche+,trans	gauche+,gauche-	trans,gauche+	trans,trans	trans,gauche-	gauche-,gauche+	gauche-,trans	gauche-,gauche-	gauche+,gauche+	gauche+,trans	gauche+,gauche-	trans,gauche+	trans,trans	trans,gauche-	gauche-,gauche+	gauche-,trans	gauche-,gauche-
20-40 ns	5	25	0	31	24	0.4	0.1	7.2	7.27	0	0	0	2.4	8.8	1.4	0.6	75	12	0	0	0	3.7	46	0	0.7	48	1.6
40-60 ns	0	0	0	59	40	0.3	0	0	0.01	0	0	0	17	53	0	0	30	0	1.6	20	0	5.4	32	0	0.1	40	0.7
60-80 ns	0	0	0	48	44	0.4	0.1	2.8	4.3	0	0	0	0.1	0	0	0.1	92	7.4	0	0	0	0	0.5	0	0.2	93	6.2
80-100 ns	0	0	0	51	49	0.2	0	0	0.01	0	0	0	0.2	32	0	0.1	66	1.7	0	0	0	5.4	26	0.1	0.7	61	7
100-120 ns	14	63	4	11	5.1	0	0.3	3.3	0	0	0	0	2.4	41	0	1.1	48	7.5	0	0	0	5.9	8.9	0.5	4.1	70	11

	MonB - ILE2									MonB - ILE9									MonB - ILE16								
20-40 ns	7	0	0	91	1.2	0.1	0.2	0	0.51	0	0	0	24	45	0.6	0	27	3	0.3	3.5	0	11	39	0	0.3	46	0.2
40-60 ns	0	0	0	94	6.1	0.1	0	0	0.01	0	0	0	21	45	0	1.4	33	0.5	0	0	0	18	64	0	0	17	0
60-80 ns	0	0	2	67	4.2	25	1.2	0	1.14	0	0	0	53	47	0	0	0	0	0	0	0	29	64	1.3	0	0.9	4.3
80-100 ns	30	52	1	1.9	15	0.1	0	0	0	0	0	0	46	54	0.3	0	0	0	0	0	0	21	75	0.1	0	0.2	4.2
100-120 ns	10	2	1	45	10	18	5.5	0.2	8.5	0	0	0	44	45	0.4	0.2	10	0.3	0	0	0	26	49	0.1	0	5.5	19

Table 4.2 Continued

Isoleucine Rotamers (NOE-based restraints)

	MonA - ILE2									MonA - ILE9									MonA - ILE16								
	gauche+,gauche+	gauche+,trans	gauche+,gauche-	trans,gauche+	trans,trans	trans,gauche-	gauche-,gauche+	gauche-,trans	gauche-,gauche-	gauche+,gauche+	gauche+,trans	gauche+,gauche-	trans,gauche+	trans,trans	trans,gauche-	gauche-,gauche+	gauche-,trans	gauche-,gauche-	gauche+,gauche+	gauche+,trans	gauche+,gauche-	trans,gauche+	trans,trans	trans,gauche-	gauche-,gauche+	gauche-,trans	gauche-,gauche-
20-40 ns	11	87	2	0	0	0	0	0	0	0	0	0	1.7	5.5	0.1	0.2	82	11	0.1	1.5	0	0.3	0.4	0	0.4	89	8.8
40-60 ns	10	74	1	0	0	0	0	0	15.5	0	0	0	2	6	0	0.1	84	7.9	0	0	0	0.8	0.5	0	0.3	78	20
60-80 ns	12	81	2	0	0	0.1	0	1.7	3.48	0	0	0	0	4.4	0	0	87	8.9	1.2	34	0	0	0.2	0	0	39	26
80-100 ns	3	64	0	3.9	3.1	0	1.4	22	1.24	2.8	75	0	0	0	0	0	22	0	3.2	60	0	2.5	4.1	0	0	29	1.2
100-120 ns	4	91	0	0	0	0	0.5	4.5	0	3.3	97	0	0	0	0	0	0	0	0	0	0	1.9	4.5	0	0.6	65	28

	MonB - ILE2									MonB - ILE9									MonB - ILE16								
20-40 ns	0	0	0	34	66	0	0	0	0	0	0	0	0	0.6	0	2.3	93	3.8	12	88	0	0	0	0	0	0	0
40-60 ns	0	0	0	27	72	0	0	0	0	0	0	0	1.2	0.1	0	6.2	80	12	9.4	91	0	0	0	0	0	0	0
60-80 ns	0	0	0	33	36	31	0	0	0	0	0	0	0.1	1.2	0	0.2	91	7.7	8.1	92	0	0	0	0	0	0	0
80-100 ns	0	0	0	39	4.2	56	1.2	0	0	0	0	0	0	1.5	0.1	1	76	21	12	88	0	0	0	0	0	0	0
100-120 ns	0	0	0	36	1.3	59	2.5	0.4	0	0	0	0	0.9	1.8	0	2	62	33	9.2	91	0	0	0	0	0	0	0

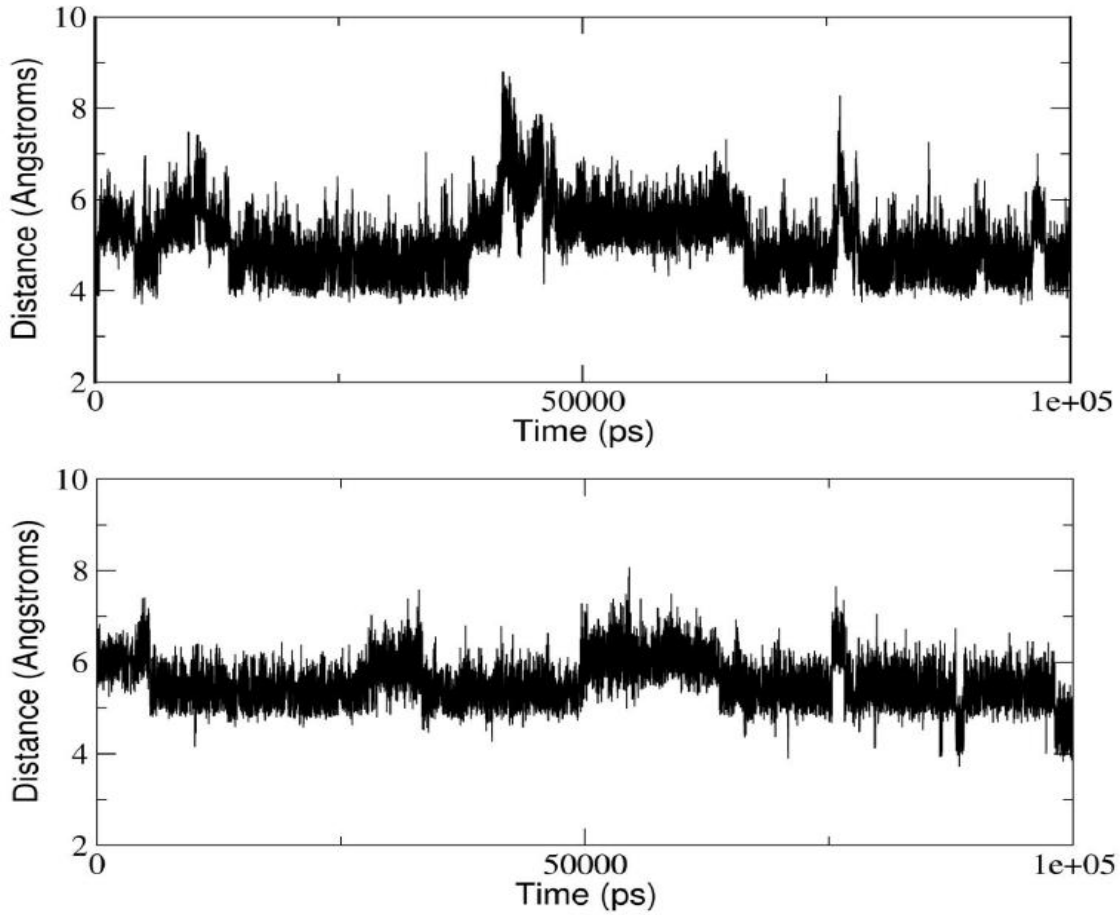


Figure 4.10 - Intrahelical distance between the Cβ atoms of Ile9 of Monomer A and Monomer B for the unrestrained (top) and restrained (bottom) models of the IAAL-E3/K3 coiled-coil simulated using the ff99SB force field. As assessed, the unrestrained model average distance was calculated to be 5.05 Å (stdev. 0.66 Å) and the restrained model average was calculated at 5.47 Å (stdev. 0.45 Å).

Table 4.3 - Overview of distance measurements from NOE restrained and unrestrained models of the IAAL-E3/K3 coiled coil as simulated using the ff99SB and ff03 force fields. Distance measurement were taken as an average over a 100 ns sampling region and compared to fourteen selected NOE distances chosen to span the entire reference molecule.

	Interactions		NMR	ff99SB		ff99SB (NOE restraints)		ff03		ff03 (NOE restraints)	
	Pair1	Pair2	NOE Distance (Å)	Distance (Å)	Stdev (Å)	Distance (Å)	Stdev (Å)	Distance (Å)	Stdev (Å)	Distance (Å)	Stdev (Å)
Acidic Monomer	3 Ala -Hα	6 Glu -HN	4.36	5.21	1.94	3.47	0.32	4.93	0.33	3.58	0.30
	4 Ala -Hα	7 Lys -HN	3.47	4.04	0.70	3.36	0.21	3.47	0.39	3.39	0.20
	6 Glu -Hα	8 Glu -HN	4.63	4.43	0.33	4.18	0.26	4.42	0.29	4.31	0.25
	8 Glu -Hα	10 Ala -HN	4.46	4.37	0.29	4.30	0.21	4.56	0.25	4.33	0.19
	11 Ala -Hα	14 Lys -HN	3.71	3.59	0.33	3.49	0.23	3.52	0.30	3.43	0.24
	14 Lys -Hα	17 Ala -HN	3.86	3.49	0.36	3.42	0.28	3.45	0.34	3.43	0.26
	16 Ile -Hα	19 Leu -HN	4.31	3.55	0.52	3.27	0.30	3.43	0.48	3.15	0.25
Basic Monomer	2 Ile -Hα	5 Leu -HN	4.64	7.00	1.53	3.42	0.31	5.52	2.23	3.38	0.32
	6 Lys - Hα	10 Ala -HN	4.85	4.31	0.41	4.27	0.31	4.41	0.45	4.11	0.26
	7 Glu -Hα	10 Ala -HN	3.93	3.49	0.32	3.56	0.28	3.67	0.32	3.43	0.27
	10 Ala -Hα	13 Lys -HN	4.90	3.48	0.27	3.49	0.26	3.47	0.29	3.43	0.25
	12 Leu -Hα	14 Glu -HN	4.70	4.28	0.29	4.29	0.26	4.26	0.30	4.21	0.28
	13 Lys -Hα	15 Lys -HN	5.00	4.37	0.31	4.35	0.30	4.53	0.29	4.48	0.26
	16 Ile -Hα	19 Leu -HN	4.94	3.57	0.45	3.49	0.30	3.66	0.68	3.31	0.28

overall van der Waals energetics are similar in the restrained and unrestrained models, the number of surface contacts in the folded coiled-coil have decreased in the unrestrained model. This is reflected in a rate of increase in nonpolar surface area of 0.52 $Å^2$ per ns over the first 100 ns of simulation time.

In the N-terminal region, the terminal solvent-exposed isoleucine residues shift towards their force field energetic minimum and adopt the *gauche*+, *trans* rotamer. A difference between rotamer choices of the restrained and unrestrained models is evident in the terminal isoleucine residue in the basic monomers.

Two distinct populations exist in the basic monomer of the restrained model at position 2: *trans*, *gauche*+ and *trans*, *gauche*-. The intra-helical measurements from Table 4.2 show that the N-terminal region maintains NOE distances while retaining van der Waals contacts (Table 4.3). The close proximity of isoleucine residues is forced by the distance restraints. This proximity in turn leads to contact between opposing isoleucine residues and the contacts drive a change in rotamers in order to minimize energy and create favorable packing; the frequent contact in the N-terminus maintains the *trans*, *gauche*- isoleucine rotamer population. The unrestrained model showed a strong basic N-terminal isoleucine rotamer population in the *gauche*+, *trans* population as well as minor populations in the *trans*, *gauche*+ and *trans*, *trans* rotamers. The frequent separation of the termini, in the unrestrained model, drove the rotamers towards their force field energetic minimum (*gauche*+, *trans*) while the occasional contact is reflected in the maintenance of the *trans*, *gauche*+ and *trans*, *trans* populations.

In the unrestrained model, the changes in interior isoleucine rotamer populations lead to an increased interhelical distance and decreased surface contacts outside of the hydrophobic core, weakening surface cohesion and refolding. The combination of poor surface contacts due to side chain torsion biases combined with the high entropy inherent in coiled-coils (69) lead to additional fraying from the termini inward. Free energy decomposition of the first 100 ns of production simulation showed that in the unrestrained model the van der Waals energetics consistently increased (becoming less stable) for the first six residues at the N-terminus of the acidic monomer and the last four residues of the C-terminus of acidic monomer. For the basic

monomer this increased to the first nine residues of N-terminus which lead to an increase in van der Waals energy and final five residues of the C-terminus engaged in nonclassical folding and interactions. At 160 ns, structural deformation in the dimer occurs. At 324 ns, the coiled-coil monomers completely separate.

The ff03 and ff99SB isoleucine rotamer parameters vary due to the additional parameterization of CT-CT-C-N dihedral to the ff03 parameter set. This dihedral affects both the isoleucine and leucine side-chain torsions and this results in significant differences compared to the ff99SB models. Both the unrestrained ff99SB and ff03 models show NOE violations in the terminal regions. The ff03 models show a lack of specificity in regard to coiled-coil orientation as seen with the simulations of NMR models 2 and 3. In the frayed, solvent-exposed isoleucine termini, the sidechains show a preference for *trans*, *gauche*+ and *trans*, *trans* rotamers (compared to the *gauche*+, *trans* population of ff99SB, see Tables 4.1 and 4.4). In the ff99SB models, the restraints forced a change in internal energetics and rotamers which in turn lead to an increase in intrahelical distance. In the ff03 models, the change of internal energetics to compensate for the additional restraint energy induced an orientation change in the dimer which can be visualized by the immediate increase in RMSD from the initial frame (see Figure 4.8b).

Application of the NOE restraints exposes biases in the isoleucine side-chain rotamers for both ff99SB and ff03 force fields. NOE distance restraints changed the isoleucine rotamer populations in order to maintain favorable hydrophobic contacts. The distance restraints in the ff99SB model led to improved binding at the N-terminus and along the helix which prevented separation of the monomers in our observed timescales and constrained the isoleucine side chains more in line with the experimental (and natural) structure.

Conclusions

Early work with force field parameterization focused on improving secondary structure behavior in proteins by adjusting internal backbone dihedral parameters, for example by tuning phi/psi with AMBER or adding 2D-spline corrections in CHARMM. In this paper, we examined the successes and failures of modern AMBER molecular dynamics force fields on the highly

Table 4.4 - Decomposition of van der Waals energetics (in kcal/mol) of the IAAL-E3/K3 coiled coil as modeled using the ff99SB force field with and without NOE dihedral and distance restraints. Energetic terms were separated into dimer specific (top) and van der Waal interactions between the dimers (second to top). Contacts between opposing isoleucine residues (second to bottom), and van der Waals between bounding leucine residues of isoleucine pairs (bottom) are also seen. Values shown show the average energy and standard deviation as measured during a stable 100 ns sampling region.

a. Interhelical van der Waals Energetics

Force Field Model	MonA Energy (kcal/mol)	Stdev	MonB Energy (kcal/mol)	Stdev
ff99SB	-27.12	7.06	-28.54	7.10
ff99SB (NOE restraints)	-34.32	6.01	-36.67	6.48

b. Intrahelical van der Waals Energetics

Force Field Model	Energy (kcal/mol)	Stdev
ff99SB	-45.80	5.15
ff99SB (NOE restraints)	-45.98	5.90

c. Isoleucine - Isoleucine van der Waals Energetics

Force Field Model	ILE2 (MonA)- ILE2 (MonB)	Stdev	ILE9 (MonA)- ILE9 (MonB)	Stdev	ILE16 (MonA)- ILE16 (MonB)	Stdev
ff99SB	-0.19	0.46	-2.64	0.71	-2.38	0.71
ff99SB (NOE restraints)	-1.14	0.68	-1.88	0.66	-1.80	0.65

d. Isoleucine - Leucine van der Waals Energetics

Force Field Model	ILE2 (MonA)- ILE2 (MonB)	Stdev	ILE9 (MonA)- ILE9 (MonB)	Stdev	ILE16 (MonA)- ILE16 (MonB)	Stdev
ff99SB	-3.68	1.19	-12.04	1.40	-11.25	1.56
ff99SB (NOE restraints)	-5.84	1.32	-12.05	1.24	-11.51	1.44

stable IAAL-E3/K3 coiled-coil and examined causes for failure to reproduce published secondary structure.

We found that MD models of the IAAL-E3/K3 coiled-coil structure reproduced known biases in several of the force fields tested. Models of IAAL-E3/K3 modeled using the ff94 force field were rigidly α-helical. Ramachandran plots of the ff96 model showed more β-sheet character than any other force field. In the literature the ff99 force field has been reported to have an α-helical bias. Secondary structure analysis of the IAAL-E3/K3 coiled-coil modeled using the ff99 force field showed a transition from classical α-helix to π-helix structure in the interior domain. Interior unraveling and loss of secondary structure on a relatively short time frame was seen with the ff99, ff99-mod1, ff99-mod2, and ff02pol.r1 force fields. In the ff99 and ff99-mod1 force fields the structural deformation begins in the termini and spreads to the interior. In the ff99-mod2 and ff02pol.r1 force field models, variance in the interior regions from the NMR model is clearly seen from the beginning of their trajectories. The force fields ff99SB, ff03, and ff03ua manage to model coiled-coil secondary and tertiary structure reasonably well on short time scales. In these models some uncoiling and loss of secondary structure is evident in the four terminal residues of the N and C termini of both monomers and dynamic folding and unfolding is seen in this region throughout the entire simulation. At approximately 160 ns the IAAL-E3/K3 model using the ff99SB force field undergoes significant secondary structure loss that appears to have originated with terminal unfolding events. Comparison of the ff99SB and ff03 models with NOE-derived distance restraints show violations at the coiled-coil termini. When these NOE intramolecular distance and dihedral main-chain restraints were applied to the force field models, the ff99SB model showed stabilization of secondary structure and RMSD measurements within the range of thermal fluctuations. Application of the restraints to the ff03 model resulted in a change of orientation from parallel to anti-parallel, consistent with the loss of orientation specificity seen in the unrestrained models. Further analysis of the ff99SB model using only main-chain dihedral restraints showed structural deformation suggesting that energetic sources for the destabilization of the model were not limited to Ψ and Φ parameters. Examination of internal energetic contributions of bond, angle, and dihedral terms over 100 ns show a decline in energy

of the dihedral term without reaching a minimum prior to structural deformation. Having ruled out main-chain contributions with the application of main-chain NOE restraints, we considered ordered side-chain contributions to the instability. Rotamer analysis of hydrophobic isoleucine side chains show significant rotameric differences between unrestrained models and models with applied NOE-derived restraints. Because of the large number of isoleucine residues in the hydrophobic core, a misbalance in the parameterization of the isoleucine side-chains could account for deformation and eventual separation of the dimers. Similarly, the successes of molecular dynamics approaches for other coiled-coils models may be explained by the rarity of isoleucine residues in their primary sequence. Analysis of the effects of the NOE restraints in the ff99SB model show that distance restraints lead to changes in isoleucine side-chain rotamer populations in order to decrease interhelical distance and optimize van der Waals contacts. In the N-terminus, NOE restraints forced additional contact between the fraying hydrophobic residues and this improved binding and prevented dimer separation. Future improvements to both the ff99SB and ff03 force fields with regard to isoleucine sidechains will be necessary to address structural differences between the computational models and the published experimental structure and the IAAL-E3/K3 is likely a good model system to assess force field improvements.

References

1. Schueler-Furman O, Wang C, Bradley P, Misura K, Baker D. Progress in modeling of protein structures and interactions. Science 2005; 310: 638-642.

2. Ponder J, Case D. Force fields for protein simulations. Advances in Protein Chemistry 2003; 66: 27-86.

3. Halgren T, Damm W. Polarizable force fields. Current Opinion in Structural Biology 2001; 11: 236-242.

4. Hansson T, Oostenbrink C, van Gunsteren W. Molecular dynamics simulations. Current Opinion in Structural Biology 2002; 12: 190-196.

5. van Gunsteren W, Bakowies D, Baron R, et al. Biomolecular modeling: goals, problems, perspectives. Angewandte Chemie-International Edition 2006; 45: 4064-4093.

6. Mackerell A. Empirical force fields for biological macromolecules: overview and issues. Journal of Computational Chemistry 2004; 25: 1584-1604.

7. Case DA, Cheatham TE, 3rd, Darden T, et al. The Amber biomolecular simulation programs. Journal of Computational Chemistry 2005; 26: 1668-1688.

8. Cornell WD, Cieplak P, Baylyl CI, et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. Journal of the American Chemical Society 1995; 117: 5179-5197.

9. Jorgensen W, Tirado-Rives J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. Journal of the American Chemical Society 1988; 110: 1657-1666.

10. Bayly C, Cieplak P, Cornell W, Kollman P. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. Journal of Physical Chemistry B 1993; 97: 10269-10280.

11. Cieplak P, Cornell W, Bayly C, Kollman P. Application of the multimolecule and multiconformational resp methodology to biopolymers: charge derivation for DNA, RNA, and proteins. Journal of Computational Chemistry 1995; 16: 1357-1377.

12. Garcia AE, Sanbonmatsu KY. Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds. Proceedings of the National Academy of Sciences, USA 2002; 99: 2782-2787.

13. Kollman PA, Dixon R, Cornell W, Fox T, Chipot C, Pohorille A. The development/application of a 'minimalist' organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data. In: Wilkinson A, Weiner P, van Gunsteren WF, editors. Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications. Dordrecht: Kluwer Academic Publishers; 1997. p. 83-96.

14. Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? Journal of Computational Chemistry 2000; 21: 1049-1074.

15. Wang L, Duan Y, Shortle R, Imperiali B, Kollman PA. Study of the stability and unfolding mechanism of BBA1 by molecular dynamics simulations at different temperatures. Protein Science 1999; 8: 1292-1304.

16. Ono S, Kuroda M, Higo J, Nakajima N, Nakamura H. Calibration of force-field dependency in free energy landscapes of peptide conformations by quantum chemical calculations. Journal of Computational Chemistry 2002; 23: 470-476.

17. Kamiya N, Higo J, Nakamura H. Conformational transition states of a beta-hairpin peptide between the ordered and disordered conformations in explicit water. Protein Science 2002; 11: 2297-2307.

18. Higo J, Ito N, Kuroda M, Ono S, Nakajima N, Nakamura H. Energy landscape of a peptide consisting of alpha-helix, 3(10)-helix, beta-turn, beta-hairpin, and other disordered conformations. Protein Science 2001; 10: 1160-1171.

19. Okur A, Strockbine B, Hornak V, Simmerling C. Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins. Journal of Computational Chemistry 2003; 24: 21-31.

20. Sorin EJ, Pande VS. Empirical force-field assessment: The interplay between backbone torsions and noncovalent term scaling. Journal of Computational Chemistry 2005; 26: 682-690.

21. Sorin EJ, Pande VS. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. Biophysical Journal 2005; 88: 2472-2493.

22. Simmerling C, Strockbine B, Roitberg AE. All-atom structure prediction and folding simulations of a stable protein. Journal of the American Chemical Society 2002; 124: 11258-11259.

23. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 2006; 65: 712-725.

24. Duan Y, Wu C, Chowdhury S, et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. Journal of Computational Chemistry 2003; 24: 1999-2012.

25. Shell MS, Ritterson R, Dill KA. A test on peptide stability of AMBER force fields with implicit solvation. Journal of Physical Chemistry B 2008; 112: 6878-6886.

26. Zhou R. Free energy landscape of protein folding in water: Explicit vs. implicit solvent. Proteins: Structure, Functions, and Genetics 2003; 53: 148-161.

27. Lwin TZ, Luo R. Force field influences in beta-hairpin folding simulations. Protein Science 2006; 15: 2642-2655.

28. Lee M, Duan Y. Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent model. Proteins: Structure, Function, and Genetics 2004; 55: 620-634.

29. Best R, Buchete N, Hummer G. Are current molecular dynamics force fields too helical? Biophysical Journal 2008; 95: 7-9.

30. Wickstrom L, Okur A, Simmerling C. Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. Biophysical Journal 2009; 97: 853-856.

31. Fawzi N, Phillips A, Ruscio J, Doucleff M, Wemmer D, Head-Gordon T. Structure and dynamics of the A 21–30 peptide from the interplay of NMR experiments and molecular simulations. Journal of the American Chemical Society 2008; 130: 6145-6158.

32. Oostenbrink C, Villa A, Mark A, Van Gunsteren W. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. Journal of Computational Chemistry 2004; 25: 1656-1676.

33. Oostenbrink C, Soares T, van der Vegt N, van Gunsteren W. Validation of the 53A6 GROMOS force field. European Biophysics Journal 2005; 34: 273-284.

34. Camilloni C, Provasi D, Tiana G, Broglia R. Exploring the protein G helix free-energy surface by solute tempering metadynamics. Proteins: Structure, Function, and Bioinformatics 2008; 71: 1647-1654.

35. Zagrovic B, Gattin Z, Lau J, Huber M, van Gunsteren W. Structure and dynamics of two ß-peptides in solution from molecular dynamics simulations validated against experiment. European Biophysics Journal 2008; 37: 903-912.

36. Cao Z, Lin Z, Wang J, Liu H. Refining the description of peptide backbone conformations improves protein simulations using the GROMOS 53A6 force field. Journal of Computational Chemistry 2009; 30: 645 - 660.

37. Kaminski G, Friesner R, Tirado-Rives J, Jorgensen W. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. Journal of Physical Chemistry B 2001; 105: 6474-6487.

38. Rizzo R, Tirado-Rives J, Jorgensen W. Estimation of binding affinities for HEPT and nevirapine analogues with HIV-1 reverse transcriptase via Monte Carlo simulations. Journal of Medicinal Chemistry 2001; 44: 145-154.

39. Jacobson M, Kaminski G, Friesner R, Rapp C. Force field validation using protein side chain prediction. Journal of Physical Chemistry B - Condensed Phase 2002; 106: 11673-11680.

40. Jorgensen W, Ulmschneider J, Tirado-Rives J. Free energies of hydration from a generalized Born model and an ALL-atom force field. Journal of Physical Chemistry B 2004; 108: 16264-16270.

41. Felts A, Gallicchio E, Wallqvist A, Levy R. Distinguishing native conformations of proteins from decoys with an effective free energy estimator based on the OPLS all-atom force field and the Surface Generalized Born solvent model. Proteins: Structure, Function, and Genetics 2002; 48: 404-422.

42. MacKerell Jr A, Feig M, Brooks 3rd C. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. Journal of Computational Chemistry 2004; 25: 1400-1415.

43. MacKerell Jr A, Feig M, Brooks III C. Improved treatment of the protein backbone in empirical force fields. Journal of the American Chemical Society 2004; 126: 698-699.

44. Buck M, Bouguet-Bonnet S, Pastor R, MacKerell A. Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme. Biophysical Journal 2006; 90: 36-38.

45. Li X, Hassan S, Mehler E. Long dynamics simulations of proteins using atomistic force fields and a continuum representation of solvent effects: Calculation of structural and dynamic properties. Proteins 2005; 60: 464-484.

46. Stavrakoudis A. Molecular dynamics simulations of an apoliprotein A–I derived peptide in explicit water. Chemical Physics Letters 2008; 461: 294-299.

47. Zagrovic B, Lipfert J, Sorin E, et al. Unusual compactness of a polyproline type II structure. Proceedings of the National Academy of Sciences, USA 2005; 102: 11698-11703.

48. Nilges M, Brunger AT. Successful prediction of the coiled coil geometry of the GCN4 leucine zipper domain by simulated annealing: comparison to the X-ray structure. Proteins 1993; 15: 133-146.

49. Landschulz W, Johnson P, McKnight S. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 1988; 240: 1759-1764.

50. Keating AE, Malashkevich VN, Tidor B, Kim PS. Side-chain repacking calculations for predicting structures and stabilities of heterodimeric coiled coils. Proceedings of the National Academy of Sciences, USA 2001; 98: 14825-14830.

51. Lindhout DA, Litowski JR, Mercier P, Hodges RS, Sykes BD. NMR solution structure of a highly stable de novo heterodimeric coiled-coil. Biopolymers 2004; 75: 367-375.

52. Yang L, Tan CH, Hsieh MJ, et al. New-generation amber united-atom force field. Journal of Physical Chemistry B 2006; 110: 13166-13176.

53. Wang ZX, Zhang W, Wu C, Lei H, Cieplak P, Duan Y. Strike a balance: optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. Journal of Computational Chemistry 2006; 27: 781-790.

54. Jorgensen W, Chandrasekhar J, Madura J, Impey R, Klein M. Comparison of simple potential functions for simulating liquid water. Journal of Chemical Physics 1983; 79: 926-935.

55. Joung IS, Cheatham TE, 3rd. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. Journal of Physical Chemistry B 2008; 112: 9020-9041.

56. Ryckaert J, Ciccotti G, Berendsen H. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics 1977; 23: 327-341.

57. Barth P, Schoeffler A, Alber T. Targeting metastable coiled-coil domains by computational design. Journal of the American Chemical Society 2008; 130: 12038-12044.

58. Essmann U, Perera L, Berkowitz M, Darden T, Lee H, Pedersen L. A smooth particle mesh Ewald method. Journal of Chemical Physics 1995; 103: 8577-8593.

59. Berendsen H, Postma J, van Gunsteren W, DiNola A, Haak J. Molecular dynamics with coupling to an external bath. Journal of Chemical Physics 1984; 81: 3684-3690.

60. Remler D, Madden P. Molecular dynamics without effective potentials via the Car-Parrinello approach. Molecular Physics 1990; 70: 921-966.

61. Sagui C, Pedersen L, Darden T. Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. The Journal of Chemical Physics 2004; 120: 73-87.

62. Car R, Parrinello M. Unified approach for molecular dynamics and density-functional theory. Physical Review Letters 1985; 55: 2471-2474.

63. Ramachandran G, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. Journal of Molecular Biology 1963; 7: 95-99.

64. Lovell SC, Davis IW, Arendall WB, 3rd, et al. Structure validation by C-alpha geometry: phi,psi and C-beta deviation. Proteins 2003; 50: 437-450.

65. Dunbrack R, Jr, Cohen F. Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Science 1997; 6: 1661-1681.

66. Pendley S, Yu Y, Cheatham III T. Molecular dynamics guided study of salt bridge length dependence in both fluorinated and non-fluorinated parallel dimeric coiled-coils. Proteins: Structure, Function, and Bioinformatics 2009; 74: 612 - 629.

67. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. Journal of Computational Chemistry 2004; 25: 1605-1612.

68. Zhou NE, Kay CM, Hodges RS. Disulfide bond contribution to protein stability: positional effects of substitution in the hydrophobic core of the two-stranded alpha-helical coiled-coil. Biochemistry 1993; 32: 3178-3187.

69. Dragan A, Privalov P. Unfolding of a leucine zipper is not a simple two-state transition. Journal of Molecular Biology 2002; 321: 891-908.

CHAPTER 5

EXPLICIT SOLVENT EFFECTS AND PARAMETERIZATION CORRECTIONS IN MM- PBSA CALCULATIONS

Abstract

MM-PBSA is a molecular dynamics post processing energy calculation method which has shown some promise in estimating binding and hydration free energies. Over the past two decades, this method has been applied to the calculation of solvation and binding free energies for many kinds of proteins, nucleic acids, and biomolecules. Our attempts to use the standard MM-PBSA framework to calculate the free energy of binding of parallel coiled-coil dimers were unsuccessful both in regards to the experimental measurements and free energy ranking of individual dimers. We suggest that the solvation terms were the source of error in the calculation. To further isolate the terms and approximations responsible for the errors, free energy of solvation estimates using MM-PBSA were compared to experimental measurements for several simple organic compounds. With these models, the errors appear to originate in the surface area approximation to the nonpolar contribution to the free energy of solvation. It was found that improvements to this measurement can be made by separating the van der Waals energetic estimation from the surface area approximation and including a polar surface area term to the nonpolar contributions to the solvation energy. Unfortunately, these approximations break down with complex alkanes containing two or more polar groups due to the complexity of solute-solvent interactions. An approach to correct these errors may be found by considering the replacement of surface area approximations with explicit measurements of solvent rotational and translational entropy.

Introduction

MM-PBSA is an energy post processing calculation approach first developed by Kollman and Case (1). This approach attempts to separate and calculate individual enthalpic and entropic contributions to the total free energy of a biomolecular system. It combines molecular mechanical energies (MM), a continuum solvent Poisson Boltzmann model (PB), a solvent accessible surface area term (SA) to calculate nonpolar contributions to the solvation free energy, and an estimate of the solute molecule's entropy using normal mode calculations or quasi-harmonic analysis (1). Many successful applications of the MM-PBSA approach for calculating relative free energies and binding free energies have been reported in the literature (1-9). MM-PBSA provides insights into the free energy of binding, which is a particularly challenging problem for other free energy techniques (10). In principle MM-PBSA is more applicable than many other approaches to the calculation of free energy because it can be applied to any two conformations or states of the same molecule and considers only differences between two endpoints without the need to construct intermediates along an energetic pathway (11). Alternative combination approaches to calculating free energy include Linear Interaction Energy (LIE) methods and ES/IS approaches. LIE methods are typically used to calculate the free energy of binding and consider atom by atom energy decomposition of the molecule and its surrounding environment in the bound and unbound states. Linear response weighted proportionality coefficients for explicit and implicit solvent electrostatic and van der Waals terms are typically calculated based on experimental data or similarity to previously solved systems (12). An inclusion of a surface area term has also be included in more recent calculations (13). ES/IS approaches typically use a combination of explicit solvent simulations to solve for conformational energy (and quasi-harmonic approaches to calculate the conformational entropy) and an estimate of the average solvation free energy (free energy of cavity formation + solute-solvent van der Waals energetics + polar interactions between the solute and solvent) (14, 15). While this approach is similar to MM-PBSA, key differences in the solvation free energy approach do exist. First, the energy of solute-solvent van der Waals is calculated during a molecular dynamics simulation (14, 15). Second, polar interactions between the solute and solvent do not rely on Poisson-Boltzmann or General Bourne approximations.

Instead, the polarization free energy is found by calculating the work done as the charges of the molecule of interest are “turned on”, typically with thermodynamic integration or free energy perturbation calculations (14).

MM-PBSA approaches were pursued here to calculate the free energy of binding of small, parallel coiled-coil dimers based on Hodges’ very stable IAAL-E3/K3 heterodimer (16). Coiled-coils are a well studied tertiary protein structural motif consisting of two or more α-helices wrapped around each other (17, 18). The coiled-coil structural motif is easily identified in the protein primary structure as a heptad repeat where individual positions are denoted **abcdefg**. Amino acid residues in positions **a** and **d** are typically hydrophobic in character and comprise the protein-protein binding domain and amino acid residues at **e** and **g** are typically polar and often form salt bridge interactions to stabilize the coiled-coil structure (19). More review of coiled-coils is discussed in our previous work (20) and elsewhere (21-24). Molecular dynamics (MD) trajectories for the dimer and individual monomers were used to estimate the free energy of binding as the difference between the free energy of the complex and the free monomers. Initial results from our models showed unrealistically favorable values for the free energy of binding even in models that were shown to spontaneously disassociate in simulation (IAAL-E3/O3, IAA(hFLeu)-E3/O3). Previous studies suggest that while the free energies are often quantitatively incorrect they can qualitatively predict the correct order or ranking of binding ligands (2, 4, 25, 26). Validation of this observation in our system was attempted by comparing our simulation results to an experimental study of coiled-coil stability by Jelesarov and Bosshart where free energies of binding were determined using isothermal calorimetry on various coiled-coil heterodimers (27). The MM-PBSA methodology was found to be quantitatively and qualitatively incorrect for this system (Table 5.1).

A review of the literature suggests that our system is not unique. Pearlman found that MM-PBSA poorly calculated and ranked the free energy of binding of a series of 16 ligands to p38 MAP kinase compared to thermodynamic integration (TI), one window free energy grid (OWFEG), and Dock Energy Score (11). Comparison of thermodynamic integration with MM-PBSA showed that MM-PBSA experienced problems with first solvation shell energetics when

Table 5.1 – MM-PBSA results for the free energy of binding (kcal mol^{-1}) for the small IAAL-E3/K3 derived coiled-coils and medium sized Jelesarov and Bosshart parallel coiled-coil dimers (AB, A12B, A12B12, and AB12). Free energies were calculated as the difference between the dimer absolute free energy and the free energies of individual monomers taken from separate trajectories. Overall, the free energies of binding were unrealistically negative (favorable) and failed to reproduce experimental values or correctly rank the dimers with respect to their free energies of binding.

Coiled-Coil Dimers	***Calc. $\Delta G_{binding}$ (kcal/mol)***	***Exp. $\Delta G_{binding}$ (kcal/mol)***
AB	-41.34	-10.69
A12B	-23.60	-9.72
AB12	-34.36	-10.51
A12B12	-33.54	-8.44
IAAL-E3/K3	-15.13	-9.60
IAA(hFLeu)-E3/K3	-25.37	NA
IAAL-O3/K3	-10.42	NA
IAA(hFLeu)-O3/K3	-4.19	NA
IAAL-R3/K3	-13.18	NA
IAA(hFLeu)-R3/K3	-25.24	NA
IAAL-H3/K3	-25.81	NA
IAA(hFLeu)-H3/K3	-20.44	NA

ranking the binding of RNA aptamer with theophylline and its analogs (28). Levy et al. found that the standard surface area model was not accurate enough for high resolution protein studies of protein folding and binding (29). Chen and Brooks found that the current surface area based nonpolar models have severe limitations, including insufficient description of the conformational dependence of solvation, over-estimation of the strength of pair-wise nonpolar interactions, and incorrect prediction of anti-cooperativity for three-body hydrophobic associations (30). They suggest that improvement can be made in the length-scale dependence of hydrophobic association and solvent screening of solute-solute dispersion interactions. Other corrections to the MM-PBSA approach have been proposed recently including hybrid linear response/MM-PBSA approaches (31), optimized radii for Poisson-Boltzmann calculations (32), the separation of van der Waals terms from the solvent accessible surface area approximation (10, 29), the inclusion of first solvent shell waters and ions in the MM-PBSA calculation (33, 34), atomic-based surface area terms (35, 36), and the separation of surface area dependent cavity and dispersion contributions in cyclic alkanes using a surface integral approach (37).

To gain further insights into the source of these errors and to consider corrections that would hopefully improve MM-PBSA methodology, MM-PBSA approaches were pursued to calculate the free energy of solvation of several small organic molecules with increasing complexity. We report here our results using MM-PBSA calculations on simple to more complex nonpolar and polar alkanes and suggest that an approach that includes the separation of van der Waals energy from the surface area approximation and includes a polar surface area term may be more appropriate (or more representative) for the calculation of free energies of solvation. Experimental measurements of free energies of solvation reported in this article are taken from references (38-47).

Methods

Molecule Parameterization

Initial approaches to MM-PBSA calculations for small organic molecules applied the RESP charges from Rizzo et al. (48) to parameters from the General AMBER force field (GAFF) (49). Later improvements to the assignment of RESP atomic charges used multiple orientations

of the molecules in an approach consistent with the original Cornell et al. (50) and the general AMBER force field (GAFF) (49, 51). The molecular structures were optimized using the Gaussian 03 software (52) at the HF/6-31+G* level (53). An SCF convergence criterion of 10-8 with tight optimization was used to ensure a fully minimized molecular structure. This minimized structure was then used to calculate a molecular electrostatic potential (MEP) on a three-dimensional grid using the GAMESS quantum mechanics software package (54) (again at the HF/6-31+G* level). Between 4 and 16 distinct orientations of small organic molecule structures were calculated and exported to the AMBER RESP program (55) which was used to fit atom centered RESP charges (55, 56) to the MEP. The parameterization was greatly facilitated by the RED II program which provides an automated method to create the MEP and fit the RESP charges (57).

Atom types for the novel molecules were chosen consistent with the general AMBER force field (GAFF). Torsional and angle parameters were assigned consistent with existing parameters from the GAFF force field. No new angle or dihedral force field parameters were required beyond those already available in GAFF.

Molecular Dynamics Simulations

All simulations of the organic molecules were completed using the general AMBER force field (GAFF) with RESP charges from the AMBER 9.0 modeling suite (58). Simulations of coiled-coil dimers used the ff99SB force field (59) from the same modeling suite.

Aqueous phase simulations. Organic molecules in this study were solvated by surrounding the compound with at least a 10 Å water layer in all directions within a truncated octahedron using explicit solvent. This amounts to between 793-1548 TIP3P (60) waters. Explicit Na+ and Cl- salt ions using the Aqvist parameters (61) were added to neutralize the system as needed. Energy minimization was performed for 500 steps first in the system with restrained substrate atoms (50 kcal mol^{-1} $Å^{-2}$) and then in an unrestrained system. Initial minimization was followed by heating to 298K at constant volume over a period of 10 ps using harmonic restraints of 2 kcal mol^{-1} $Å^{-2}$ on the solute atoms. Subsequent unrestrained equilibration at 298K followed for 500 ps. Bond lengths involving hydrogen atoms were constrained with SHAKE (62, 63) with a geometric tolerance for the constraint of 0.00001 Å during the coordinate

resetting. Periodic boundary conditions were applied using the particle mesh Ewald method (PME) with a less than 1 Å charge grid and cubic B-spline interpolation (64).

Production molecular dynamics (MD) simulations were performed with a 2 fs time step and a direct space non-bonded cutoff of 10 Å with the pair list of atomic interactions built out to 11 Å and heuristic update of the pair list triggered when any atom moved more than 0.5 Å since the previous update. During production runs, the center of mass translational motion of the entire system was removed after the initial velocity assignments and subsequently every 5000 MD steps. Constant temperature was maintained using weak temperature coupling to a heat bath with a 2 ps time constant (65). Following equilibration, pressure (1 atm) was maintained using isotropic position scaling with a 1.0 ps pressure relaxation time (65). 15 ns of production MD for each organic molecule of interest was recorded.

Gas phase simulations. Energy minimization of organic molecules was performed for 1000 steps using a 50 kcal mol^{-1} $Å^{-2}$ restraint in vacuum phase (igb = 0). Minimization was followed by restrained equilibration (50 kcal mol^{-1} $Å^{-2}$); heating the system to 298K in 10 ps. SHAKE was applied to bonds involving hydrogen atoms to ensure consistency with aqueous simulations. Production simulation was performed using a 2 fs time step for a total simulation time of 26 ns. A pair list of atomic interactions built out to 33 Å was applied with a heuristic update of the pair list triggered when any atom moved more than 0.5 Å since the last update. Temperature scaling used Langevin dynamics at a collision frequency of 2.0 ps^{-1}. For Langevin dynamics, the seed for the pseudo random seed generator used the current clock date and time to avoid synchronization artifacts (66).

MM-PBSA

Implementation of MM-PBSA for the determination of hydration free energies follows the methodology for use in AMBER as described for Gohlke and Case (10). Terminology used in this paper with respect to MM-PBSA will follow those definitions (10).

For solvation free energy estimation, 5000 snapshots were taken during the final 5 ns of production MD. Normal mode analysis calculations to determine entropy were completed using a subset of 50 snapshots spanning the range of the original set. All waters and ions were stripped.

Polar contribution to the solvation free energy was calculated using the Poisson-Boltzmann (PB) equations as implemented in Delphi II. For the calculations, two grid points per Å were used and the solute filled 80% of the grid box. Convergence in the Poisson-Boltzmann potential was determined when the absolute change in the PB potential at the grid points was less than 10^{-4} kT per unit charge (10). Atomic parse radii consistent with prior AMBER Delphi parameterization were used.

Molecular surface area (MSA) was calculated using the molsurf program implemented in AMBER. The surface tension proportionality constant was set to 0.00542 kcal mol^{-1} $Å^{-2}$ and the free energy of nonpolar solvation for a point solute *b* was set to 0.92 kcal mol^{-1}. This is consistent with the use of Delphi as the PB solver (10).

Entropy rotation, translation, and vibration contributions were calculated using normal mode analysis. Minimization of each snapshot in the gas phase used the conjugate gradient method with a distance dependant dielectric of 4r until the RMS of the elements of the gradient vector was less than 10^{-6} kcal mol^{-1} $Å^{-1}$. For the initial calculations, S_{config}, the configurational entropy contribution from side-chain reorganization, was neglected. Later calculations (see below) included an approximation of the configurational entropy of the solute.

Solute-Solvent van der Waals Contributions

For some calculations, van der Waals energetics between atoms in the solute and solvent were measured using the AMBER ANAL module to calculate the energetic interactions in all 5000 snapshots sampled. Calculation of the solute-solvent interactions used the dimensions of the periodic box as the cutoff of the van der Waal interactions.

Polar Surface Area

Polar surface area was calculated using the fast double cubic lattice method in the NSC approach (67) as implemented in the VEGA ZZ software (68). Surface area of solvent exposed atoms is represented as a dot surface where each dot is assigned a polar or apolar designation based on atom type. Polar surface area is then calculated as the sum of the surface area of all points designated as polar. Reported values of polar surface area were calculated as an average

of measured polar surface for an ensemble of 1000 snapshots spanning the last 5 ns of simulation time.

Water Shell Calculations

Calculations of the number of water molecules in the first and second solvation shells were performed using the ptraj watershell subroutine that is included in the AMBER Tools distribution. Default distance settings were maintained; the range of the first shell was set to a maximum of 3.5 Å from the molecule while the range of the second solvent shell was set to 5.0 Å. The water shell value recorded was a sum of all water molecules found in an ensemble of 5000 snapshots taken during the last 5 ns of production stimulation.

Water Mediated Hydrogen Bonds

To determine the average number of water mediated hydrogen bonds, the ptraj hbond subroutine was used to calculate solvent-solute hydrogen bonds on each polar atom in the solute molecule for each of the 5000 snapshots. Water mediated hydrogen bonds were counted if two or more solute polar groups were found to be bonded to the same water molecule at different polar atoms in the same time step. Initially, hydrogen bonds were counted if the distance was less than or equal to 3.5 Å with an angle less than or equal to 120 degrees. Later optimizations changed the hydrogen bond minimal angle and bond length to fit the difference seen between experimental and calculated solvation energies.

Configurational Entropy

Configurational entropy of the organic molecules in aqueous and gas phase was estimated by assigning all unique heavy-atom dihedral combinations in the molecule to three possible rotamers: gauche+, gauche-, or trans. Summation of the occupancy of the combined states across 5000 sampling snapshots provided the percentage occupancy (p_i) for each configuration state that was used to calculate the configuration entropy based on the Boltzmann Law (see equation 5.1).

$$S_{config} = p_i \log p_i \qquad [5.1]$$

Estimates of configuration entropy using angles in molecules that did not contain four unique atoms to form a dihedral plane were not attempted because no discrete population differences were seen between molecules in the aqueous and gas phases.

Thermodynamic Integration Free Energy Calculations

Calculations of the relative free energy of hydration were completed for some small molecules using thermodynamic integration (69). On the basis of a thermodynamic cycle, octane ($\lambda = 0$) was perturbed in both gas (vacuum) and aqueous environments to 2-octanone ($\lambda = 1$). Subsequent studies looked at the conversion of 2-octanone to 2,4-octanedione, 2,5-octanedione, 2,6-octanedione, and 2,7-octanedione. Relative free energy of hydration values were determined by subtracting gas phase perturbation measurements from aqueous phase measurements. Aqueous simulations were performed using a particle mesh Ewald treatment of electrostatics, as previously described. Equilibrated structures were allowed to relax during 6 ns of molecular dynamics simulation followed by thermodynamic integration sampling for a minimum of 6 ns at each sampling point. Convergence of thermodynamic integration sampling was visualized by a plateau in the $\partial V/\partial \lambda$ time course. For these calculations 20 sampling points of λ were used, based on Gaussian quadratures (0.00344, 0.01801, 0.04388, 0.08044, 0.12683, 0.18197, 0.24457, 0.31315, 0.38611, 0.49617, 0.50383, 0.61389, 0.68685, 0.75543, 0.81802, 0.87317, 0.91956, 0.95612, 0.98199, and 0.99656), sampled over 3 ns.

Results and Discussion

The MM-PBSA approach allows estimation of conformational free energy differences for many different types of molecules. Our attempts to use this approach to calculate the free energy of binding of coiled-coils was not feasible due to errors in both quantitative and qualitative measurements (see Table 5.1). As MM-PBSA has shown significant promise in free energy of binding measurements in other applications, a study of the source of error in this calculation was started by considering what made this system and calculation unique compared to other systems where MM-PBSA had been successfully utilized. In our study, we focused on the free energy of binding of parallel, coiled-coil dimers. These molecules tend to contain many charged and polar

residues which allow the formation of salt bridges. Secondarily, coiled-coils undergo significant structural reorganization upon binding (27). In our study, free coiled-coil monomers were completely unfolded prior to binding. Errors would most likely propagate when measuring water interactions between charged residues as well as large changes with respect to the surface area of the molecule. For this reason, the free energy of solvation term (ΔG_{solv}) in the MM-PBSA calculation was the focus of these corrections. In this study, we drew heavily upon the MM-PBSA Ras-Raf paper published by Golhke and Case (10) as well as the recent follow up by Rizzo and Case (48). To allow for clarity, terminology and initial parameters will be adopted from the former paper.

Our initial attempt was to use the published, default parameters in MM-PBSA to reproduce experimental free energies of solvation; i.e. the free energy difference as a molecule is transferred from gas phase into an aqueous environment.

To this end, models of small, linear nonpolar alkanes were developed using RESP atomic charges from Rizzo et al. (48) and parameters from the General AMBER Force Field (GAFF) for 12 nonpolar linear and branched alkanes which include models of methane, ethane, propane, butane, 2-methylpropane, 2,2-dimethylpropane, pentane, 2-methylpentane, 2,4-methylpentane, hexane, heptane, and octane. Our models differed from the original Rizzo approach in that parse radii were used in the Poisson Boltzmann (PB) calculation to determine the polar contribution to the free energy of solvation.

As mentioned by Rizzo and Case, prior approaches to calculating the free energy of solvation considered the approximation that the free energy of solvation could be estimated by summing polar and nonpolar contributions (45, 48, 70).

$$\Delta G_{\mathrm{solv}} = \Delta G_{\mathrm{solv,polar}} + \Delta G_{\mathrm{solv,nonpol}} \qquad [5.2]$$

The polar term or polar contribution to the free energy of solvation was determined using a grid based difference solution to the Poisson-Boltzmann (PB) equation. The nonpolar contributions to the free energy of solvation were calculated using an approximation to the total surface area of the molecule considered. This approximation was first suggested by Lee and

Richards who found a correlation between solvent accessible surface area and the experimental free energies of solvation (71). Later studies by Hermann built upon Lee and Richards' work by parameterizing the surface area approximation using small, straight alkanes to minimize polar contributions (72). This original fit of the nonpolar contribution to solvation free energy using experimental ΔG_{solv} values of small, straight alkanes plotted against their molecular surface area (MSA) set the surface tension proportionality constant, γ, to 0.00542 kcal mol^{-1} $Å^{-2}$ and the free energy of point solute b to 0.92 kcal mol^{-1} (72).

$$\Delta G^{SA}_{solv,nonpolar} = \gamma MSA + b \quad [5.3]$$

Free energies of solvation were calculated for the 12 nonpolar alkanes using equation 5.2 and Hermann's values for nonpolar energetic calculations. Values corresponded reasonably with Rizzo's results (R^2=0.98) and the experimental free energy of solvation for all compounds except ethane (see Figure 5.1a). The average absolute error from the experimental free energy of solvation was found to be 0.78 kcal mol^{-1} with a maximum absolute error (ethane) of 3.81 kcal mol^{-1}.

Phase Specific Enthalpy and Entropy Differences

Hermann's surface area approximation allowed the minimization of polar contributions to the free energy but neglected any differences arising from the solute's interaction with the environment. To determine if conformational differences between the gas phase and aqueous phase molecules were significant and contributed to a difference in the free energy of solvation, an alternative approach to calculating free energies of solvation was applied.

$$\Delta G^{*}_{solv} = G_{aq} - G_{gas} \quad [5.4]$$

$$G_{aq} = \langle U_{MM} \rangle_{aq} + G_{solv,PB_{(\varepsilon=80)}} + G_{solv,nonpol} - TS_{solute(aq)} \quad [5.5]$$

$$G_{gas} = \langle U_{MM} \rangle_{gas} + G_{solv,PB_{(\varepsilon=1)}} - TS_{solute(gas)} \quad [5.6]$$

$$\Delta G_{solv,polar} = G_{solv,PB_{(\varepsilon=80)}} - G_{solv,PB_{(\varepsilon=1)}} \quad [5.7]$$

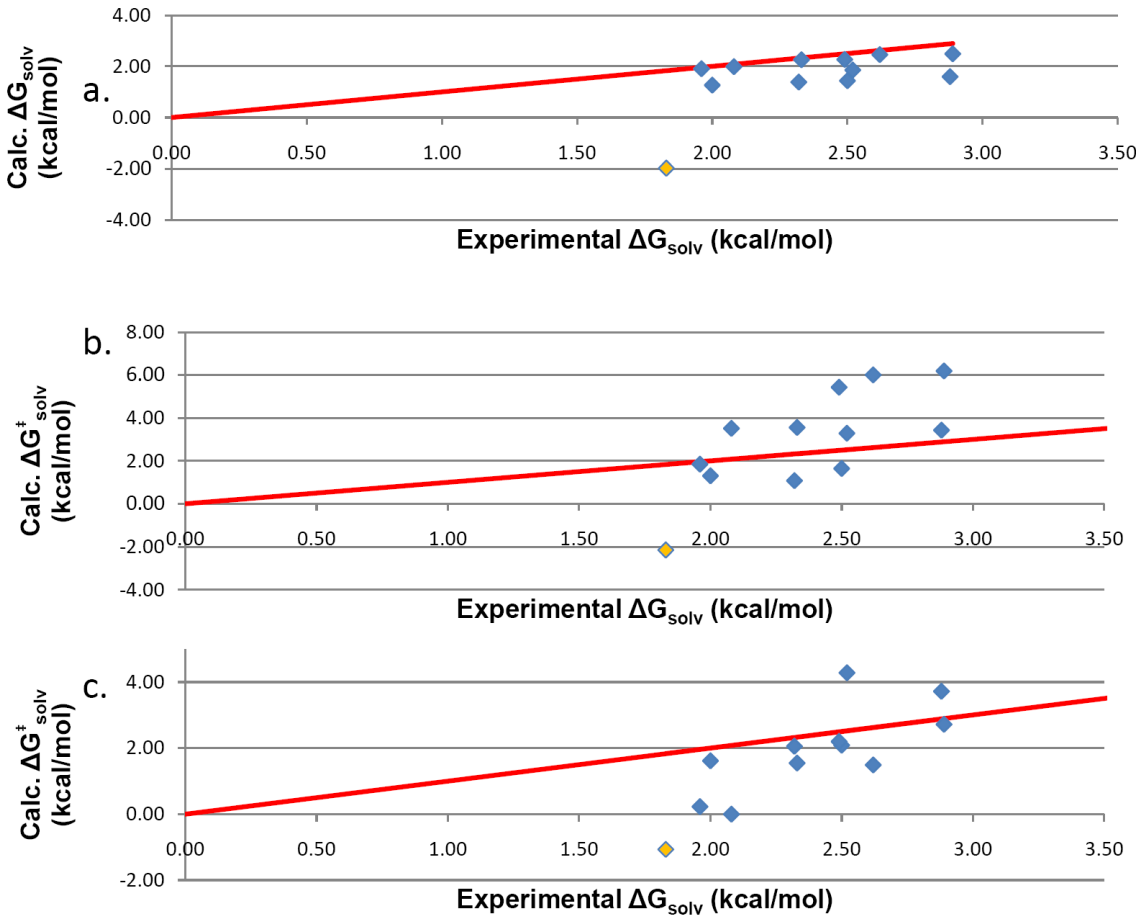


Figure 5.1 - Plot of experimentally measured and calculated solvation free energies for a small set of nonpolar alkanes. Solvation free energies were calculated using MM-PBSA at AMBER specific settings for Delphi as the Poisson Boltzmann solver(1). These settings specify that the surface tension proportionality constant γ be set to 0.00542 kcal mol^{-1} $Å^{-2}$ and the free energy of solvation for a point solute b be set to 0.92 kcal mol^{-1}. Initially, the calculated free energy of solvation (ΔG_{solv}) was found as a sum of polar and nonpolar contributions to the free energy of solvation (a). Later calculations ($\Delta G^{‡}_{solv}$) incorporated environmental conformational differences by including gas and aqueous phase solute-specific molecular mechanics potential energy and translational, rotational, and vibrational entropies (b and c). Figures 1a and 1b were generated using the Rizzo et al. atomic charges derived from a single orientation (a and b) while the atomic charges of molecules in Figure 5.1c were determined used between 4 and 16 orientations from a minimized QM derived structure. In all graphs, ethane (colored yellow) showed the greatest deviation from experimental measurements.

$$\Delta G^{\ddagger}_{solv} = \Delta G_{solv,polar} + \Delta G_{solv,nonpolar} + \left[\left(\langle U_{MM}\rangle_{aq} - \langle U_{MM}\rangle_{gas}\right) - T\left(S_{solute(aq)} - S_{solute(gas)}\right)\right] \quad [5.8]$$

For this approach, the free energy of solvation was calculating by subtracting the absolute free energy of the molecule in gas phase from the aqueous (aq) free energy using separate MD trajectories. This differs from MM-PBSA in its use of separate phase trajectories. The inclusion of solute entropy ($S_{tran,rot,vib}$) and molecular mechanics potential energy, $<U_{MM}>$, for the solute molecule (calculated in aqueous and gas phase MD trajectories) was introduced into this calculation in order to account for environment specific molecule rearrangement (73). The polar contribution to the free energy of solvation was calculated as the difference between the PB continuum solvent approximations taken at the dielectrics of 80 (aqueous) and 1 (gas phase). The solute potential energy can be further broken down into covalent terms (bonds, angles, and dihedral energetics) and non-covalent terms (van der Waals and electrostatics). In Figure 5.1b, experimental and calculated free energies of solvation are compared as calculated using equation 5.4 (average absolute error of 1.7 kcal mol^{-1}). It becomes clear from the data (Table 5.2) that significant entropic and conformational differences do exist and that these differences increase with the complexity of the molecule. The calculated values no longer trend with the experimental results. The magnitude of the deviation from the experimental measurements was decreased significantly (average absolute error of 1.1 kcal mol^{-1}) when atomic charges were refit by the authors using the RED II scripts but they still failed to correlate with experimental data (compare Figures 5.1b and 5.1c). It is evident that parameterization of the surface area term neglecting conformational dependence of the environment (i.e. solute conformational changes following solvation) can lead to significant errors in calculated free energies for large alkanes and organic molecules (see Figure 5.1c) (73, 74).

To determine if the surface area approximation could be refit by considering environmental differences to solute conformation and entropy, the molecular surface area was fit against the nonpolar contribution to the solvation free energy as defined below.

$$\Delta G^{SA,fit}_{solv,nonpol} = \Delta G^{exper.}_{solv} - \Delta G_{solv,polar} - \left[\left(\langle U_{MM}\rangle_{aq} - \langle U_{MM}\rangle_{gas}\right) - T\left(S_{solute(aq)} - S_{solute(gas)}\right)\right] \quad [5.9]$$

Table 5.2 - Experimentally measured and calculated solvation free energies ($\Delta G^{\ddagger}_{solv}$) for a series of nonpolar organic molecules (atomic charges from Rizzo et al.) using the approach defined in Equation 5.8. This entails the addition of solute-specific molecular mechanics potential energy and entropy terms in order to sample environmental conformational differences. As can be seen in the difference between the calculated and measured free energies, the absolute error tends to increase with the size and complexity of the organic molecules suggesting the original approach to calculating solvation free energies may be under-parameterized. All energetic terms are reported in kcal mol^{-1} units.

Compounds	*Gas Phase*		*Aqueous Phase*				*Free Energy of Solvation*		
	U_{MM}	$-TS_{tran,rot,vib}$	U_{MM}	$\Delta G_{solv,polar}$	$\Delta G_{solv,nonpol}$	$-TS_{tran,rot,vib}$	Calc. ΔG_{solv}.	Exper. ΔG_{solv}	Difference
methane	1.67	-14.72	1.71	-0.43	1.70	-14.72	1.31	2.00	0.69
ethane	3.92	-17.26	3.73	-3.89	1.91	-17.25	-2.16	1.83	3.99
propane	0.07	-19.71	-0.26	-0.17	2.08	-19.45	1.85	1.96	0.11
butane	11.91	-22.79	12.22	-0.25	2.24	-21.58	3.52	2.08	-1.44
2-methylpropane	-43.64	-21.47	-44.09	-0.82	2.21	-21.33	1.08	2.32	1.24
2,2-dimethylpropane	-71.33	-23.27	-71.48	-0.89	2.33	-22.92	1.64	2.50	0.86
pentane	8.45	-25.41	8.06	-0.15	2.41	-23.72	3.56	2.33	-1.23
2-methylpentane	-24.54	-26.88	-24.54	-0.66	2.52	-25.45	3.29	2.52	-0.77
2-4-dimethylpentane	-75.42	-29.25	-75.79	-1.03	2.63	-27.05	3.43	2.88	-0.55
hexane	12.29	-29.01	12.25	-0.30	2.57	-25.82	5.43	2.49	-2.94
heptane	18.98	-31.41	19.06	-0.28	2.74	-27.95	6.00	2.62	-3.38
octane	18.01	-33.67	18.10	-0.41	2.90	-30.06	6.19	2.89	-3.30

Plotting the molecular surface area against the nonpolar contribution to the free energy of solvation resulted in a very poor fit (R^2 correlation coefficient = 0.038) and a negative slope on a linear trend-line (see Figure 5.2a). Since our training set were very similar molecules and showed little variance in their respective experimental free energies of solvation, seven linear alkanes with a single polar group were added to the original, nonpolar molecule set. Despite this addition, the correlation remained weak (R^2 = 0.22) with a negative slope (see Figure 5.2b).

Direct Calculation of Solute-Solvent van der Waals

Golhke and Case cite studies that attempt to improve the surface area fit by separating out solute-solvent van der Waals contributions ($\Delta U_{solv,vdw}$) from the nonpolar contributions to the free energy of solvation ($\Delta G_{solv,nonpol}$) (10, 29, 75, 76).

$$\Delta G^{*}_{solv,nonpol} \approx \Delta U_{solv,vdw} + \Delta G_{solv,cav} \quad [5.10]$$

Van der Waals interactions can extend below the molecule surface and do not scale linearly. By directly calculating the van der Waals interactions and refitting the difference between our calculation and experimental results to the molecular surface area a strong correlation (R^2=0.94) was found that improved beyond the initial nonpolar fit (see Figure 2c). Average and maximum unsigned error were 0.84 kcal mol^{-1} and 1.66 kcal mol^{-1} respectively.

Alternative Surface Area Approximations

In the early research on surface area approximations, Lee and Richards stated that the reorganization component of the free energy of cavity formation ($\Delta G_{solv,cav}$) was directly proportional to the number of water molecules in the first solvation shell around the molecule and that the molecular surface area of the molecule is proportional to the number of water molecules in first solvation shell (71). If we consider two hypothetical molecules which have identical surface area but differ in regard to polarity at the molecule surface, one would imagine that charges near the surface of the molecule would disrupt the solvation shell and lead to solvent-solute interactions. Using the assumption that surface area was directly proportional to entropy loss due to ordering of the solvent, this assumption would fail to account for the difference in

a.

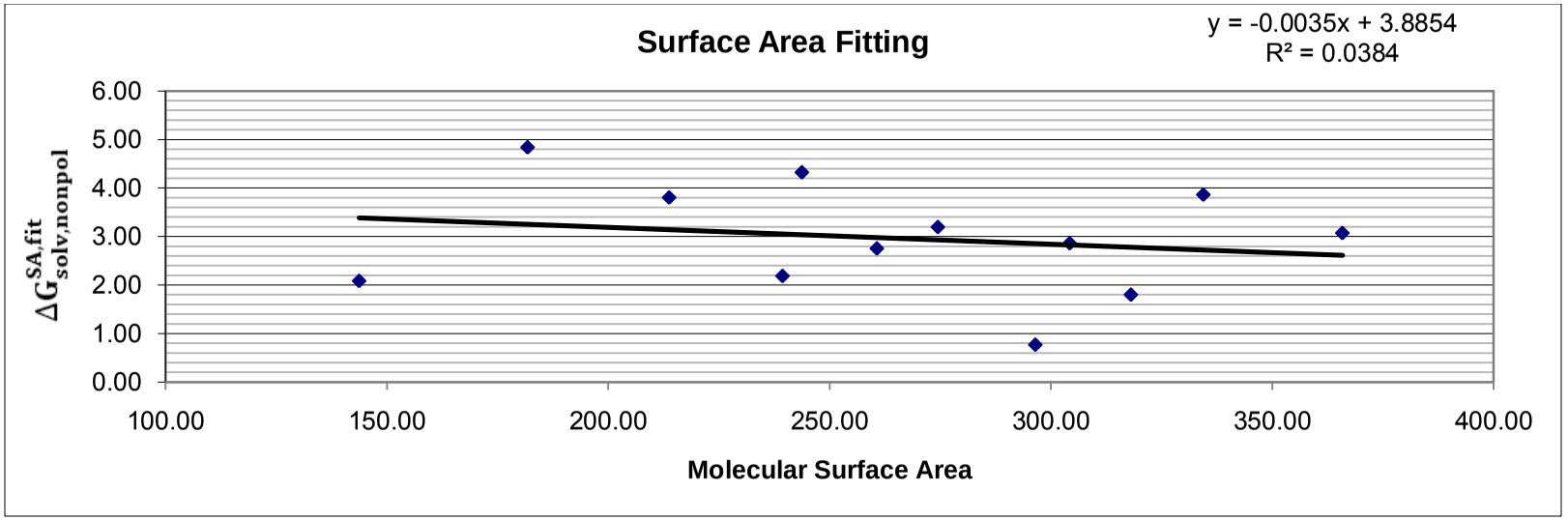


Figure 5.2 - Fitting of the nonpolar contribution of the free energy of solvation ($\Delta G^{SA,fit}_{solv,nonpol}$) to the molecular surface area of a test set of nonpolar alkanes. The nonpolar contribution of the free energy of solvation estimates non-Coulombic interactions between the solvent and solute. A poor correlation was found when using the original set of twelve nonpolar alkanes (a, R^2 = 0.038) as well as combined set which include the nonpolar alkanes and alkanes with a single polar group (b, R^2 = 0.23). Significant improvement is seen when the solute-solvent van der Waals interactions are calculated separately (c, R^2=0.94). When the solute-solvent van der Waals were calculated separately, the surface proportionality constant, γ, of the free energy of cavity formation ($G_{solv,cav}$) was fit at 0.0573 kcal mol^{-1} Å^{-2} and the intercept, b, at -2.3347 kcal mol^{-1}. Values of $G_{solv,cav}$ on the y axis were determined using $\Delta G_{solv,cav} = \Delta G^{*}_{solv,nonpol} - \Delta U_{solv,vdw}$. Calculation of the SA,PSA approximation set the surface proportionality constant, γ_1, at 0.0541 kcal mol^{-1} Å^{-2}, the polar surface proportionality constant, γ_2, at -0.0184 kcal mol^{-1} Å^{-2}, and the intercept, b, at -1.159 kcal mol^{-1} as determined by a fit to a combined set of nonpolar and polar molecules containing a single polar atom.

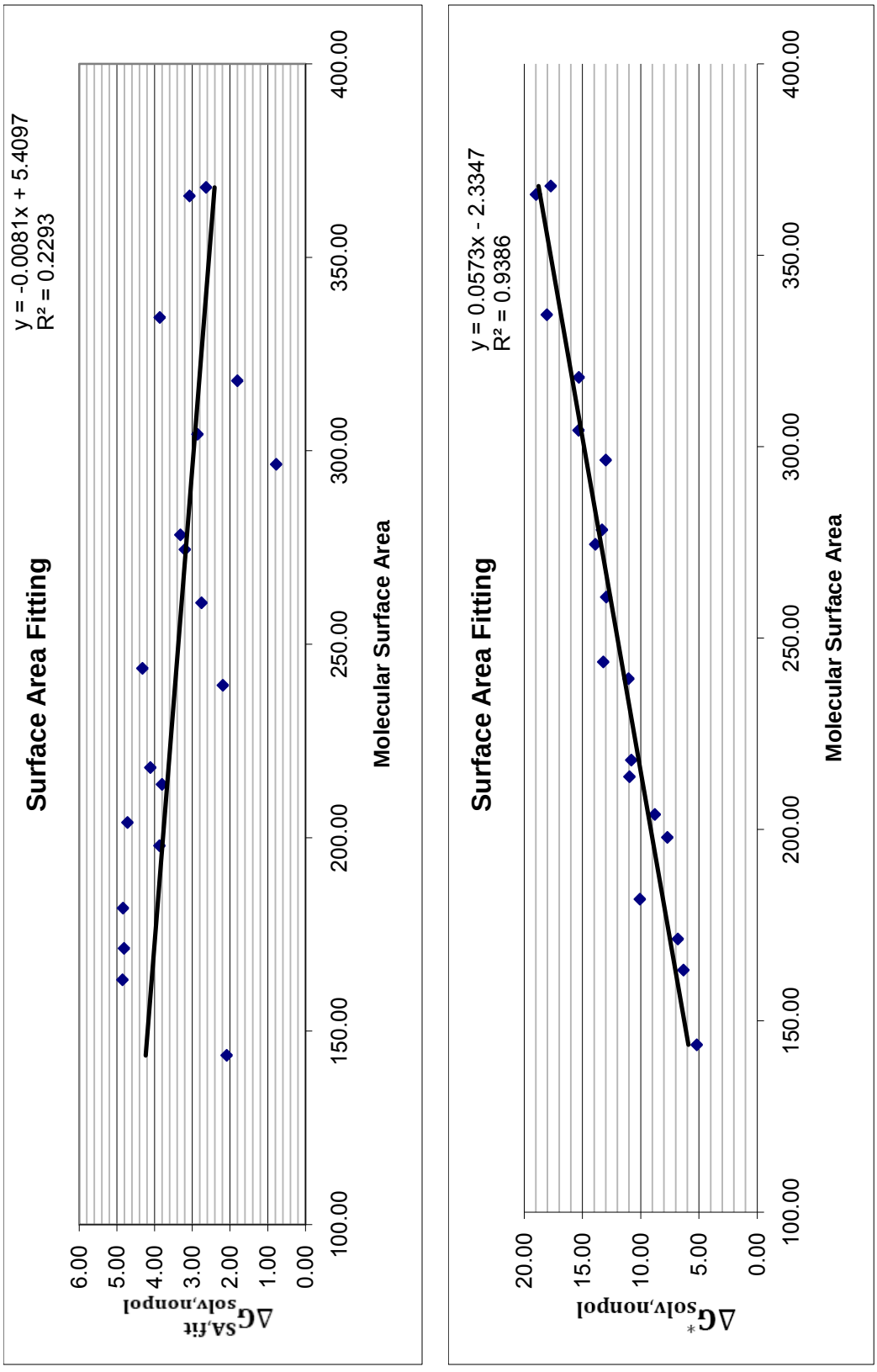


Figure 5.2 Continued

d.

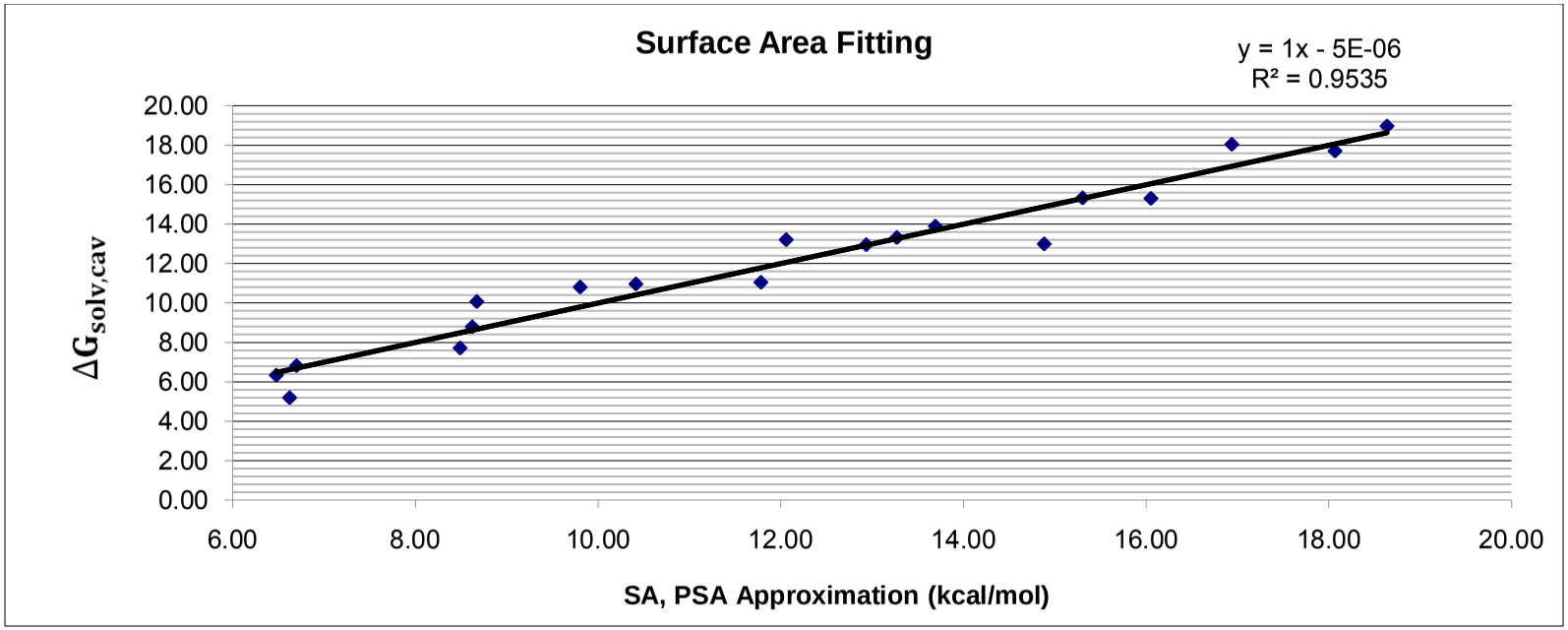


Figure 5.2 Continued

solvent-solute interactions in these two models. Several variables and combinations of variable were explored to give a better fit to the residual data than the simple, initial molecular surface area model (see Table 5.3). The two best fits used a combination of the molecular surface area and the polar surface area (MSA, PSA) and the number of water molecules in the first solvation shell and the polar surface area (Solvation Shell, PSA). In order to keep the calculations simple and focused on solute-centric approaches, the model that focused on molecular surface area and polar surface area was chosen. Comparison of experimental and calculated terms for the training set showed strong agreement (R^2 = 0.95) with a 0.64 kcal mol^{-1} average unsigned error from the experimental free energy of solvation (see Figure 5.2d and Table 5.4) with the maximum deviation found with 2-methylpentane (error = -1.99 kcal mol^{-1}). The best fit using the molecular surface and polar surface area approximation were found in the polar molecules.

Fourteen additional molecules composed of two or more polar groups were simulated and added to the training set. This addition of molecules led to large deviations from the experimental free energies of solvation (average unsigned error = 2.05 kcal mol^{-1}) and it was determined that the model was insufficiently parameterized to correctly calculate the free energy of solvation for these molecules (see Table 5.4).

Potential Sources of Error

The errors that appeared in our model may have developed from several sources. Experimental measurements of solvation energies in organic molecules containing two or more polar groups are rare due to the difficulty in obtaining completely dehydrated molecules(39). Molecules such as 1,2,3-propanetriol show large deviations in published measurements(77). These errors may result from the limits of the experimental approach. The interactions of two polar groups can also introduce interactions that we have not previously considered. Configurational entropy which is not typically measured in MM-PBSA calculations may play a role in the differences seen between calculations and measurements. Proximal charges will be highly affected by the dielectric constant of the environment. In gas phase, the dielectric constant is near 1.0 and like charges or elements of similar polarity would be repulsed while atoms with opposite charges and polarity would attract. Aqueous environments with a higher dielectric, 80,

Table 5.3 - Further improvements to fit of the cavity forming contribution of the free energy of solvation ($\Delta G_{solv,cav}$) by fitting to simple and compound variables. The sum of residues squared for the fit to each of these variable sets is shown including the molecular surface area (MSA), the number of water molecules in the first solvation shell, the number of waters in both the first and second solvation shell, the number of waters in the first solvation shell and the number of hydrophilic atoms, the molecular surface area and the number of water molecules in the first solvation shell, the number of waters in the solvation shell and the polar surface area (PSA), and the molecular surface area and the polar surface area. A lower residual sum of squares indicates a better fit between calculated and experimental values.

Variables	***Residual Sum of Squares***
MSA (Original approach)	11.79
Watershell1	11.61
MSA, Watershell1	11.08
Watershell1, Number of Hydrophillic Atoms	6.79
Watershell1, PSA	6.34
MSA, PSA	6.23

Table 5.4 - Experimentally measured and calculated solvation free energies (kcal mol^{-1}) for the combined set of nonpolar organic molecules, organic molecules with a single polar group, and organic molecules with two or more polar groups using a fit to molecular surface area and polar surface area ($\Delta G_{solv,\ SA\text{-}PSA}$).

Compounds	*Gas Phase*		*Aqueous Phase*				*Free Energy of Solvation*		
	U_{MM}	$-TS_{tran,rot,vib}$	$U_{MM} + \Delta G_{solv,polar}$	$-TS_{tran,rot,vib}$	$\Delta U_{solv,vdw}$	$\Delta G_{solv,SA\text{-}PSA}$	Calc. ΔG_{solv}	Exper. ΔG_{solv}	Difference
methane	1.69	-14.72	1.61	-14.72	-3.11	6.62	3.43	2.00	-1.43
ethane	3.93	-17.27	0.91	-17.25	-5.23	8.67	0.43	1.83	1.40
propane	5.90	-19.75	3.76	-19.45	-7.16	10.41	1.41	1.96	0.55
butane	9.95	-22.53	6.75	-21.58	-8.88	12.06	0.93	2.08	1.15
2-methylpropane	-10.75	-21.75	-11.03	-21.34	-8.86	11.78	3.06	2.32	-0.74
2,2-dimethylpropane	-6.83	-23.44	-7.61	-22.92	-10.20	12.94	2.48	2.50	0.02
pentane	12.42	-25.30	9.96	-23.71	-10.66283	13.69	2.16	2.33	0.17
2-methylpentane	8.11	-27.25	8.05	-25.45	-12.12	14.88	4.51	2.52	-1.99
2-4-dimethylpentane	-0.50	-28.89	-1.26	-27.05	-13.50	16.05	3.63	2.88	-0.75
hexane	15.46	-28.50	12.37	-25.79	-12.45	15.30	2.48	2.49	0.01
heptane	18.83	-31.31	14.22	-27.94	-14.20	16.94	1.50	2.62	1.12
octane	21.75	-34.22	17.42	-30.07	-15.90	18.64	2.56	2.89	0.33
methanol	5.85	-17.23	-4.31	-17.03	-1.48	6.48	-4.97	-5.11	-0.14
ethanol	-3.86	-19.84	-13.09	-19.49	-3.84	8.49	-4.23	-5.01	-0.78
methylamine	0.25	-17.07	-9.20	-17.00	-2.00	6.70	-4.68	-4.56	0.12
ethylamine	-16.01	-19.38	-25.36	-19.25	-4.07	8.62	-4.67	-4.50	0.17
propanal	8.88	-21.92	0.84	-21.43	-6.69	9.80	-4.44	-3.44	1.00
3-pentanone	19.20	-26.74	11.64	-25.91	-10.00	13.27	-3.45	-3.41	0.04
2-octanone	7.71	-34.85	-0.55	-32.10	-15.07	18.07	-2.52	-2.88	-0.36
acetamide	-88.65	-20.27	-103.74	-20.40	-4.56	7.93	-11.85	-9.71	2.14
propionamide	-37.54	-23.19	-51.91	-23.16	-6.55	9.79	-11.10	-9.41	1.68
2-methoxyethanol	24.23	-25.60	10.78	-23.60	-6.02	11.03	-6.44	-6.77	-0.33
2-methoxyethanamine	24.07	-24.43	9.48	-23.52	-6.09	11.17	-8.60	-6.55	2.05
1,1-diethoxyethane	-13.72	-34.24	-18.55	-30.31	-14.11	16.98	1.97	-3.27	-5.24
1,2-diethoxyethane	21.00	-34.21	12.92	-30.53	-14.30	17.62	-1.08	-3.53	-2.45
1,1-dimethoxyethane	-26.87	-27.12	-33.70	-25.54	-10.27	13.23	-2.29	-2.93	-0.64
1,2-dimethoxyethane	25.91	-28.58	16.71	-25.80	-10.35	13.70	-3.06	-4.84	-1.78
1,2,3-propanetriol	44.76	-28.80	22.35	-25.09	-2.85	9.90	-11.65	-9.22	2.43
2-ethoxyethanol	22.10	-28.81	9.34	-25.92	-8.04	12.98	-4.92	-6.61	-1.69
2-propoxyethanol	32.49	-31.83	19.17	-28.04	-9.38	14.60	-4.33	-6.42	-2.09
2-butoxyethanol	24.75	-33.92	14.51	-30.17	-12.28	16.27	-2.51	-6.27	-3.76
N-methylacetamide	-8.85	-24.51	-22.06	-23.28	-7.25	10.76	-8.47	-10.08	-1.61
N-methylformamide	1.44	-21.24	-11.38	-20.78	-5.63	8.77	-9.23	-10.00	-0.77

would buffer these electrostatic effects and this may lead to differences in the conformations available and the rigidity or fluidity of the molecules which adopt them (76, 78). Finally, the introduction of two or more polar group can lead to unique solute-solvent interactions such as water mediated hydrogen bonds. Water molecules in charge stabilized water mediated hydrogen bonds would lose translational and rotational freedom of movement and the errors seen may reflect an unmeasured loss of entropy in the solvent. To determine if any of these sources may have contributed to errors, we considered them separately.

Errors arising from failing to consider configurational entropy. Measurements of configurational entropy were considered by determining all unique dihedral angles in the molecules and then calculating and classifying them into three potential rotameric states: trans, gauche+, or gauche-. Data was collected during the terminal 5 ns of production simulation and the percent occupation of each configuration was determined. A configurational state was defined using all unique dihedrals in the molecule. Configurational entropy was calculated using the Boltzmann law as described earlier. These calculations show that the maximum energetic difference seen between gas and aqueous phase configurations amounted to 0.4 kcal mol^{-1} which cannot account for differences as large as 5.2 kcal mol^{-1}

Errors arising from experimental measurements. Experimental errors were considered by using thermodynamic integration (TI), instead of experimental values, to calculate solvation free energies for a series of alkanes with multiple polar groups and then comparing MM-PBSA calculations against the TI calculations. Thermodynamic integration has been shown to be as accurate as sub kcal mol^{-1} measurements in published studies (79). The relative solvation energy was calculated as octane was first converted to 2-octanone and then 2-octanone was converted to 2,4-octanedione, 2,5-octanedione, 2,6-octanedione, and 2,7-octanedione. For the conversion of octane to 2-octanone, the thermodynamic integration measurements were within 0.15 kcal mol^{-1} of the experimental published value. Comparison of the two calculation methodologies against each other (see Table 5.5) show that MM-PBSA produced similar errors against TI values as those seen in the experiment derived measurements. It can be concluded

Table 5.5 - Comparison of MM-PBSA calculated solvation free energies against thermodynamic integration (TI) derived solvation free energies (all values are reported in kcal mol^{-1} units) for organic molecules containing two or more functional groups. Thermodynamic integration has been shown to calculate free energies accurately to within sub kcal mol^{-1} range of experimentally derived measurements. Since similar errors are seen with MM-PBSA models against TI derived measurements and MM-PBSA models against experimentally derived solvation free energies (see Table 5.4, final section) of organic molecules with two or more polar groups, we can rule out experimental error as the main contribution

Compounds	*Gas Phase*		*Aqueous Phase*				*Free Energy of Solvation*		
	U_{MM}	$-TS_{tran,rot,vib}$	$U_{MM} + \Delta G_{solv,polar}$	$-TS_{tran,rot,vib}$	$\Delta U_{solv,vdw}$	$\Delta G_{solv,SA-PSA}$	Calc. ΔG_{solv}	TI ΔG_{solv}	Difference
2,4-octanedione	-17.14	-35.55	-30.32	-33.94	-15.04	17.59	-9.02	-6.23	2.79
2,5-octanedione	-1.36	-36.27	-14.59	-34.20	-14.61	17.53	-8.23	-6.74	1.50
2,6-octanedione	11.13	-37.30	-3.77	-34.23	-14.71	17.64	-8.90	-7.26	1.64
2,7-octanedione	-10.01	-36.93	-25.96	-34.23	-14.46	17.54	-10.17	-8.57	1.59

.

that experimental error is not the source of error seen in free energy of solvation calculations of alkanes with multiple polar groups.

Errors arising from complex solute-solvent interactions: The final consideration for the introduction of errors in the calculation of solvation free energies was loss of entropy due to solvent-solute interactions. We considered water-mediated hydrogen bonds to be a major contributor due to the translational and rotational constraints that must be maintained on a water molecule in a stable interaction. In order to limit the number of potential contributing variables, we limited the molecules sampled to alkanes containing carbonyl or ether groups. Water molecules participating in water mediated hydrogen bonds were counted during the terminal 5 ns of production simulation and were plotted against differences seen between the experimental (and TI calculated) and the MM-PBSA calculated results. Optimization of restrictions for bond angle and distance between hydrogen bond donors and acceptors did not allow for a correlation between water mediated hydrogen bonds and the errors seen in MM-PBSA calculations of alkanes with multiple polar groups. While it is evident that this approach to calculating loss of entropy in the solvent did not correlate with the errors observed, we are not prepared to state that loss of solvent entropy is not a major contributor to the errors seen. Indeed, the greatest deviation seen from our polar surface area model occurs with molecules where two electronegative groups face the same plane and are separated by two methylene groups, which is consistent with optimal spacing for interactions with water molecules. The measurement of solvent entropy loss by counting waters proximal to polar groups may not be sufficient quantitative to discount this variable. The authors' conclude that approaches that focus on solute measurements may have reached the limits of their practicality and accuracy in predicting complex solvent-solute interaction. While the inclusion of a polar surface term does improve accuracy it fails to consider how solvent entropy is affected by the charge of polar atoms, it considers the total surface area while failing to address the distribution and topology of the polar atoms, and the polar surface area term cannot describe and account for water mediated hydrogen bonds and local solvent entropy effects. Future directions to improve free energy

measurements may need to become solvent focused and explore explicit measurements of solvent translational and rotational movement.

When direct calculations of the solute-solvent van der Waals and the polar surface area approximation were applied to the original coiled-coil dimers to calculate the free energy of binding (see Table 5.6) some improvement was seen, but this approach still failed to correctly order the dimers with respect to binding energies. An additional source of error in this calculation (and calculations involving larger biomolecules) may result from inaccuracies of the force field applied. For example, errors in the side-chain torsion energetic barriers (which are typically not parameterized) may affect packing and the binding affinities of the individual monomers resulting in variance from experimental values. Similarly, if the force field used with coiled-coils fails to correctly represent the loss of secondary structure of the unbound monomer or insufficient simulation was performed to sample the unfolded state, calculated values for the free energy of binding would be incorrect.

Conclusion

In this chapter we report on the difficulties that we experienced when using the MM-PBSA energetic methodology to calculate the free energy of binding of small, parallel coiled-coils. While there has been significant success in the use of MM-PBSA, many researchers including ourselves have experienced difficulties in obtaining useful results. A few of those researchers have attributed the errors to the surface area term of the nonpolar contributions to the free energies of solvation. This possibility was explored by reproducing the free energies of solvation of several simple organic molecules using MM-PBSA and comparing those results to published, experimental-derived values.

It was shown that surface area fitting approaches which do not include molecular mechanics and entropy terms to account for environmental conformational differences may lead to errors when approximating the free energy of solvation.

Removing the van der Waals terms from the surface area approximation and independently calculating the solute-solvent van der Waals energetics greatly improved the

Table 5.6 – MM-PBSA results for the free energy of binding (kcal mol^{-1}) in representative coiled-coil dimers. Values in the initial column were calculated using the original approach (see Table 5.1) using a molecular surface area approximation to the nonpolar contribution to free energy of solvation. Approach 2 separates the surface area approximation of the nonpolar contribution to the solvation free energy ($\Delta G^{SA}_{solv,nonpol}$) to a direct calculation of the solute-solvent van der Waals energetics, $\langle U_{MM} \rangle$, and the free energy of cavity formation ($\Delta G_{solv,cav}$) (see Equation 5.10). Approach 3 builds upon the preceding approaches and considers a combined molecular surface area/polar surface area approximation ($\Delta G_{solv,\ SA\text{-}PSA}$) to the free energy of cavity formation ($\Delta G_{solv,cav}$) (see Table 5.3).

Coiled-Coil Dimers	*MM-PBSA original*	*MM-PBSA approach 2*	*MM-PBSA approach 3*	*Exper. $\Delta G_{binding}$*
AB	**-41.34**	-58.22	-43.69	-10.69
A12B	**-23.60**	-40.70	-25.75	-9.72
AB12	**-34.36**	-56.81	-35.94	-10.51
A12B12	**-33.54**	-56.18	-38.32	-8.44
IAAL-E3/K3	**-15.13**	0.25	-4.41	-9.60
IAA(hFLeu)-E3/K3	**-25.37**	-23.95	-17.40	NA
IAAL-O3/K3	**-10.42**	-5.57	3.99	NA
IAA(hFLeu)-O3/K3	**-4.19**	-7.96	-0.14	NA
IAAL-R3/K3	**-13.18**	-26.21	-19.04	NA
IAA(hFLeu)-R3/K3	**-25.24**	-36.04	-26.01	NA
IAAL-H3/K3	**-25.81**	-35.05	-27.89	NA
IAA(hFLeu)-H3/K3	**-20.44**	-31.67	-23.50	NA

correlation of the molecular surface area to the nonpolar contribution to the free energy of solvation.

We speculated that a significant source of the error which results from using a surface area approach is due to insufficient consideration of local interactions between water molecules and polar or ionic residues near the molecule surface. As a first attempt to model that behavior, a compound term to include both molecular surface area and polar surface area was introduced. With nonpolar alkanes and alkanes with a single polar residue, this additional term was sufficient to model free energies of solvation. Unfortunately, this term was under-parameterized when considering alkanes containing two or more polar groups.

In test cases of alkanes with two of more polar group, three potential sources of error for the MM-PBSA calculation were considered: errors arising due to experimental considerations, errors arising from failing to consider configurational entropy, and errors arising from complex solvent-solute interactions. Considering these potential causes independently we were able to eliminate the former two sources of error but felt that our model for testing complex solvent-solute interactions was not sufficiently robust to eliminate it as a variable.

Future studies to improve the MM-PBSA methodology will consider improved measurements of configurational entropy and methods to directly calculate translational and rotational entropies of water molecules as an alternative to surface area approaches to calculate the nonpolar contributions to the free energy of solvation.

References

1. Cheatham TE, 3rd, Srinivasan J, Case DA, Kollman PA. Molecular dynamics and continuum solvent studies of the stability of polyG-polyC and polyA-polyT DNA duplexes in solution. Journal of Biomolecular Structure and Dynamics 1998; 16: 265-280.

2. Kuhn B, Kollman P. Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. Journal of Medicinal Chemistry 2000; 43: 3786-3791.

3. Donini OA, Kollman PA. Calculation and prediction of binding free energies for the matrix metalloproteinases. Journal of Medicinal Chemistry 2000; 43: 4180-4188.

4. Wang J, Morin P, Wang W, Kollman PA. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. Journal of the American Chemical Society 2001; 123: 5221-5230.

5. Huo S, Wang J, Cieplak P, Kollman PA, Kuntz ID. Molecular dynamics and free energy analyses of Cathepsin D-Inhibitor interactions: insight into structure-based ligand design. Journal of Medicinal Chemistry 2002; 45: 1412-1419.

6. Kuhn B, Donini O, Huo S, Wang J, Kollman PA. MMPBSA applied to computer-assisted ligand design. *Free Energy Calculations in Rational Drug Design*. New York: Kluwer Academic/Plenum Publishers; 2001. p. 243-251.

7. Kuhn B, Gerber P, Schulz-Gasch T, Stahl M. Validation and use of the MM-PBSA approach for drug discovery. Journal of Medicinal Chemistry 2005; 48: 4040-4048.

8. Gohlke H, Kiel C, Case D. Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the Ras–Raf and Ras–RalGDS complexes. Journal of Molecular Biology 2003; 330: 891-913.

9. Kollman P, Massova I, Reyes C, et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Accounts of Chemical Research 2000; 33: 889-898.

10. Gohlke H, Case DA. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. Journal of Computational Chemistry 2004; 25: 238-250.

11. Pearlman DA. Evaluating the molecular mechanics poisson-boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP kinase. Journal of Medicinal Chemistry 2005; 48: 7796-7807.

12. Hansson T, Marelius J, Åqvist J. Ligand binding affinity prediction by linear interaction energy methods. Journal of Computer-Aided Molecular Design 1998; 12: 27-35.

13. Su Y, Gallicchio E, Das K, Arnold E, Levy R. Linear interaction energy (LIE) models for ligand binding in implicit solvent: theory and application to the binding of NNRTIs to HIV-1 reverse transcriptase. Journal of Chemical Theory and Computation 2007; 3: 256-277.

14. Vorobjev Y, Almagro J, Hermans J. Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. Proteins: Structure Function and Genetics 1998; 32: 399-413.

15. Vorobjev Y, Hermans J. ES/IS: estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model. Biophysical Chemistry 1999; 78: 195-205.

16. Lindhout DA, Litowski JR, Mercier P, Hodges RS, Sykes BD. NMR solution structure of a highly stable de novo heterodimeric coiled-coil. Biopolymers 2004; 75: 367-375.

17. Adamson JG, Zhou NE, Hodges RS. Structure, function and application of the coiled-coil protein folding motif. Current Opinion in Biotechnology 1993; 4: 428-437.

18. Lupas A. Coiled coils: new structures and new functions. Trends in Biochemical Sciences 1996; 21: 375-382.

19. Lupas AN, Gruber M. The structure of alpha-helical coiled coils. Advances in Protein Chemistry 2005; 70: 37-78.

20. Pendley S, Yu Y, Cheatham III T. Molecular dynamics guided study of salt bridge length dependence in both fluorinated and non-fluorinated parallel dimeric coiled-coils. Proteins: Structure, Function, and Bioinformatics 2009; 74: 612-629.

21. Cohen C, Parry D. Alpha-helical coiled coils--a widespread motif in proteins. Trends in Biochemical Sciences 1986; 11: 245-248.

22. Woolfson D. The design of coiled-coil structures and assemblies. Advances in Protein Chemistry 2005; 70: 79-112.

23. Zhou N, Zhu B, Kay C, Hodges R. The two-stranded -helical coiled-coil is an ideal model for studying protein stability and subunit interactions. Biopolymers 2004; 32: 419-426.

24. Gillingham A, Munro S. Long coiled-coil proteins and membrane traffic. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research 2003; 1641: 71-85.

25. Combelles C, Gracy J, Heitz A, Craik D, Chiche L. Structure and folding of disulfide-rich miniproteins: Insights from molecular dynamics simulations and MM-PBSA free energy calculations. Proteins: Structure, Function, and Bioinformatics 2008; 73: 87-103.

26. Massova I, Kollman P. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. Perspectives in Drug Discovery and Design 2000; 18: 113-135.

27. Jelesarov I, Bosshard H. Thermodynamic characterization of the coupled folding and association of heterodimeric coiled coils (leucine zippers). Journal of Molecular Biology 1996; 263: 344-358.

28. Gouda H, Kuntz I, Case D, Kollman P. Free energy calculations for theophylline binding to an RNA aptamer: comparison of MM-PBSA and thermodynamic integration methods. Biopolymers 2003; 68: 16-34.

29. Levy RM, Zhang LY, Gallicchio E, Felts AK. On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute-solvent interaction energy. Journal of the American Chemical Society 2003; 125: 9523-9530.

30. Chen J, III C. Implicit modeling of nonpolar solvation for simulating protein folding and conformational transitions. Physical Chemistry Chemical Physics 2008; 10: 471-481.

31. Zhou Z, Bates M, Madura J. Structure modeling, ligand binding, and binding affinity calculation (LR-MM-PBSA) of human heparanase for inhibition and drug design. Proteins: Structure, Function, and Bioinformatics 2006; 65: 580-592.

32. Swanson J, Adcock S, McCammon J. Optimized radii for Poisson- Boltzmann calculations with the AMBER force field. Journal of Chemical Theory and Computation 2005; 1: 484-493.

33. Cheatham III T, Ryjac?ek F, Lankas F, van Meervelt L, Hobza P, Sponer J. Molecular dynamics simulations and thermodynamics analysis of DNA- drug complexes. Minor groove binding between 4, 6-Diamidino-2-phenylindole and DNA duplexes in solution. Journal of the American Chemical Society 2003; 125: 1759-1769.

34. Štefl R, Cheatham T, Špa ková N, et al. Formation pathways of a guanine-quadruplex DNA revealed by molecular dynamics and thermodynamic analysis of the substates. Biophysical Journal 2003; 85: 1787-1804.

35. Eisenberg D, McLachlan A. Solvation energy in protein folding and binding. Nature 1986; 319: 199-203.

36. Ooi T, Oobatake M, Nemethy G, Scheraga H. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proceedings of the National Academy of Sciences 1987; 84: 3086-3090.

37. Zacharias M. Continuum solvent modeling of nonpolar solvation: Improvement by separating surface area dependent cavity and dispersion contributions. Journal of Physical Chemistry A 2003; 107: 3000-3004.

38. Baumeister T, Cordes F. A new model for the free energy of solvation and its application in protein ligand scoring: Technical Report 04-51, Zuse Institute Berlin (ZIB), 2004.

39. Nicholls A, Mobley D, Guthrie J, et al. Predicting small-molecule solvation free energies: an informal blind test for computational chemistry. Journal of Medicinal Chemistry 2008; 51: 769-779.

40. Li J, Zhu T, Hawkins G, et al. Extension of the platform of applicability of the SM5. 42R universal solvation model. Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta) 1999; 103: 9-63.

41. Hine J, Mookerjee P. The intrinsic hydrophilic character of organic compounds. Correlations in terms of structural contributions. Journal of Organic Chemistry 1975; 40: 292-298.

42. Hawkins G, Cramer C, Truhlar D. Parametrized model for aqueous free energies of solvation using geometry-dependent atomic surface tensions with implicit electrostatics. Journal of Physical Chemistry B 1997; 101: 7147-7157.

43. Cabani S, Gianni P, Mollica V, Lepori L. Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution. Journal of Solution Chemistry 1981; 10: 563-595.

44. Duffy E, Jorgensen W. Prediction of properties from simulations: free energies of solvation in hexadecane, octanol, and water. Journal of the American Chemical Society 2000; 122: 2878-2888.

45. Sitkoff D, Sharp K, Honig B. Accurate calculation of hydration free energies using macroscopic solvent models. The Journal of Physical Chemistry 1994; 98: 1978-1988.

46. Viswanadhan V, Ghose A, Singh U, Wendoloski J. Prediction of solvation free energies of small organic molecules: additive-constitutive models based on molecular fingerprints and atomic constants. Journal of Chemical Information and Computer Science 1999; 39: 405-412.

47. Ben Naim A, Marcus Y. Solvation thermodynamics of nonionic solutes. The Journal of Chemical Physics 1984; 81: 2016-2027.

48. Rizzo R, Aynechi T, Case D, Kuntz I. Estimation of absolute free energies of hydration using continuum methods: accuracy of partial charge models and optimization of nonpolar contributions. Journal of Chemical Theory Computation 2006; 2: 128-139.

49. Pearlman D, Case D, Caldwell J, et al. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Computer Physics Communications 1995; 91: 1-41.

50. Cornell WD, Cieplak P, Baylyl CI, et al. A second generation force field for the simulation of proteins, nucleic acids, and organic Molecules. Journal of the American Chemical Society 1995; 117: 5179-5197.

51. Pigache A, Cieplak P, Dupradeau F. Automatic and highly reproducible RESP and ESP charge derivation: application to the development of programs RED and X RED. 227th ACS National Meeting; 2004; 2004.

52. Frisch MJ, Trucks GW, Schlegel HB, et al. Gaussian 2003. Gaussian, Inc.: Pittsburgh, PA; 2003.

53. Hehre W, Ditchfield R, Pople J. Self—consistent molecular orbital methods. XII. further extensions of gaussian—type basis sets for use in molecular orbital studies of organic molecules. The Journal of Chemical Physics 1972; 56: 2257-2261.

54. Granovsky AA. PC GAMESS version 7.0. 2006.

55. Bayly CI, Cieplak P, Cornell WD, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges- the RESP model. Journal of Physical Chemistry 1993; 97: 10269-10280.

56. Cieplak P, Cornell W, Bayly C, Kollman P. Application of the multimolecule and multiconformational resp methodology to biopolymers: charge derivation for DNA, RNA, and proteins. Journal of Computational Chemistry 1995; 16: 1357-1377.

57. Zaffran TC, P.;, Dupradeau FY. R.E.D. II. 2005.

58. Case D, Darden T, Cheatham III T, et al. AMBER 9. University of California, San Francisco; 2006.

59. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 2006; 65: 712-725.

60. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparisons of simple potential functions for simulating liquid water. Journal of Chemical Physics 1983; 79: 926-935.

61. Aqvist J. Ion-water interaction potentials derived from free energy perturbation simulations. Journal of Physical Chemistry 1990; 94: 8021-8024.

62. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. Journal of Computational Physics 1977; 23: 327-341.

63. Barth E, Kuczera K, Leimkuhler B, Skeel RD. Algorithms for constrained molecular dynamics. Journal of Computational Chemistry 1995; 16: 1192-1209.

64. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh ewald method. Journal of Chemical Physics 1995; 103: 8577-8593.

65. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. Journal of Computational Physics 1984; 81: 3684-3690.

66. Cerutti D, Duke R, Freddolino P, Fan H, Lybrand T. Vulnerability in popular molecular dynamics packages concerning Langevin and Andersen dynamics. Journal of Chemical Theory and Computation 2008; 4: 1669-1680.

67. Eisenhaber F, Lijnzaad P, Argos P, Sander C, Scharf M. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. Journal of Computational Chemistry 1995; 16: 273-284.

68. Pedretti A, Villa L, Vistoli G. VEGA--an open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming. Journal of Computer-Aided Molecular Design 2004; 18: 167-173.

69. Mezei M. The finite difference thermodynamic integration, tested on calculating the hydration free energy difference between acetone and dimethylamine in water. Journal of Chemical Physics 1987; 86: 7084-7088.

70. Still W, Tempczyk A, Hawley R, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. Journal of the American Chemical Society 1990; 112: 6127-6129.

71. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. Journal of Molecular Biology 1971; 55: 379-400.

72. Hermann R. Theory of hydrophobic bonding. 11. The correlation of hydrocarbon solubility in water with solvent cavity surface area. Journal of Physical Chemistry 1972; 76: 2754-2759.

73. Yang C, Sun H, Chen J, Nikolovska-Coleska Z, Wang S. Importance of ligand reorganization free energy in protein-ligand binding-affinity prediction. Journal of the American Chemical Society 2009; 131: 13709-13721.

74. Mobley D, Dill K, Chodera J. Treating entropy and conformational changes in implicit solvent simulations of small molecules. Journal of Physical Chemistry B 2008; 112: 938-946.

75. Ashbaugh H, Kaler E, Paulaitis M. A “universal” surface area correlation for molecular hydrophobic phenomena. Journal of the American Chemical Society 1999; 121: 9243-9244.

76. Gallicchio E, Kubo M, Levy R. Enthalpy-entropy and cavity decomposition of alkane hydration free energies: Numerical results and implications for theories of hydrophobic solvation. Journal of Physical Chemistry B 2000; 104: 6271-6285.

77. Sandberg L, Casemyr R, Edholm O. Calculated hydration free energies of small organic molecules using a nonlinear dielectric continuum model. Journal of Physical Chemistry B 2002; 106: 7889-7897.

78. Ashbaugh HS, Kaler EW, Paulaitis ME. Hydration and conformational equilibria of simple hydrophobic and amphiphilic solutes. Biophysical Journal 1998; 75: 755-768.

79. Shirts M, Pitera J, Swope W, Pande V. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. The Journal of Chemical Physics 2003; 119: 5740-5761.

CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS

This series of studies set out to determine whether modern MD simulation approaches would be accurate enough to assist in the design of engineered coiled-coils for therapeutic applications. Initially we looked at the fluorinated coiled-coils from the work of Y. Bruce Yu (currently at the University of Maryland) which lead to a formulation of a series of coiled-coil models. These models were based on Hodges' IAAL-E3/K3 coiled-coil (1, 2) in order to test relative free energies of binding and began simulations using the ff99 force field with both explicit and implicit solvent models. While our early experiments with larger coiled-coils looked promising, we quickly ran into two very serious barriers to the continuation of our work using these small coiled-coils. First, it became obvious early on that the ff99 force field, which was the most current force field at that time, failed to reproduce the structure of our model protein, IAAL-E3/K3. Second, our approach to calculate the relative free energies of binding using the MM-PBSA methodology was reporting values that were overly negative (stabilizing). Through contact with the Simmerling lab we were able to test early versions of their ff99-mod1 and ff99SB force fields and found that they greatly improved the stability of the IAAL-E3/K3 dimer. Our work applying these and other AMBER force fields is considered in Chapter 4 of this dissertation. We found that none of the force fields tested were able to reproduce the structure of IAAL-E3/K3 stably during long trajectories. The Simmerling lab is currently developing a force field that addresses the side-chain torsional bias that we discovered and future MD simulations applying that force field look promising. In this dissertation, we used both the ff99SB and ff03 force field with shorter trajectories and carefully monitored the stability over time to minimize the introduction of simulation artifacts into these studies.

The second barrier involved errors in energy calculation methodologies. The free energy of binding approach used in MM-PBSA subtracts the free energy of the dimer from the free energy of the individual monomers. In coiled-coils, much of secondary structure forms (folds) only upon association or dimerization. Since the nonpolar contribution to the free energy of solvation is directly proportional to the surface area of the molecule, large changes to the surface area of the molecule would propagate any errors that exist in that term. This is precisely what occurs in coiled-coils where the surface area can change significantly between the bound (folded) and unbound (unfolded) species. Several approaches were used to try to get around these errors including: fixing the MM-PBSA methodology, using thermodynamic integration to calculate the free energy of binding, using helicity calculations to estimate relative free energies of binding, and the calculation of a potential of mean force (PMF). In Chapter 5, we discussed some approaches used to try to correct for these errors in the MM-PBSA calculation which included: the direct calculation of van der Waals energetics between the protein and the solvent, calculations of configurational entropy, and the use of a polar surface area term to improve solvent entropy calculations. While our approach did improve calculations of the free energy of hydration, they did not solve the errors nor allow correct ranking when applied towards the free energy of binding. We concluded that future approaches may need to consider direct calculations of the solvent entropy based on the movement of individual solvent molecules. Some recent papers have suggested the use of atom based surface area terms (3-5) or hydropathic interaction analysis (HINT) (6-8) as a replacement for the molecular surface area approach that is typically used. Atom based surface approaches are similar to our polar surface approach in their ability to designate regions as hydrophobic or hydrophilic and may further improve our approach by considering atom type specific interactions with the solvent. Hydropathic interactions analysis is an increasingly popular approach which calculates solvent and solute interactions based on solvent exposure of chemically active groups (such as thiol and hydroxyl groups). One concern in using either of these approaches as a replacement to the current surface area approximation is that neither of them considers complex solvent structure and entropy loss due to the proximity of polar atoms across the solute surface. Another concern is our reliance on experimental

measurements of the free energy of hydration. Experimental measurements show greater and greater variance with an increasing number of polar groups in the molecule. For future improvements, a transition from experimental measurements to thermodynamic integration calculations of the free energy of hydration may be necessary to eliminate error from the calculation and fitting methodologies. A final direction that can be explored in future studies is the water model used in simulations. For the experiments reported in this dissertation, we used TIP3P waters as our water model for MM-PBSA calculations. TIP3P is a very basic model and while it has been used successfully in many simulations, a transition to a more exact model, such as TIP4P$_{EW}$ or SPC/E, in future studies may remove errors that arise from inaccuracies in the calculation of solvation entropies.

In addition to our attempts to improve or fix MM-PBSA we considered the use of thermodynamic integration (TI) to calculate the relative free energy of binding for coiled-coils using a thermodynamic cycle. Mixed results were seen using this approach in the two studies of coiled-coils we pursued. In Chapter 3, we showed thermodynamic integration successfully used to calculate the free energy of binding of anti-parallel coiled-coils composing the binding domain of the Bcr protein. These coiled-coils showed significant helical character in the monomeric form which likely lead to our success in making those calculations. Thermodynamic integration was also attempted using parallel coiled-coils based on the IAAL-E3/K3 design as described in Chapter 2. These coiled-coils lost all secondary structure in the monomeric form and TI calculations deviated orders of magnitude from the experimental measurements. The failure in the application of thermodynamic integration approaches in this latter study could be attributed to two possible conditions. First, insufficient equilibration time between individual sampling points could lead to incorrect TI energetic values. Coiled-coils show significant conformational change between bound and unbound states and the simulation time required to adequately sample the folding transition (at each individual sampling step) may not have been sufficient. Second, the accuracy of the TI calculation is directly reliant on the accuracy of the force field to model the chemical transition. As was mentioned earlier, improvements to the protein force field are needed and this may have affected the accuracy of the calculation.

Although not well described in prior chapters an additional energetic approach using umbrella sampling and a potential of mean force (PMF) calculation of the binding energy for the Bcr coiled-coils was attempted. This approach focused on calculating the resistance to the forced separation of coiled-coil monomers from a starting bound trajectory as an approximation to the binding energy. Five distance restraints and one dihedral angle restraint were used to force the separation and define the separation vector to prevent nonspecific interactions. Despite the restraints, the coiled-coil monomers denatured in order to maintain surface contacts. Energetic calculations were unusually high reflecting the additional cost of protein folding as well as separation. Ultimately this approach was abandoned in favor of thermodynamic integration calculations.

Experimentally, the use of helicity or ellicipity measurements are often used for calculating the free energy of binding in coiled-coils (9, 10). These measurements typically use circular dichroism calculations directly or in conjunction with a thermal or chemical denaturant (9, 11, 12). In our studies using the Bcr-Abl and fluorinated coiled-coils, we used the DichroCalc program developed by Bulheller and Hirst (13) to calculated circular dichroism (CD). Secondary approaches to measure association include percent helicity approximations by determining secondary structure using the DSSP algorithm (14) and calculation of α-helix specific hydrogen bonds. The success of this approach to approximate free energy of binding assumes equivalent helicity in the coiled-coils in monomeric form. Future directions may consider algorithms to subtract CD calculations of the unbound monomers from the CD calculation of the bound dimer as an approach to further improve this metric.

Our study of fluorinated compounds (Chapter 2) was designed to address several questions. The use of fluorinated amino acids in coiled-coils were considered to be isomorphic to their non-fluorinated complements despite the large size difference between hydrogen and fluorine atoms. We were interested in determining if the increased size of the fluorinated amino acids affected the ideal salt bridge length and the packing of the hydrophobic core. Fluorinated amino acids were also reported to improve the resistance of a protein to thermal and chemical denaturation and we explored causes for that stabilization. Using thermodynamic integration, we

found that an increase in stability and ellipticity of coiled-coil was due to differences in hydrophobicity of the fluorinated and nonfluorinated amino acids. 5,5,5,5',5',5'-hexafluoroleucine was 1.1 kcal/mol less stable in aqueous solvent than leucine. This increase in hydrophobicity would drive association (or prevent denaturation). Improved association would lead to additional formation of secondary structure or improved ellipticity. We also found that while fluorinated amino acids are not isomorphic, changes in their preferred rotamers actually lead to similar geometries in the hydrophobic core. These differences in rotamer preferrence do decrease the configurational entropy (flexibility) of the core and result in a small entropic penalty to the free energy of the coiled-coil. We found that fluorinated amino acids can contribute to an electrostatic stabilization of salt bridge interactions as well. Energetic decomposition of the IAA(hFLeu)-E3/R3 coiled-coil dimer showed an stabilization of -1.8 kcal/mol due to an interaction between Arg29, Glu13, and an adjacent hexafluoroleucine over its nonfluorinated complementary dimer, IAAL-E3/R3. Finally, our force field model of the IAAL-E3/K3 dimer was able to show us errors in the leucine rotamer assignments that were made in the NMR structure. Future studies of this system could further explore the electrostatic stabilization that we saw in IAA(hFLeu)-E3/R3 coiled-coil dimer and determine if we could reproduce that stabilization on both sides of the salt bridges possibly through the use of a fluorinated isoleucine analog. Recently the use of intrahelical hydrogen bonds has been introduced in the design of parallel, coiled-coil. Computational approaches could greatly aid the energetic analysis of this design to determine the stabilization and energetic results of incorporating this feature in the protein motif.

The Bcr-Abl collaboration (see Chapter 3) with the Lim group was a natural continuation of our work with coiled-coils and an increase in the complexity of our studies. The Bcr-Abl protein contains an anti-parallel coiled-coil region as opposed to the IAAL-E3/K3 parallel coiled-coil derivatives we had designed. Antiparallel coiled-coils are less well-known and consequently not often used in engineering. These coiled-coils often share mixtures of stabilizing and destabilizing elements in order to favor binding while destabilizing the parallel orientation of that binding. This collaboration focused on designing a mutant that would improve binding to the wild type Bcr-Abl while ensuring specificity against self-association. We considered several different

engineering approaches including designing disulfide bonds, increasing alanine in the backbone, and rationally designed point mutations. Stability and the free energy of binding were estimated using helicity calculations and later more refined energetic approaches including MM-PBSA and thermodynamic integration calculations. Computational analysis showed improved binding of the designed mutant coiled-coils to the wild type protein beyond the wild type dimer and very low association to form the homodimer mutant. Experimental results using nuclear importation and two-hybrid assays confirmed improved binding of the CC-CCmut1 heterodimer over the oncoprotein. Further experimental assays showed a decrease in the phosphorylation in the native oncoprotein and downstream activity in transformed cells following transcription of the CCmut1 protein construct. This decrease in activity of the oncoprotein induces caspase activity and leads to cell death through apoptosis.

References

1. Litowski J, Hodges R. Designing heterodimeric two-stranded alpha-helical coiled-coils. Effects of hydrophobicity and alpha-helical propensity on protein folding, stability, and specificity. Journal of Biological Chemistry 2002; 277: 37272-37279.

2. Lindhout D, Litowski J, Mercier P, Hodges R, Sykes B. NMR solution structure of a highly stable de novo heterodimeric coiled-coil. Biopolymers 2004; 75: 367-375.

3. Eisenberg D, McLachlan A. Solvation energy in protein folding and binding. Nature 1986; 319: 199-203.

4. Ooi T, Oobatake M, Nemethy G, Scheraga H. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proceedings of the National Academy of Sciences 1987; 84: 3086-3090.

5. Rizzo R, Aynechi T, Case D, Kuntz I. Estimation of absolute free energies of hydration using continuum methods: accuracy of partial charge models and optimization of nonpolar contributions. Journal of Chemical Theory Computation 2006; 2: 128-139.

6. Kellogg G, Semus S, Abraham D. HINT: a new method of empirical hydrophobic field calculation for CoMFA. Journal of Computer-Aided Molecular Design 1991; 5: 545-552.

7. Cozzini P, Fornabaio M, Marabotti A, Abraham D, Kellogg G, Mozzarelli A. Simple, intuitive calculations of free energy of binding for protein- ligand complexes. 1. Models without explicit constrained water. Journal of Medicinal Chemistry 2002; 45: 2469-2483.

8. Abraham D, Kellogg G, Holt J, Ackers G. Hydropathic analysis of the non-covalent interactions between molecular subunits of structurally characterized hemoglobins. Journal of Molecular Biology 1997; 272: 613-632.

9. Jelesarov I, Bosshard H. Thermodynamic characterization of the coupled folding and association of heterodimeric coiled coils (leucine zippers). Journal of Molecular Biology 1996; 263: 344-358.

10. Hodges R, Saund A, Chong P, St-Pierre S, Reid R. Synthetic model for two-stranded alpha-helical coiled-coils. Design, synthesis, and characterization of an 86-residue analog of tropomyosin. Journal of Biological Chemistry 1981; 256: 1214-1224.

11. Thompson K, Vinson C, Freire E. Thermodynamic characterization of the structural stability of the coiled-coil region of the bZIP transcription factor GCN4. Biochemistry 1993; 32: 5491-5496.

12. Zitzewitz J, Bilsel O, Luo J, Jones B, Matthews C. Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. Biochemistry 1995; 34: 12812-12819.

13. Bulheller B, Hirst J. DichroCalc--circular and linear dichroism online. Bioinformatics 2009; 25: 539-540.

14. Rost B, Sander C. Prediction of protein secondary structure at better than 70% accuracy. Journal of Molecular Biology 1993; 232: 584-584.